University of Mannheim

School of Social Sciences

New Methods for Job and Occupation Classification

by

Malte Schierholz

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Sozialwissenschaften
der Universität Mannheim

Dekan der Fakultät für Sozialwissenschaften	Prof. Dr. Michael Diehl
Erstbetreuerin	Prof. Dr. Frauke Kreuter
Zweitbetreuer	Prof. Dr. Matthias Schonlau
Gutachter	Prof. Dr. Florian Keusch
Gutachter	Prof. Dr. Jörg Drechsler
Tag der Disputation	6. Juni 2019

Contents

I Overview **1**

1 Introduction **2**

1.1 Motivation 2

1.2 Contributions 6

1.3 Acknowledgments 7

2 Summary of Dissertation Papers **8**

2.1 Occupation Coding During the Interview 8

2.1.1 Research Question 8

2.1.2 Results 10

2.2 An Auxiliary Classification with Work Activity Descriptions for Occupation Coding 11

2.2.1 Key Idea and Current Deficits 11

2.2.2 Principles for Development 12

2.3 Machine Learning for Occupation Coding - A Comparison Study 13

2.3.1 Algorithms 13

2.3.2 Results 15

3 Outlook and Perspectives **16**

4 Bibliography **20**

II Contributions **25**

1 Occupation Coding During the Interview **26**

1.1 Introduction 26

1.2 Background 28

1.3 Data and methods 31

1.4 Results and evaluation 38

1.5 Summary and conclusion 49

Appendix 55

Part A: Additional Results . . . 56
Part B: Questionnaire . . . 60
Part C: Instructions for Validation . . . 93
Part D: Behavior Coding Manual . . . 100

2 Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung 103
2.1 Einleitung . . . 108
2.2 Hintergrund: "Beruf" in der Statistik . . . 112
2.3 Leitgedanken zur Hilfsklassifikation . . . 125
2.4 Diskussion . . . 141
Anhang . . . 147
Formatierungsvorgaben . . . 148
Organisatorisches Vorgehen . . . 155
Ausgewählte Schwierigkeiten . . . 156
Beispiel: Verwendetes Material zur Bearbeitung der KldB-Berufskategorie 92122 Dialogmarketing - Fachkraft . . . 160
Excursion: Extensional Definition versus Prototypical Definition . . . 189

3 Machine Learning for Occupation Coding - A Comparison Study 191
3.1 Introduction . . . 191
3.2 Algorithms . . . 197
3.3 Research Questions . . . 208
3.4 Data . . . 213
3.5 Analytical Strategy . . . 214
3.6 Results . . . 217
3.7 Discussion . . . 227
Appendix . . . 239
Part A: Study Descriptions and Data . . . 241
Part B: Evaluation Metrics . . . 247
Part C: Model Tuning . . . 251
Part D: Detailed Results . . . 271
Part E: Similarity-based Reasoning: Connecting Approximate String Matching and a Hierarchical Bayesian Model . . . 308

Eidesstattliche Versicherung 324

Part I

Overview

Chapter 1

Introduction

1.1 Motivation

To find out how people live, ask about their occupation. Many answers will be quite informative, revealing more than just the type of worked performed (i.e., the tasks, duties, materials, and procedures used at work). The answer may also convey (stereotypical) information about the education (professional knowledge, required certificates) and about the social context (e.g., workplace, industry, membership in professional organizations, salary, occupational identity, legal privileges, public expectations, etc.). Dostal et al. (1998), Weeden and Grusky (2005), von der Hagen and Voß (2010) provide further background what is meant by 'occupation'. The concept is used by sociologists to derive various measures of social position (prestige scores, socio-economic indexes, class schemes) (e.g., Ganzeboom and Treiman, 2003), by epidemiologists to infer measures of work-related exposure to chemical agents (e.g., McGuire et al., 1998), and in uncountable other scientific endeavors. All this makes 'occupation' a standard demographic variable, collected in many surveys and censuses alike. Yet, conventional procedures to capture occupation are time-consuming because respondents typically describe their occupations using their own words and, in a subsequent step, human coders need to classify (=code) thousands of verbal responses according to extensive classification schemes.

The concept of 'occupation' is rather vague and can be operationalized in many different ways. Two classifications commonly used in Germany are the 2008 *International Standard Classification of Occupation* (ISCO, International Labour Office 2012), consisting of 436 occupational categories at the most detailed level, and the 2010 *German Classification of Occupations* (GCO, Bundesagentur für Arbeit 2011), consisting of 1286 occupational categories. Both classifications conceptualize 'occupation' similarly, emphasizing the tasks and duties performed in a given job (called the 'occupational activity' in the following).

The overarching goal of this thesis is the development of a novel instrument for precise, high-quality measurement of occupation in accordance with principles from both official

Figure 1.1: Several candidate job titles are suggested upon searching for "Schlosser" (metal worker). The complete list after clicking "Show All" contains 41 job titles.
Source: Screenshot from http://jobboerse.arbeitsagentur.de (accessed Feb. 2019)

classifications. The questions posed to respondents should minimize respondent burden. Post-interview coding processes should be as efficient as possible.

Two examples illustrate current coding procedures and their weaknesses. The first example, shown in Figure 1.1, stems not from survey research but is taken from the job search website of the German Federal Employment Agency. Visitors can enter some text, triggering several candidate job titles to appear. If the visitor selects a job title, there exists a link to a single 2010 GCO category, which is being coded. There are several concerns with this process. Firstly, if a visitor enters a job but cannot find an appropriate job title in the subsequent step, he will need to run another search or give up. In response, the third contribution in this thesis discusses algorithms meant to improve the suggestions. Secondly, the questionnaire design literature (e.g., Krosnick and Presser, 2010) recommends that all response options should be easily understood, unambiguous, and interpreted in the same way by everyone. In addition, response options should be mutually exclusive. All this helps to ensure that informants will always select the same response option if they wish to express identical pieces of information. By contrast, the job titles in the example are rather

technical, non-intuitive, do not emphasize occupational activities, and some of them relate to occupations with recognized vocational training programs, making it unlikely that the first requirement is met and difficult to (dis-)prove the second requirement. In response, the second contribution in this thesis develops alternative answer options. Thirdly, if a data analyst uses the 2010 GCO for his analysis, he will probably rely on category definitions from the classification to understand the content of the data. Yet, the job titles originally selected by informants and the category definitions used by the analyst have different connotations. Thus, the data analyst risks to misunderstand what the informant wished to express when selecting the job title. Since the answer options developed in the second contribution are closely aligned with the 2010 GCO, this is less of an issue with improved answer options.

The second example comes from the intended area of application, survey research. Some advice on how to measure occupation in German surveys has been published by Geis (2011), Paulus and Matthes (2013), Statistisches Bundesamt (2016), and Züll (2016). While most practitioners deviate from these recommendations (see appendix part A in the third contribution), the general procedure is often as follows. Respondents are typically asked two open-ended questions and, ideally, several closed questions. The responses need to be coded after the interview. To select the most appropriate category, deliberate decisions by one or more classification experts are usually regarded as optimal, but the costs may be prohibitive. In practice, verbal answers are sometimes compared automatically with a list of job titles specifically created for this purpose, or they can be coded according to rules and conventions coders have created together with their colleagues, possibly using specialized software for computer-assisted coding. To code more difficult cases, coders may need to consult the category definitions from an occupational classification, or they decide on the most appropriate category in a group discussion. Again, there are several concerns with this process. Firstly, manual coding is time-consuming. Secondly, it is inefficient to ask all respondents several questions about their job if, for some respondents, a single question would suffice to determine the correct category. Thirdly, other respondents do not provide all the information needed for unambiguous coding, despite being asked several questions, leading to errors in coding. Some studies have shown low rates of agreement between coders (Campanelli et al., 1997, Elias, 1997, Bound et al., 2001, Massing et al., 2019, further details are provided in contributions 1 and 3), questioning the reliability and validity of manual occupation coding. Finally, since outsiders can rarely replicate all the coding decisions, transparency and reproducibility of the coding process are yet another concern. In reaction to these shortcomings, an alternative classification procedure is developed in this thesis.

The first contribution develops a new instrument, a prototype for occupation coding during the interview, and reports its results. The prototype *combines* two specific developments that have been used separate from another before. Firstly, interviewers have coded occupations *during the interview* so that respondents can specify their responses further if

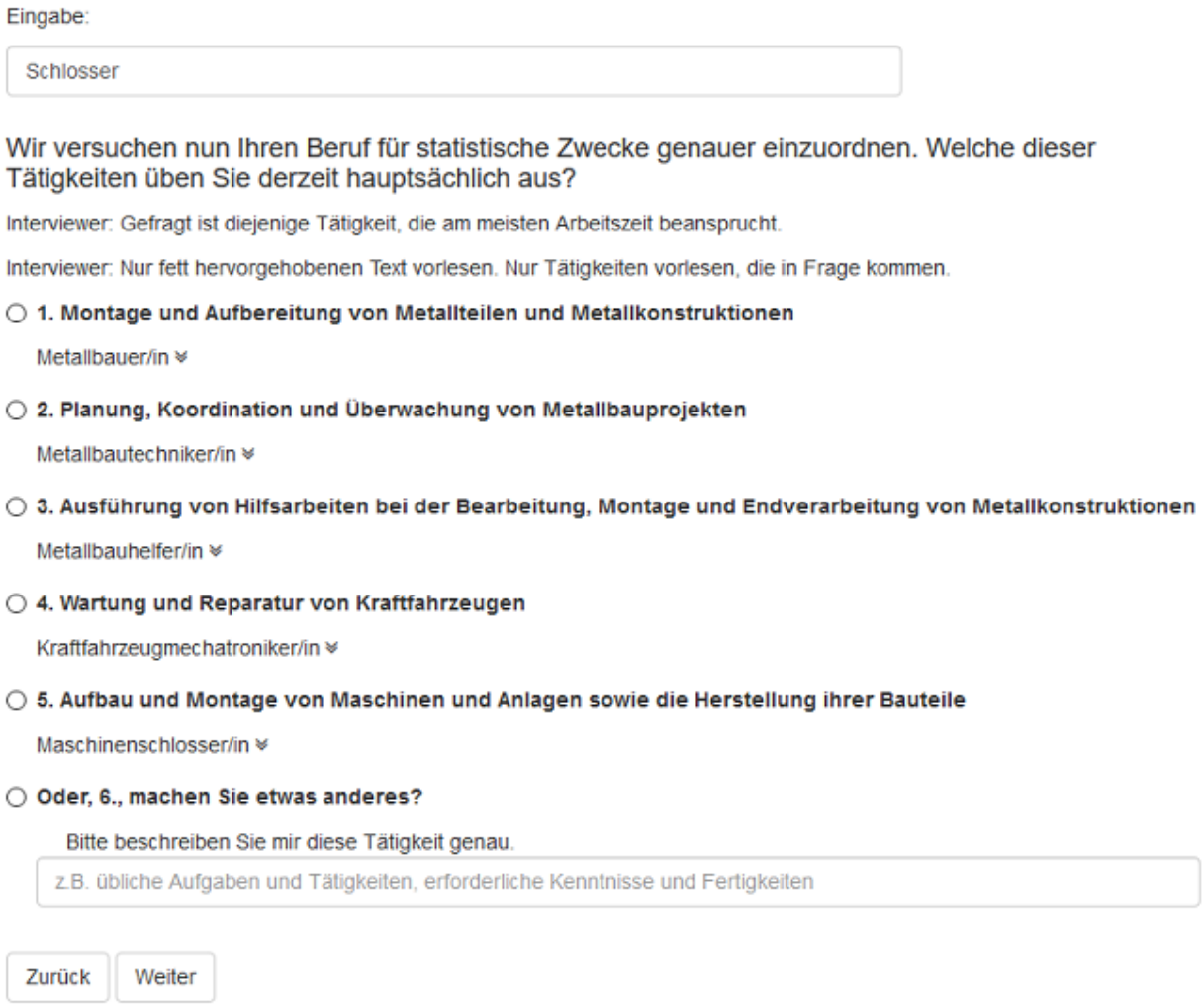


Figure 1.2: The new instrument: Several candidate occupational activities are suggested upon searching for "Schlosser" (metal worker). Compare with Figure 1.1.

needed. Secondly, post-interview coding has been done using *machine learning algorithms* and training data rather than simply matching answers with job titles from a coding index. The overall experiences with this prototype were promising. The subsequent contributions (contributions 2 and 3) react to shortcomings of the prototype and refine the instrument further. Thus, the two developments mentioned above run through all three contributions as a common thread, albeit contributions 2 and 3 make developments that could stand on their own. In particular, the second contribution develops answer options meant to be suggested during the interview, but possibly useful for existing coding software as well. The third contribution compares various machine learning algorithms and develops a new one, which will help to increase the performance wherever such algorithms are applied. Integrating all three contributions, Figure 1.2 provides an impression what the final instrument developed in this thesis looks like. It is expected that this novel instrument will help improve current shortcoming in occupational data collection.

1.2 Contributions

The following contributions are part of this thesis. A summary is provided in chapter 2.

	Full references and related work
Contribution 1	Malte Schierholz, Miriam Gensicke, Nikolai Tschersich, Frauke Kreuter (2018). Occupation coding during the interview, *Journal of the Royal Statistical Society: Series A* **181**(2): 379–407. URL: https://doi.org/10.1111/rssa.12297 • An earlier version is available in – Malte Schierholz, Miriam Gensicke, Nikolai Tschersich (2016). Occupation coding during the interview, *IAB-Discussion Paper* 17/2016, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, 30 p. URL: https://www.iab.de/389/section.aspx/Publikation/k160512302
Contribution 2	Malte Schierholz, Lorraine Brenner, Lea Cohausz, Lisa Damminger, Lisa Fast, Ann-Kathrin Hörig, Anna-Lena Huber, Theresa Ludwig, Annabell Petry, Laura Tschischka (2018). Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung * Leitgedanken und Dokumentation, *IAB-Discussion Paper* 13/2018, Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, 43 p. URL: https://www.iab.de/183/section.aspx/Publikation/k180509301 • The key principles were summarized and published in – Malte Schierholz (2018). Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung, *AStA Wirtschafts- und Sozialstatistisches Archiv* **12**(3): 285–298. URL https://doi.org/10.1007/s11943-018-0231-2 • The latest version of the auxiliary classification is available at https://github.com/malsch/occupationCodingAuxco
Contribution 3	Malte Schierholz (to be submitted). Machine Learning for Occupation Coding - A Comparison Study, 47 p. • An associated R-package is available at https://github.com/malsch/occupationCoding

1.3 Acknowledgments

Funding for this work has been provided by grant KR 2211/3-1 from the German Research Foundation, by the German Institute for Employment Research (IAB), and by the Mannheim Centre for European Social Research (MZES) at the University of Mannheim.

Chapter 2

Summary of Dissertation Papers

The first contribution motivates, develops, and evaluates a first prototype for "Occupation Coding During the Interview" (Schierholz, Gensicke, Tschersich and Kreuter, 2018). The subsequent contributions refine the idea.

The second contribution titled "An Auxiliary Classification with Work Activity Descriptions for Occupation Coding" develops improved answer options that will be suggested to respondents (Schierholz, Brenner, Cohausz, Damminger, Fast, Hörig, Huber, Ludwig, Petry and Tschischka, 2018).

The third contribution "Machine Learning for Occupation Coding - A Comparison Study"—first published as part of this thesis—compares several algorithms and develops a new one, aiming to make optimal suggestions.

Each contribution is summarized in the following.

2.1 Occupation Coding During the Interview

The paper describes a first prototype implementing occupation coding during the interview. The prototype was tested in a telephone survey commissioned by the Institute for Employment Research and conducted by Kantar Public.

2.1.1 Research Question

Schierholz, Gensicke, Tschersich and Kreuter (2018) illustrate their proposed technique for interview coding with the following example:

> [C]onsider a respondent who answers 'vice director [*sic*] and teacher' when asked about his job activities. On the basis of this *verbatim* answer and, if desired, further input from the interview, a computer algorithm searches for possible occupations and calculates associated probabilities at the time of the interview. The job titles that were found to be most likely are then suggested in closed-ended question format to

> the interviewer, who in turn asks the respondent to select the most appropriate occupation among these suggestions. The suggestions for the above-mentioned example are shown in Fig. 1 [MS: see p. 27]. Since we cannot guarantee that the algorithm will always suggest an accurate job title, suggestions are complemented by a last answer option 'or do you work in a different occupation?'. If this option is chosen, further questions should be asked to gather additional details about the person's job; if not, coding is complete. In the example, the job title 'Teacher—elementary school' was selected, capturing a detail, the school type, that was not provided in the original verbal response. (p. 381)

Importantly, the new technique was tested in an interviewer-administered survey and designed specifically for this purpose. This differs from the example mentioned above ("Schlosser", see Figure 1.1), which has been implemented as part of a public website. On a website, visitors can navigate through dozens of job titles. By contrast, the new instrument suggests at most five job titles. Moreover, there is no auto-completion feature built in as in the "Schlosser"-example, but the job titles are suggested as a follow-up question after the verbal answer has been saved. This question sequence helps interviewers because they are used to this format and ensures the collection of codable information from all respondents.

The intended area of application is statistical data collection with surveys. From this perspective it is meaningful to compare the new instrument with current practice in this field. Schierholz, Gensicke, Tschersich and Kreuter (2018) continue by summarizing their goals as follows.

> With this new approach, we pursue three fundamental objectives to improve current shortcomings in the data collection process. First, we aim to reduce *coding errors* that arise from missing data or contradictory information provided by respondents. Respondents' verbal answers are sometimes ambiguous and difficult to code, in particular when survey questions are not aligned with the theoretical concepts that underlie occupational classification systems. Answer options often help to clarify the meaning of survey questions. Suggesting a limited number of answer options from the occupational classification based on initial verbal responses thus is expected to improve the measurement, while limiting respondent burden. Second, we seek to maximize the number of interview-coded answers to *minimize efforts for coding the residual cases* after the interview. Third, we aim to *save valuable interview time*, thereby reducing the respondent burden. The closed-ended question that is shown in Fig. 1 [MS: see p. 27] can often replace the additional open-ended question about occupation that is used in many questionnaires so that the total interview duration decreases. A final key advantage of the new instrument is the supervised learning algorithm that predicts possible job titles. The predictions are based on training data from past studies and can be *improved as more data become available.* (pp. 381-382)

Note how this algorithm differs from more traditional algorithms used for occupation coding. Often, these algorithm match verbal answers with a predefined list of job titles (Geis, 2011, Elias et al., 2014). If a similar or identical job title is found in this list, the corresponding code is assigned. By contrast, the algorithm used here is a machine learning algorithm, which makes predictions based on previously coded data. While such algorithms have been developed before, a key innovation in this paper is its use for occupation coding at the time of the interview.

The specific algorithm used to predict possible job titles originates from the author's Master thesis (Schierholz, 2014). It was later adapted for this specific application. 26 scores are calculated in parallel using a range of different matching techniques and statistical methods. Each of these scores, by design, should correlate with the probability that a given job title will be selected. Tree boosting is used to combine the 26 scores into a single prediction, an approach known as stacking in the machine learning literature. The algorithm is described in detail and evaluated as part of this contribution.

2.1.2 Results

The paper proves the feasibility of interview coding. Implementing the complex interview coding instrument in a telephone survey is possible. 72.4% of the respondents selected a job title during the interview, reducing the workload for post-interview coding. The new instrument can shorten the total interview duration by a few seconds because it would replace a standard question which is only being asked for post-interview coding purposes.

To assess the quality, two professional coders coded the data and two student assistants checked the correctness of professional coding and interview coding. Interview coding is competitive with one professional coder, but slightly worse than the other. However, the differences are small and may not matter in practice. The goal to reduce coding errors with interview coding is not achieved. Two examples are described in detail, illustrating possible weaknesses of interview coding.

In 13.6% of the cases, respondents receive suggestions by the algorithm but do not select a job title. Although the proposed system has no benefits for these respondents, they have to bear the burden of an additional question, increasing the length of their interviews. To counter this problem, one could tweak the algorithm to make suggestions only if the verbal answer meets certain conditions, preventing poor suggestions. If the algorithm was improved, the size of this problematic group would be reduced to 5.6% of the population; however, the proportion of respondents who select a job title would decrease to 61.3% if fewer respondents received suggestions.

The paper describes a prototype for occupation coding during the interview. As such, several features were tested that turned out to be worthless. Other factors are mentioned that should be improved upon. In particular, the two subsequent contributions in this thesis

follow directly from concerns expressed in this first paper. Schierholz, Gensicke, Tschersich and Kreuter (2018) write:

> When respondents choose one of the job titles suggested, it is too often not the most appropriate. Respondents frequently select general job titles that are not entirely wrong but link to suboptimal GCO categories. These inappropriate job titles stem from the *Dokumentationskennziffer*, which is therefore not well suited for coding during the interview. To preclude the possibility that respondents select an incorrect category, we recommend the development of an auxiliary classification that describes answer options more precisely. (pp. 403-404)

This leads directly to the second contribution in this thesis, which develops such an auxiliary classification.

There is also the idea that the algorithm could be improved further and more training data will help as well to obtain more useful suggestions. The third contribution thus explores different algorithms, which are being trained with larger training data.

2.2 An Auxiliary Classification with Work Activity Descriptions for Occupation Coding

The second contribution develops answer options describing occupational activities meant to be suggested to respondents.

To provide a conceptual basis, the contribution provides some background knowledge about occupations, job titles, and occupational activities, aiming to explain what researchers mean if they want to measure 'occupation'. It describes the principles that underlie occupational classifications and compares the 2010 GCO and the 2008 ISCO. Answer options were developed in close alignment with these official classifications, explaining why the product is called an 'auxiliary' classification.

2.2.1 Key Idea and Current Deficits

Occupational classifications describe several principles relevant for occupational measurement, but do not contain exact specifications on how to classify employed persons. The second contribution proposes the following gold standard for accurate coding, building on principles mentioned in occupational classifications.

Employed persons should be classified

- on the basis of the occupational activity actually performed (not on the basis of available answers given by respondents)

- into the most appropriate occupational category at the finest level of the classification (category definitions should be used to determine the most appropriate category)

- whereby in ambiguous situations the coding decision should be based on the subset of occupational activities in which the employee spends most of his working time.

This gold standard for coding is not achievable in practice because coders do not know all necessary details about a person's occupational activity. Few respondents describe their occupational activity in sufficient detail. Many answers are rather imprecise or may depend on the respondents' vocational training, which is not of interest.

As an alternative, one could give all category definitions to the respondent and let him choose the most appropriate category. Of course, this is highly impractical because the category definitions are hundreds of pages long. Approximating this approach, the auxiliary classification summarizes the category definitions from both official classifications, allowing respondents to choose the most appropriate category summary.

Besides this gold standard argument, three practical issues have lead to the development of the auxiliary classification:

- In many data collections it is desired to code answers simultaneously into both the GCO and the ISCO. Every category from the auxiliary classification is linked with both official classifications, supporting this use case.

- Many job titles are imprecise, overly general, or ambiguous. This makes them unsuited to communicate about work activities in an unambiguous way.

- There exist many different job titles that describe very similar and overlapping jobs. Computer-assisted coding systems promise a clear display of relevant answer options, but large numbers of job titles would thwart this goal.

2.2.2 Principles for Development

Put in highly simplified terms, the categories from the auxiliary classification were created by summarizing the content of the official category definitions. A few thousand person hours were needed to do this work as careful as possible. Nine student assistants helped with this task, supervised by the author.

Both classifications, GCO and ISCO, state that their respective categories are mutually exclusive. The auxiliary categories are created as the intersections between any two categories from GCO and ISCO. Mathematically, it follows that categories from the auxiliary classification are also mutually exclusive, fulfilling a common requirement for answer options in surveys. Moreover, if the correct category from the auxiliary classification gets selected, the underlying categories from GCO and ISCO must be correct as well.

Describing the intersections between the GCO and ISCO categories has been challenging. Auxiliary categories must be understood by interviewers and respondents. For this reason the category summaries in each auxiliary category need to be short and clear. In addition, the language needs to be precise, because answer options described in vague words would tend to overlap (Tourangeau et al., 2000).

The auxiliary classification developed as part of this thesis consisted of 1226 categories, but has not been tested yet. After its first publication the auxiliary classification has been updated to facilitate comprehension. The most recent version is available at https://github.com/malsch/occupationCodingAuxco.

2.3 Machine Learning for Occupation Coding - A Comparison Study

Occupation coding can be time-consuming and expensive if thousands or millions of answers need to be coded. Computers are commonly used to automate the task, either applying *computer-assisted coding* or *automated coding* (United Nations Statistical Commission and Economic Commission for Europe, 1997, Speizer and Buckley, 1998). With computer-assisted coding, a computer program suggests a few relevant categories and a human coder selects the most appropriate one. With automated coding, a human decision is not needed, but the computer program selects the highest-ranked category all by itself. Both approaches typically depend on algorithms calculating scores and ranking the suggested categories by their likelihood to be correct.

Interview coding is a special form of computer-assisted coding. To make it work, an algorithm calculating scores and suggesting possible categories is needed. This was done using a machine learning approach in the first contribution, but the authors call for an improved algorithm, which should be trained with additional data. The third contributions describes subsequent developments towards this goal.

2.3.1 Algorithms

Occupational data has certain characteristics distinguishing it from other data sets. On the one hand, it is high-dimensional with 10.000s of predictors (words) that might be mentioned by respondents and 100s or 1000s of outcome categories defined in occupational classifications. On the other hand, most verbal answers are short (one-word-long job titles) and sometimes misspelled. These characteristics are unusual in machine learning. Still, an optimal algorithm must take these factors into account, complicating its development.

Based on an extensive review of the literature, the paper distinguishes two types of algorithms. The classical approach consults a coding index to find an appropriate category

for a given answer. If the verbal answer from the respondent is identical (algorithm 1) or similar (algorithm 2, Elias et al., 2014) to an entry from the coding index, the corresponding code is assigned. The machine learning approach, in contrast, does not depend on a coding index but calculates possible codes from previously coded data, known as training data. Two general strategies how training data can be used have been reported in the occupation coding literature. The first strategy calculates similarities between a new verbal answer to be coded and all previous answers from the training data. This can be implemented in various ways, differing in how the similarities are calculated and in how the final scores are calculated from a set of most similar training observations (algorithm 3 by Creecy et al. (1992) and algorithm 4 by Gweon et al. (2017) implement two such variants). The second strategy, based on loss minimization, estimates a function linking the verbal input with the outcome categories. The resulting function will depend on a predetermined functional form, on the loss function, and on the optimization algorithm used (algorithm 5 refers to regularized multinomial logistic regression as implemented by Friedman et al. (2010) and algorithm 6 refers to XGBoost, a tree boosting algorithm, as implemented by Chen and Guestrin (2016)). Strengths and weaknesses of each algorithm are demonstrated in an extensive comparison.

Predicting an occupational category is only possible if the verbal answer to be coded is not completely unknown to the system. Both the coding index and the training data provide overlapping but distinct information about possible categories, suggesting that the results could be improved if the predictions were based on both sources. Linking misspelled answers to these sources should also help, possible by using string similarities, but not yet common in the occupation coding literature using machine learning. To improve upon existing algorithms, a novel algorithm is proposed that makes use of both the coding index and the training data and is based on string similarity calculations (algorithms 7-9, different only in the similarity calculations used). A final algorithm under comparison (algorithm 10) is an ensemble of algorithms 6 to 9.

Schierholz summarizes the paper as follows:

> This paper reviews and compares various techniques to calculate such scores, focusing on algorithms from machine learning. The algorithms under comparison are carefully selected and several less-promising ones have been discarded. A novel algorithm is introduced that may stimulate future research. For the evaluation we take an applied perspective, asking how many answers could be coded automatically and how often the assigned categories would be identical with manual coding. Using five data sets we have at our disposal, we showcase how much the results depend on the respective choice of training and test data. This allows us to identify the most competitive algorithms. (p. 5)

All algorithms except algorithm 2 are implemented as part of an R package, available

at https://github.com/malsch/occupationCoding.

2.3.2 Results

The paper contains a plethora of results, demonstrating how much the results differ depending on the training data available and the intended application (automated coding, computer-assisted coding, interview coding). Since this thesis focuses on interview coding, only the key results relevant for this application are highlighted in the following.

To report results about future expected performance, the test data should be representative of the intended application. The data set collected in the first contribution was chosen because it represents the population of interest, adult employed persons. Also, the quality from manual coding is believed to be high in this data set because coders had access to information from several additional variables about the respondents' occupations.

Algorithms 2, 8, and 10 perform best in different situations, depending on the training data used. Algorithm 2 is best if a few training observations are available. Algorithm 8 is competitive if a medium number of training observations are available. The best results are obtained when using as many training observations as possible after pooling different data sets. In this situation algorithm 10 outperforms all others.

It is recommended to suggest categories for 75% of the respondents having the highest scores. The remaining 25% should not receive suggestions because these suggestions are probably incorrect, making it unlikely for respondents to select an appropriate category, and interview time would be wasted. For the subset of respondents who receive suggestions, it is of interest how many will select a category. While this remains to be tested in practice, a useful proxy measurement is available. It is found that coders selected for 84% of the respondents in this subset a category that would have been shown to respondents as one of the top five highest-ranked suggestions, indicating that the suggestions are usually relevant. Since respondents will see the suggestions but coders did not, this number is a conservative estimate of how often respondents will select one of the categories suggested. Given that respondents tend to endorse statements in questionnaires, known as acquiescence (Krosnick and Presser, 2010), one can speculate that respondents will select one of the suggested categories more often than coders, even if the selected category is only loosely fitting. These numbers presuppose that occupational categories from the 2010 GCO would be suggested. In practice, it is planned to link the GCO categories with the auxiliary classification (contribution 2) and suggest auxiliary categories instead.

Chapter 3

Outlook and Perspectives

The first contribution to this thesis demonstrates the feasibility and promises of interview coding. The subsequent contributions fix different shortcomings of the first prototype. In particular, the second contribution develops answer options tailored for interview coding and the third contribution explores optimal algorithms to make the suggestions. Both contributions change the original prototype in fundamental ways.

Improving the original prototype is ongoing research. The final instrument should work in different modes of data collection, i.e., in computer-assisted personal interviews, in computer-assisted telephone interviews, and in web surveys. End users should find its handling simple and effective. Programmers should be able to integrate the instrument into the questionnaire without technical hurdles. While not yet finished, I have programmed a web service to accomplish these goals. Figures 1.2 and 3.1 demonstrate what it currently looks like. Both examples were selected to allow comparisons with other figures in this thesis. The next step will be to test the improved instrument in practical settings.

At the current point in time, prior to a renewed test, it is highly speculative how the updated instrument will perform. It is very likely that the instrument can be improved further—this might be necessary to fulfill highest scientific standards or to make it suitable for widespread application. Various ideas could be tested. Yet, in the face of frequent disagreement among human coders and no established gold standard for how to determine the 'correct' category, a central challenge will be to develop and/or apply criteria for evaluation, against which different instruments for occupational measurement could be judged.

On the applied side, possibilities to improve the instrument further include: Should the text box to enter verbal answers have an auto-completion feature (as in Google Search), reducing spelling errors? How many answer options should be shown? What is an optimal graphical design and which visual aids could help interviewers and respondents? Should interviewers carry out the interviews in a more standardized or in a conversational format?

On the more algorithmic side, possibilities include: It is expected that the new instrument can reduce measurement error, implying that respondents will choose different

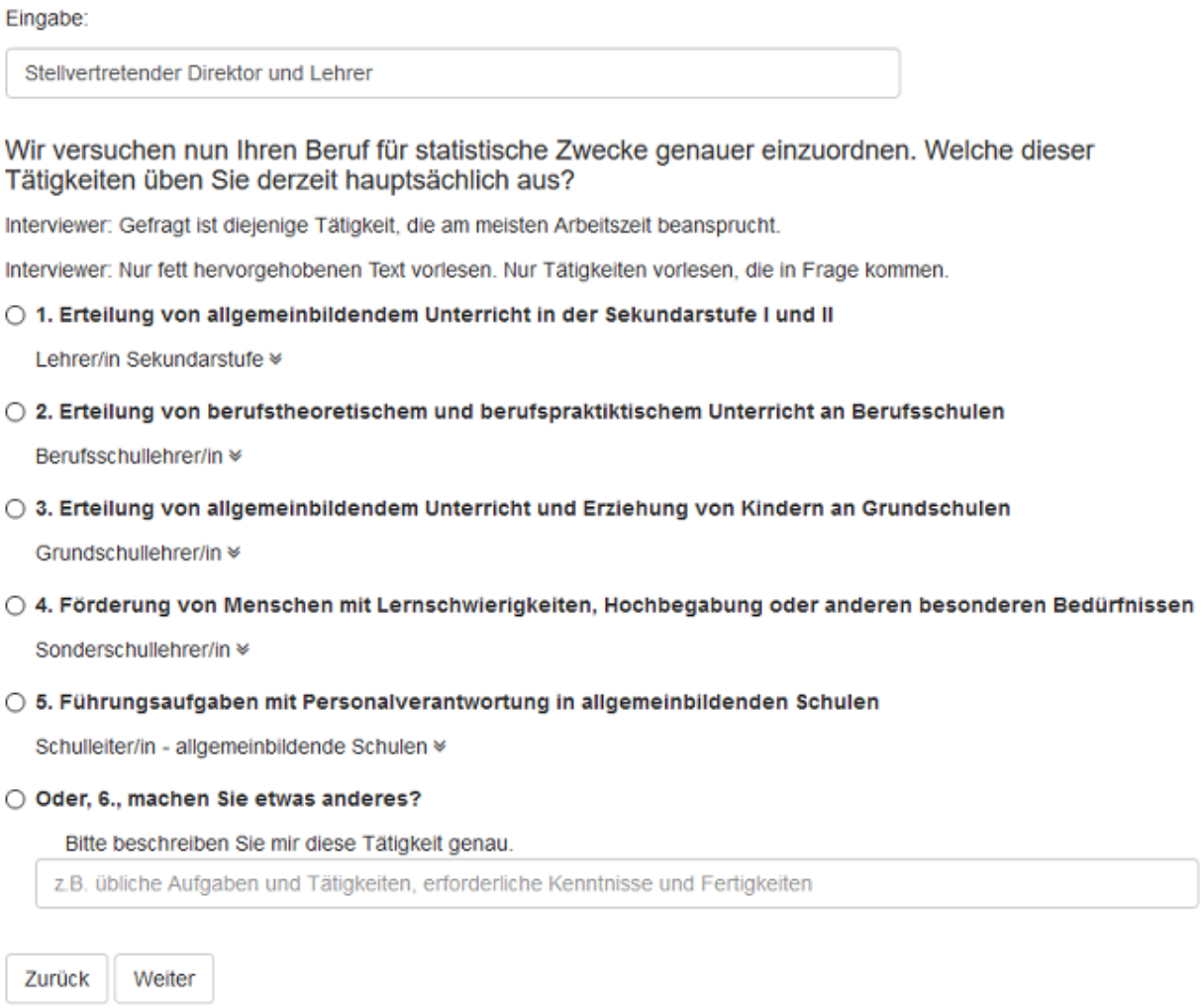


Figure 3.1: The new instrument: Several candidate occupational activities are suggested upon searching for "Stellvertretender Direktor und Lehrer" (vice director and teacher). Compare with Fig. 1 inside the first contribution (p. 27)

categories than coders currently do. Respondents choices can be seen as additional (better?) training observations that become available one by one—and techniques from 'online learning' (Shalev-Shwartz, 2007) could be implemented to update the prediction algorithm with every new observation. There is also the potential problem that the proposed machine learning algorithm suggests for any given verbal input always the same answer options in a deterministic manner, entailing that respondents have no chance to select alternative answer options. This risks to bias the selections from all respondents towards selecting one of the answer options suggested and away from any alternative answer option, paralleling the issues of interviewer effects. To alleviate this concern, the algorithm could suggest different response options at random proportional to their predicted probabilities. On average over many responses this could counterbalance the individual biases that every respondent is exposed to.

On the conceptual side, much depends on the auxiliary classification (contribution 2)

and how 'occupation' is conceptualized therein. In accordance with current standards in Germany, the auxiliary classification aims to measure 'berufliche Tätigkeiten' (occupational activities). Such occupational activities were developed based on both the 2010 GCO and the 2008 ISCO. Yet, describing the essential, characteristic aspects of an occupational activity is difficult. An evaluation of the auxiliary classification has not been done yet. To be successful, respondents should perfectly understand the occupational activities described in the auxiliary classification and, after they select an activity, no others but the associated categories from the 2010 GCO and the 2008 ISCO should be correct.

If the approach taken in the auxiliary classification using occupational activities would fail, alternative conceptualizations could be developed. I suggest two connections I was not aware of at the time when the auxiliary classification was developed. Firstly, even though contribution 2 makes a strong argument against job titles, it might be too early to abolish them completely. One study concerned with occupational identity finds that music teachers have low attachment to their occupational tasks, but identify more with their job titles (Rewolinski, 2014). This suggests that job titles might be better suited if the goal were to measure respondents' occupational identity and not the occupational activities they perform. To exemplify this point further, an economist working at a central bank might be classified according to his field of study (economist, GCO category: 91404) or according to his employer (central bank official, GCO category: 72184). A possible decision criterion might be how much he identifies with either option. Since data collected in surveys is mainly about subjective views respondents hold about themselves and their environment, not necessarily about observable facts from that environment, it might be promising to tailor an occupational question towards their occupational identity.

Secondly, the concept of 'occupational activity' is very similar with that of an 'economic activity'. The *International Standard Industrial Classifications of All Economic Activities (ISIC), Rev.4* (United Nations Department of Economic and Social Affairs, 2008) has been created to classify statistical units based on their economic activity. ISIC is widely accepted and part of an internationally harmonized system of industrial classifications. Of course, these classifications have been created to classify establishments and enterprises, not individual persons, according to their economic activity. However, large enterprises can pursue more than one economic activity and it is not uncommon that different subdivisions of an enterprise pursue different economic activities. From an atomistic view, employees may be regarded as the smallest unit in an enterprise, each pursuing their own economic activity, which is then the same as an occupational activity. Besides this argument, industrial classifications and occupational classifications have the same historical origin. Splitting the original classification into two distinct classifications for employees and establishments has been subject to much debate (International Labour Office, 1923, Meerwarth, 1925, Willms, 1983). Yet, even after decades of separate development, occupational classifications are in

large parts organized by similar principles as industrial classifications. It may be worth to elaborate the similarities and differences between both classifications further, possibly with the result to reunite both classifications.

Both alternatives, conceptualizing occupation as an 'occupational identity' or as an 'economic activity', were not explored further in this thesis. How these ideas compare with the current conceptualization as an 'occupational activity' is unknown. Which conceptualization researchers prefer may well depend on their area of research.

A novel, innovative instrument for occupation coding during the interview has been developed in this thesis. As argued in the first contribution, the instrument has the potential to increase data quality while reducing the costs of collecting the data. Although this is an exciting development in itself, it misses a larger point relating to principles. The thesis proposes a paradigm shift from traditional coding procedures to interview coding. The new process is more transparent because it does not rely on coders' subjective reasoning. Instead, asking respondents and having them select an answer option is a widely recognized standard in survey research. The new instrument complies with this standard because respondents choose the most appropriate occupational activity by themselves—according to their own knowledge about their job.

Bibliography

Bound, J., Brown, C. and Mathiowetz, N. (2001). Chapter 59 - Measurement Error in Survey Data, Vol. 5 of *Handbook of Econometrics*, Elsevier, pp. 3705 – 3843.
URL: *https://doi.org/10.1016/S1573-4412(01)05012-7*

Bundesagentur für Arbeit (2011). *Klassifikation der Berufe 2010*, Bundesagentur für Arbeit, Nuremberg.

Bundesanstalt für Arbeit (1988). *Klassifizierung der Berufe*, Bundesanstalt für Arbeit, Nuremberg.

Campanelli, P., Thomson, K., Moon, N. and Staples, T. (1997). The quality of occupational coding in the United Kingdom, *in* L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz and D. Trewin (eds), *Survey Measurement and Process Quality*, Wiley, New York, pp. 437–453.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, New York, NY, USA, pp. 785–794.
URL: *http://doi.acm.org/10.1145/2939672.2939785*

Creecy, R. H., Masand, B. M., Smith, S. J. and Waltz, D. L. (1992). Trading MIPS and memory for knowledge engineering, *Commun. ACM* **35**(8): 48–64.
URL: *http://doi.acm.org/10.1145/135226.135228*

Dostal, W., Stooß, F. and Troll, L. (1998). Beruf – Auflösungstendenzen und erneute Konsolidierung, *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* **31**(3): 438–460.
URL: *http://doku.iab.de/mittab/1998/1998_3_MittAB_Dostal_Stooss_Troll.pdf*

Elias, P. (1997). Occupational classification (ISCO-88): Concepts, methods, reliability, validity and cross-national comparability, *OECD Labour Market and Social Policy Occasional Papers 20*, OECD Publishing, Paris.
URL: *http://dx.doi.org/10.1787/304441717388*

Elias, P., Birch, M. and Ellison, R. (2014). CASCOT International version 5, *User Guide*, Institute for Employment Research, University of Warwick, Coventry.
URL: *http://www2.warwick.ac.uk/fac/soc/ier/software/cascot/internat/*

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* **33**(1): 1–22.
URL: *https://doi.org/10.18637/jss.v033.i01*

Ganzeboom, H. B. G. and Treiman, D. J. (2003). Three internationally standardised measures for comparative research on occupational status, *in* J. H. P. Hoffmeyer-Zlotnik and C. Wolf (eds), *Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables*, Springer US, Boston, MA, pp. 159–193.
URL: *https://doi.org/10.1007/978-1-4419-9186-7_9*

Geis, A. (2011). Handbuch für die Berufsvercodung, *Coding Documentation*, GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim.
URL: *http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/handbuch_der_berufscodierung_110304.pdf*

Geis, A. and Hoffmeyer-Zlotnik, J. H. (2000). Stand der Berufsvercodung, *ZUMA-Nachrichten* **24**(47): 103–128.
URL: *https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_47.pdf*

Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M. and Steiner, S. (2017). Three methods for occupation coding based on statistical learning, *Journal of Official Statistics* **33**(1): 101–122.
URL: *http://dx.doi.org/10.1515/JOS-2017-0006*

International Labour Office (1923). Systems of Classification of Industries and Occupations, *Studies and Reports, Series N, No. 1*, International Labour Office, Geneva.
URL: *https://www.ilo.org/public/libdoc/ilo/ILO-SR/ILO-SR_N1_engl.pdf*

International Labour Office (2012). *International Standard Classification of Occupations: ISCO-08*, International Labour Organization, Geneva.

Krosnick, J. A. and Presser, S. (2010). Question and questionnaire design, *in* P. V. Marsden and J. D. Wright (eds), *Handbook of Survey Research, 2nd Edition*, Emerald Group, Bingley, pp. 263–313.

Massing, N., Wasmer, M., Wolf, C. and Zuell, C. (2019). How Standardized is Occupational Coding? A Comparison of Results from Different Coding Agencies in Germany, *Journal of Official Statistics* **35**(1): 167–187.
URL: *http://dx.doi.org/10.2478/JOS-2019-0008*

McGuire, V., Nelson, L. M., Koepsell, T. D., Checkoway, H. and Longstreth, W. T. (1998). Assessment of occupational exposures in community-based case-control studies, *Annual Review of Public Health* **19**(1): 35–53. PMID: 9611611.
URL: *https://doi.org/10.1146/annurev.publhealth.19.1.35*

Meerwarth, R. (1925). Nationalökonomie und Statistik: Eine Einführung in die empirische Nationalökonomie, *Handbuch der Wirtschafts- und Sozialwissenschaften in Einzelbänden, Bd. 7*, Walter de Gruyter, Berlin.

Müller, A. (2014). The implementation of the German Classification of Occupations 2010 in the IAB Job Vacancy Survey, *IAB-Forschungsbericht 10/2014*, Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
URL: *https://www.iab.de/185/section.aspx/Publikation/k140924302*

Paulus, W. and Matthes, B. (2013). Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel, *FDZ-Methodenreport 08/2013*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
URL: *http://fdz.iab.de/187/section.aspx/Publikation/k131014a03*

Rewolinski, C. (2014). *The Measurement of Occupational Identity Among Undergraduate Preservice Music Teachers: a Test Development Study*, PhD thesis, University of North Texas, University of North Texas Libraries, Digital Library.
URL: *digital.library.unt.edu/ark:/67531/metadc699995/*

Rosch, E. (1978). Principles of categorization, *in* E. Rosch and B. Lloyd (eds), *Cognition and Categorization*, Erlbaum, Hilldsdale, NY, pp. 27–48.

Schierholz, M. (2014). Automating Survey Coding for Occupation, *FDZ-Methodenreport 10/2014*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
URL: *https://fdz.iab.de/342/section.aspx/Publikation/k141027302*

Schierholz, M., Brenner, L., Cohausz, L., Damminger, L., Fast, L., Hörig, A.-K., Huber, A.-L., Ludwig, T., Petry, A. and Tschischka, L. (2018). Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung * Leitgedanken und Dokumentation, *IAB-Discussion Paper 13/2018*, Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
URL: *https://www.iab.de/183/section.aspx/Publikation/k180509301*

Schierholz, M., Gensicke, M., Tschersich, N. and Kreuter, F. (2018). Occupation coding during the interview, *Journal of the Royal Statistical Society: Series A* **181**(2): 379–407.
URL: *https://doi.org/10.1111/rssa.12297*

Shalev-Shwartz, S. (2007). *Online learning: theory, algorithms and applications*, PhD thesis, Hebrew University of Jerusalem, Hebrew University of Jerusalem.
URL: *http://shemer.mslib.huji.ac.il/dissertations/W/JMC/001440314_1.pdf*

Speizer, H. and Buckley, P. (1998). Automated coding of survey data, *in* M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II and J. M. O'Reilly (eds), *Computer Assisted Survey Information Collection*, Wiley, New York, pp. 223–243.

Sperling, H. (1961). Zur Theorie und Methode der Berufsklassifizierung, *Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft* **81**: 705–720.

Statistisches Bundesamt (1961). *Klassifizierung der Berufe*, Kohlhammer, Stuttgart.

Statistisches Bundesamt (2016). *Demographische Standards*, Statistisches Bundesamt, Wiesbaden.

Tourangeau, R., Rips, L. J. and Rasinski, K. (2000). *The Psychology of Survey Response*, Cambridge University Press.

United Nations Department of Economic and Social Affairs (2008). *International Standard Industrial Classification of All Economic Activities (ISIC) Revision 4* , United Nations, New York.

United Nations Statistical Commission and Economic Commission for Europe (ed.) (1997). *Statistical Data Editing Volume No. 2*, United Nations, New York, chapter 6.
URL: *http://www.unece.org/stats/publications/editing/SDE2.html*

von der Hagen, A. D. and Voß, G. G. (2010). Beruf und Profession, *in* F. Böhle, G. G. Voß and G. Wachtler (eds), *Handbuch Arbeitssoziologie*, VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 751–803.
URL: *https://doi.org/10.1007/978-3-531-92247-8_26*

Weeden, K. A. and Grusky, D. B. (2005). The case for a new class map, *American Journal of Sociology* **111**(1): 141–212.
URL: *https://doi.org/10.1086/428815*

Willms, A. (1983). Historische Berufsforschung mit amtlicher Statistik. Rekonstruktion der Entwicklung der Berufsstatistik in Deutschland und Entwurf einer Klassifikation vergleichbarer Berufsfelder, 1925-1980, *VASMA-Projekt Arbeitspapier Nr. 30*, Universität Mannheim, Mannheim.

Züll, C. (2016). The Coding of Occupations, *Technical report*, GESIS – Leibniz Institute for the Social Sciences, GESIS Survey Guidelines. Mannheim.
URL: *https://doi.org/10.15465/gesis-sg_en_019*

Part II

Contributions

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Statistics in Society
Series A

J. R. Statist. Soc. A (2018)
181, *Part* 2, *pp.* 379–407

Occupation coding during the interview

Malte Schierholz,

University of Mannheim, and Institute for Employment Research, Nuremberg, Germany

Miriam Gensicke and Nikolai Tschersich

Kantar Public, Munich, Germany

and Frauke Kreuter

University of Maryland, College Park, USA, University of Mannheim and Institute for Employment Research, Nuremberg, Germany

[Received April 2016. Final revision April 2017]

Summary. Currently, most surveys ask for occupation with open-ended questions. The verbal responses are coded afterwards, which is error prone and expensive. We present an alternative approach that allows occupation coding during the interview. Our new technique uses a supervised learning algorithm to predict candidate job categories. These suggestions are presented to the respondent, who in turn can choose the most appropriate occupation. 72.4% of the respondents selected an occupation when the new instrument was tested in a telephone survey, entailing potential cost savings. To aid further improvements, we identify some factors for how to increase quality and to reduce interview duration.

Keywords: Coding; Interview coding; Measurement error; Occupation; Open-ended questions; Supervised learning

1. Introduction

Occupation is a core organizational principle in our society. Researchers from many disciplines have an interest in measuring occupation, e.g. to capture individuals' tasks and duties for economic studies, to measure the health risk from a person's job or to determine the person's status in society for sociological research, for example in terms of the '*Standard international occupational prestige scale*', the *class scheme of Erikson, Goldthorpe and Portocarero* or the '*International socio-economic index*' (see Hoffmeyer-Zlotnik and Warner (2012), page 191). Many data collections ask for occupation, including the *UK census*, which yielded almost 30 million verbal answers on employment in 2001 (Office for National Statistics, 2003), and the register-based *German 2011 census* with 3.6 million verbal answers (Loos *et al.*, 2013). The *American Community Survey* also contains questions on occupation, collecting approximately 2 million responses annually (Thompson *et al.*, 2014). Similar questions are common within many other surveys.

Unfortunately, the measurement of occupation is costly, time consuming and prone to errors. The standard approach is to ask one or two open-ended questions during the interview and sub-

Address for correspondence: Malte Schierholz, Mannheim Centre for European Social Research, University of Mannheim, 68131 Mannheim, Germany.
E-mail: Malte.Schierholz@mzes.uni-mannheim.de

© 2017 The Authors Journal of the Royal Statistical Society: Series A (Statistics in Society) 0964–1998/18/181379
Published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

sequently to code the verbal answers in a classification scheme with hundreds of categories and thousands of jobs. This coding task is non-trivial. Conrad *et al*. (2016) discussed various reasons why quality may be compromised. For example, many verbal responses are ambiguous and fit well into more than one category. Furthermore, some respondents have occupations for which no appropriate category exists. Because the target classification is fixed in advance, category modifications that could account for such difficulties are not feasible. Still, coders are typically required to decide on a single, most appropriate, job category. Several studies review the quality of coding occupational information under a variety of conditions (e.g. language, target classification, coding rules and procedures, and coder's experience) and report agreement rates for different people coding the same answers. Campanelli *et al*. (1997) employed three British expert coders to validate original codes from a number of non-experts, obtaining accuracies between 69% and 85%. Elias (1997) listed several British studies with intercoder reliabilities between 70% and 78%, with one exception from Slovenia reaching only 56%, and an international review by Mannetje and Kromhout (2003) mentioned reliabilities between 44% and 89%. Thus, the coding process entails a high degree of uncertainty that is usually ignored during data analysis. Higher quality in occupational data is clearly desirable—even more so if the new technique that we suggest here allows data collection at reduced costs.

Before going into detail, we briefly illustrate the technique proposed: consider a respondent who answers 'vice director [*sic*] and teacher' when asked about his job activities. On the basis of this *verbatim* answer and, if desired, further input from the interview, a computer algorithm searches for possible occupations and calculates associated probabilities at the time of the interview. The job titles that were found to be most likely are then suggested in closed-ended question format to the interviewer, who in turn asks the respondent to select the most appropriate occupation among these suggestions. The suggestions for the above-mentioned example are shown in Fig. 1. Since we cannot guarantee that the algorithm will always suggest an accurate job title, suggestions are complemented by a last answer option 'or do you work in a different occupation?'. If this option is chosen, further questions should be asked to gather additional details about the person's job; if not, coding is complete. In the example, the job title 'Teacher—elementary school' was selected, capturing a detail, the school type, that was not provided in the original verbal response.

With this new approach, we pursue three fundamental objectives to improve current shortcomings in the data collection process. First, we aim to reduce *coding errors* that arise from missing data or contradictory information provided by respondents. Respondents' verbal answers are sometimes ambiguous and difficult to code, in particular when survey questions are not aligned with the theoretical concepts that underlie occupational classification systems.

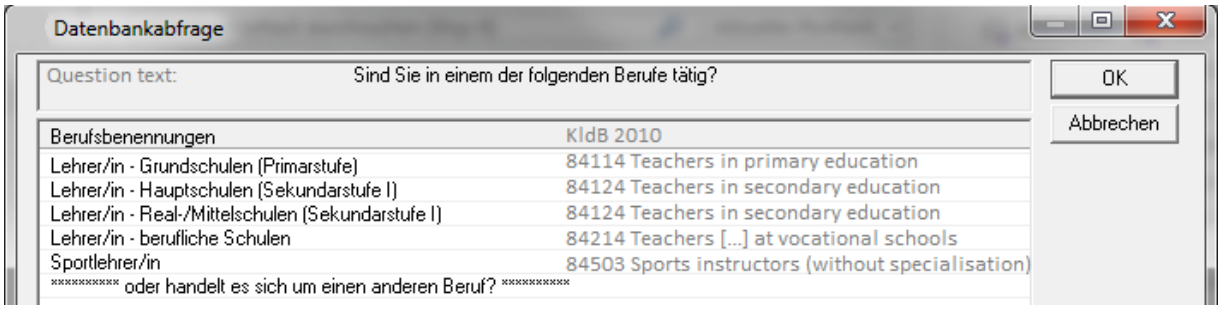


Fig. 1. Screenshot from the interview with the 'vice director and teacher': job titles in black fount were suggested to the interviewer; the text in grey fount was not shown during the interview and only added for this paper to illustrate underlying categories from the 2010 German classification of occupations; category titles are shown in abbreviated form (this example is discussed in Section 4.4.1)

Answer options often help to clarify the meaning of survey questions. Suggesting a limited number of answer options from the occupational classification based on initial verbal responses thus is expected to improve the measurement, while limiting respondent burden. Second, we seek to maximize the number of interview-coded answers to *minimize efforts for coding the residual cases* after the interview. Third, we aim to *save valuable interview time*, thereby reducing the respondent burden. The closed-ended question that is shown in Fig. 1 can often replace the additional open-ended question about occupation that is used in many questionnaires so that the total interview duration decreases. A final key advantage of the new instrument is the supervised learning algorithm that predicts possible job titles. The predictions are based on training data from past studies and can be *improved as more data become available*.

The new approach was tested in a computer-assisted telephone survey and codes occupations according to the *2010 German classification of occupations* (GCO) (Bundesagentur für Arbeit, 2011a, b), which is a detailed official classification and consists of 1286 well-documented categories subsuming 24000 job titles. Simultaneous coding according to the *2008 'International standard classification of occupations'* (ISCO) (International Labour Office, 2012) is supported in theory; in practice, the algorithm relies on a database that was not prepared for ISCO coding. Adaptions to Web surveys and other computer-assisted modes of data collection are possible, showing that many applications beyond telephone surveys and German occupations exist.

In this paper we describe the test of the new approach. To provide sufficient background and rationale, we review in Section 2 ('Background') literature on occupational coding and survey methodology more generally. In Section 3 ('Data and methods') we describe the data that were used for the test, the technical underpinnings of the new approach and the procedures to evaluate the new method. In Section 4 ('Results and evaluation') we describe the results from our test and discuss extensively the strengths and weaknesses of our approach, giving a special focus on possible modifications of the instrument to achieve even better results. Section 5 serves as a summary and compiles recommendations.

Owing to German privacy regulations, we cannot make our data public. Researchers who are interested in analysing the data on site at the Institute for Employment Research are invited to contact the first author. The computer code of our analysis is available from

`http://wileyonlinelibrary.com/journal/rss-datasets`

2. Background

Many processes are related to occupation coding during the interview. First, we characterize briefly the scientific coding tradition in Germany. Second, we argue in detail why our new instrument is expected to improve current practice. Third, we outline the main techniques that are similar to our approach.

2.1. Occupation coding in Germany

Geis and Hoffmeyer-Zlotnik (2000) have provided an overview of occupation coding in Germany and its difficulties. They argued for scientific standards in occupation coding that require coding to be carried out systematically and reliably. Consequently, Geis (2011) published a German coding manual for the international classifications from 1968 and 1988 (both outdated), which contains coding rules, conventions and a case collection to help to achieve high reliability between codings done by different coders. However, reliability does not guarantee validity and coding procedures that are optimized to achieve high reliability carry the danger of introducing systematic biases. For example, one of Geis's rules requires the coder to select the least skilled

of the plausible job categories. As a consequence, this coding procedure will underestimate the degree of professionalization in the workforce. Paulus and Matthes (2013) provided shorter instructions for coding into the 2010 German national classification. Dictionaries are available for both classifications to automate the coding process. If the *verbatim* answer from the interview matches an entry in the dictionary, then the corresponding code is assigned. With our technique of coding during the interview we do not follow this German coding tradition with its emphasis on coding rules and reliability; however, we employ professional coders who have a history of coding within the German context and we compare the new approach with current practice.

2.2. Motives for interview coding

We pursue three objectives with our new instrument:

(a) minimizing coding costs,
(b) increasing data quality and
(c) reducing the duration of the interview.

To ensure high data quality, researchers from various countries have discussed and compared various procedures for occupation coding (Biemer and Caspar, 1994; Campanelli *et al*., 1997; Bushnell, 1998; Biemer and Lyberg, 2003; Maaz *et al*., 2009; Svensson, 2012; Belloni *et al*., 2016). Although they focused on errors that arise from coding after the interview, all these researchers also observed that insufficient, low quality verbal answers from the interview are another possible source of errors. According to Hoffmann *et al*. (1995), page 13, 'the largest source of error lies in shortcomings of the verbatim raw material', as opposed to errors resulting from coding. Coming from a different perspective, the data editing literature (e.g. Granquist and Kovar (1997) and de Waal *et al*. (2011)) pointed out that it is overly expensive to correct all errors and inconsistencies in a processing step after data collection and advised researchers to improve measurement during the interview instead.

Why is it difficult to elicit suitable information from respondents concerning their occupations? Occupational classifications have organizing principles to cluster the variety of occupations into categories of 'similar' occupations. The classification then specifies each category in terms of illustrative occupations, typical tasks and boundaries to related categories. The classification thus dictates which job details are necessary for accurate coding. The current measurement process, however, is not aligned with theoretical concepts from the classification: the United Nations and International Labour Office (2010) recommended asking two open-ended questions, thereby hoping to collect sufficient details for coding according to the 2008 ISCO international classification. It is common practice in many surveys both to ask two questions and to give further instructions, requesting full details about the 'occupation' or 'job title' as well as the tasks and activities in the job (Tijdens, 2014a). We are sceptical about these standards for two reasons.

(a) No efforts are made in questionnaires to provide information about the classification structure and the boundaries between categories to make respondents' verbal answers more relatable to specific categories. Respondents can only guess which details about their jobs are requested and ambiguous answers are thus inevitable.
(b) The exact wording of an answer matters to coders, but it is unlikely that every respondent would use exactly the same words when answering the same question in a repetition study. In fact, the belief sampling model (Tourangeau *et al*., 2000), which is usually applied to attitude questions, suggests otherwise. According to this model, respondents base their answer on a small number of considerations drawn from a larger body of accessible

knowledge about his or her job. The considerations are retrieved at random depending on his or her current state of mind, with randomness leading to variability in answers. If the questions were more precise, the respondent could understand better what kind of knowledge is requested and randomness should decline.

Taking both points together, ambiguous and variable answers are thus a direct consequence of overly general questions that do not provide sufficient clues concerning what kind of information is required.

If respondents provide only vague occupational titles or if their task descriptions are not sufficiently detailed, many surveys demand from interviewers that they ask for a full job title or description (Hoffmann *et al.*, 1995; Tijdens, 2014a). The optimal kind and extent of probing are controversial. On the one hand, probes to open-ended questions in general can increase respondents' understanding of a question or encourage them to clarify their answers and to provide more complete information (e.g. Billiet and Loosveldt (1988), Conrad and Schober (2005) and Holland and Christian (2009)). On the other hand, when interviewers have some freedom concerning when and how to carry out probing, there is concern that respondents' answers are prone to interviewer effects, biasing all responses that are elicited by the same interviewer in a certain direction. To reduce this error, the standardized interviewing literature recommends avoiding probing whenever possible by using improved question wordings (e.g. Fowler and Mangione (1990), Mangione *et al.* (1992) and Schaeffer *et al.* (2010)).

Aside from this general discussion, the specific evidence for occupation suggests that probes are sometimes counterproductive. Since chances for conflicting information that is more difficult to code are higher when more information is available, coders disagree more often about the correct code when answers are longer (Conrad *et al.*, 2016). Results from Cantor and Esposito (1992), drawn from coders' comments on the interviewers' questioning strategy, point in a similar direction. Their coders criticized that interviewers probe too much in some cases, which results in adding ambiguity to the answer rather than in resolving conflicting information, whereas in other cases no probes are asked at all although coders actually desire more specific information about a certain aspect of the occupation. The bottom line is that generating useful probes is extremely difficult for interviewers. They would need to be well trained in probing and experienced with the occupational classification to elicit more useful answers, or—and this is the route that we follow in this paper—the questionnaire should provide interviewers with additional, classification-related questions that prescribe the exact wording for standardized probes.

Any effort to obtain more information about a respondent's job increases the total duration of the interview. Longer interviews have been associated with an increased respondent burden, lower rates of survey participation, more satisficing behaviour and reduced quality of data at the end of long questionnaires (e.g. Bradburn (1978), Holbrook *et al.* (2003), Galesic and Bosnjak (2009) and Roberts *et al.* (2010)). Asking two open-ended questions and possible additional probes cost a considerable amount of time. Although these efforts are necessary to obtain precise information from some respondents, valuable interview time is wasted for others who have given a precise answer already to the first question. To reduce the length of the interview, our proposed instrument is adaptive: we suggest asking only one open-ended question, whereupon the interview software evaluates the answer and decides which question is asked next.

2.3. Related techniques and instruments

Our new instrument combines two specific developments that have been implemented separately from each other.

(a) Post-interview coding of occupations is increasingly carried out by using machine learning algorithms and training data rather than simple matches of answers to prespecified words from a dictionary.
(b) Some researchers have described mechanisms for coding during the interview that enable the respondent to specify his or her response more precisely, if required.

In what follows we provide a brief overview of both approaches.

Several researchers have proposed computer systems to automate the post-survey coding process (e.g. Speizer and Buckley (1998) for a review, Measure (2014) and Gweon *et al*. (2017)). Software for computer-assisted coding suggests possible categories to the coder to make human work more efficient. Other algorithms known as 'automated coding' independently assign categories without human supervision. More difficult residual cases are left for professional manual coding to keep the error level from automated coding below some prespecified threshold. Conceptually, both systems are generally based on coding rules and large databases that contain codes for recurrent job titles (e.g. the prominent 'Computer-assisted structural coding tool' program CASCOT that was described by Elias *et al*. (2014) implements all the above mentioned). However, coding rules are created by hand, and complex coding systems require quality checks before they can be used for production. Creecy *et al*. (1992) have challenged such hand-crafted rule systems, which are expensive to develop, with their own algorithm that learns from training data, consisting of 132 247 observations. By doing so, previously coded verbal answers are used to learn coding rules automatically. Their software outperforms another coding system that was based on hand-crafted rules and used for production in the 1990 US census. More recent proposals learn from even larger training data with more than 1.5 million observations each (Jung *et al*., 2008; Thompson *et al*., 2014; Javed *et al*., 2015). The novel algorithm that we use in this study combines the learning from training data and the usage of hand-crafted databases.

Different strands of research try to code occupations directly during the interview. Hoffmeyer-Zlotnik *et al*. (2006) asked for occupation with a sequence of three filter questions: the first asks for broad occupational groups, the second specifies the occupation further and the final third question refers to specific 1988 ISCO categories. Tijdens (2014b, 2015) also avoided verbal answers with a similar 'search tree' for Web surveys. A different strategy is employed by the job portal offered on line at `http://jobboerse.arbeitsagentur.de/` by the German Federal Employment Agency, which uses textual input to autosuggest possible job titles from a database. These titles are linked to job categories from the 2010 GCO. Hacking *et al*. (2006) and Svensson (2012) also mentioned occupation coding during the interview, but their descriptions lack details.

Coding during the interview is not limited to occupation coding but is applicable to any question with a large number of answer options. Bobbitt and Carroll (1993), for example, tested a system for coding 'major field of study'. In a telephone survey, they implemented a fuzzy text search algorithm that suggests possible codes, allowing interviewers to verify codes directly with the respondents. Couper and Zhang (2016) asked for prescription drugs in Web surveys and compared three question formats: a text box for later coding, a drop box menu containing a list of 4768 drug names in alphabetical order and an autosuggested list that narrows down the large number of drugs on the basis of textual input and simple text matching. They concluded that each format has its own strengths; nevertheless, a careful design of the instrument is worthwhile.

3. Data and methods

The new tool was tested in the survey '*Selectivity effects in address handling*' that was commissioned by the German Institute for Employment Research and conducted by Kantar Public. In

October and November 2014, Kantar Public conducted in total 1208 valid computer-assisted telephone interviews; 1064 verbal answers for occupation were collected. The questionnaire covered, among others, several topics related to the respondents' current occupation and work history, the use of social media for private and professional purposes, and volunteering activities.

3.1. Sampling and data collection

A random sample of 17 001 people—some of them with multiple addresses; others without phone numbers, which needed to be identified—was drawn from a German federal database that is used in social security administration (vom Berge *et al.*, 2013). Since the primary purpose of the survey was to explore possible selectivity effects, a random subsample of 10 000 people was asked for consent to address transfer and the consenters' addresses as well as control group addresses were transferred from the German Institute for Employment Research to the survey operator. Details of the experiment are described in Sakshaug *et al.* (2016). Before the fieldwork, a notification letter was sent to 7183 available addresses.

The sampling frame covers employees, unemployed people, jobseekers, recipients of unemployment benefit II and participants in active labour market programmes. It thus accounts for a large share of the German working population. However, people who never paid contributions for social security insurance and have never received benefits from the German Federal Employment Agency are not included. This implies a specifically strong undercoverage of civil servants and self-employed people.

All 67 interviewers, the local fieldwork managers and the supervisors were trained by the central project management team of Kantar Public. The new tool was an essential part of this training.

3.2. Integration into the questionnaire

The coding process starts by asking one open-ended question about the occupation ('Please tell me your occupational activity'), followed by a sequence of approximately eight additional job-related questions that are intended to collect as many details as possible about the respondent's job. These further questions are used

(a) for manual coding to evaluate the new instrument and
(b) as covariates in our model to improve predictions.

The answer to the first open-ended question would suffice to suggest possible job titles and that the additional questions could be skipped if they were not needed to evaluate the instrument. The exact German wording and its corresponding English translation are available in the on-line appendix.

Immediately following these questions, interviewers are prompted to read the following text:

> 'We now try to classify your occupation. A database query is made for this purpose. This can take a short moment.'

The interviewer then starts the query and the algorithm computes at most five job titles to be suggested to the respondent. After a few seconds, the question generated is shown in a pop-up window (Fig. 1). The interviewer then asks into which of the following categories (job titles) the job falls, or whether the answer option 'different occupation' would be most appropriate. A random subset of less than 10% received an additional answer option 'similar occupation'. Because of the small sample size we discuss this portion of the study only in the on-line appendix. All interviewers received a quick debriefing question on the flow of the interaction

after the coding module. Details of those debriefings are also available in the on-line appendix, for brevity.

Every suggested job title (shown on the left-hand side in Fig. 1) corresponds to one category from the *Dokumentationskennziffer*, which is an internal job classification that is used by the German Federal Employment Agency in its daily operations (see Paulus and Matthes (2013) for details). This classification subdivides the 1286 categories from the 2010 GCO in 11 194 *Dokumentationskennziffer* categories. Conversely, this means that every job title is linked to exactly one category in the 2010 GCO. Thus, when a job title is selected during the interview, the *Dokumentationskennziffer* code is saved and a 2010 GCO code is automatically assigned as well. For illustration, we include these associated GCO categories in the grey fount on the right-hand side of Fig. 1. All evaluations provided below will be done on the scale of the 2010 GCO, as this is the official and well-documented German national classification. The *Dokumentationskennziffer* itself is used only as an auxiliary classification that provides the job titles for our instrument, links these job titles to the 2010 GCO and makes available a large database of search words.

Many researchers do not use the national 2010 GCO but work with the 2008 ISCO instead. As this study explores technical possibilities, we test our technology only on the 2010 GCO. However, it is worth noting that many—but not all—*Dokumentationskennziffer* categories are linked to specific ISCO categories, making it conceptually feasible to code in the 2008 ISCO and 2010 GCO at the same time during the interview. Since the ISCO with its 436 categories corresponds to only about a third the size of the GCO, we also expect improved quality evaluations if the analysis below is carried out for the 2008 ISCO.

3.3. Prediction algorithm

Possible job categories are predicted with a supervised learning algorithm that learns from training data, i.e. from verbal answers whose classification codes are already known from manual coding. Our training data come from the survey '*Working and learning in a changing world*' (Antoni *et al*. (2010); Drasch *et al*. (2012) documented the coding process). This survey interviewed 9 227 people about their employment biographies, i.e. all the jobs that they have held during their lifetime, yielding a total of 32 887 job records. Compared with other supervised learning algorithms for occupation coding, this number is exceptionally small. Because of the tiny size of our training data, 433 out of 1 286 job categories from the 2010 GCO are not covered, implying that these categories would never be suggested if the predictions were based only on these training data.

In principle, training data should be as large as possible to account for a high variety of possible verbal inputs, including misspellings, and it should also cover all contingencies how a specific input text can be coded in different categories. Such large training data were not available to us; as a consequence, many respondents provide verbal answers that cannot be matched to the training data. To obtain predictions for these respondents still, we use two databases of job titles in addition to our training data and search for possible job categories in all three sources. Resorting to additional databases should mitigate our problem of small training data, but more training observations could certainly further improve our results.

Schierholz (2014) developed the underlying prediction algorithm and evaluated its performance. To integrate Schierholz's (2014) algorithm into our new coding approach, the target classification was changed (*Dokumentationskennziffer* instead of the 2010 GCO) and some scores (see below) were streamlined. The algorithm works in three steps; the exact calculations in each step are described below. The rest of this paper can easily be understood without these technical details.

(a) Calculate scores $\theta_{lj}^{(m)} = f_m(x_l, c_j)$ for a given respondent l and all *Dokumentationskennziffer* job categories c_j, $j = 1, \ldots, 11194$. We use 26 predefined matching methods f_m, $m = 1, \ldots, 26$, that link the respondent's answers x_l to databases and training data.
(b) Predict correctness probabilities for all categories using a function $\hat{g}: (\theta_{lj}^{(1)}, \ldots, \theta_{lj}^{(26)}) \mapsto \hat{p}(c_j|l)$. This function was estimated beforehand from training data.
(c) Suggest the five most probable job categories (under some restrictions) to the respondent.

3.3.1. Calculate scores

To be useful, the scores $\theta_{lj}^{(m)}$ should be predictive of the true probability $p(c_j|l)$ that category c_j is correct for respondent l. Any supervised learning technique might be used to estimate functions f_m, the more the better, as long as the number of scores is far below the number of observations that are used. For simplicity, we use only a small set of 26 matching methods to define the functions f_m. Several scores are built on each other and, because predictions improve when more scores are used, we included all of them. The matching methods are summarized in Table 1. For example, our first matching method, f_1, selects all training data observations in which the full texts from respondent l and from the training data are identical and calculates the frequencies of each category c_j in this subset. By construction, the most frequent code that is found with this matching method is likely to have the highest probability of being correct.

To develop additional matching methods, we vary four dimensions as shown in Table 1.

(a) The input is either the respondent's answer to one of the closed questions, the *full text* (i.e. the first verbal answer after removing some special characters and replacing letters with their upper-case equivalents), or a *phrase* (i.e. the subsequence of words from the full text that has the highest frequency of appearance in a single category in the training data). In our example, the full text is 'VICE DIRECTOR AND TEACHER' and the derived phrase is 'TEACHER'.
(b) For comparison, our input texts must be either *identical* to or a *substring of* another text. The naive Bayes statistic is based on a *word-by-word* comparison.
(c) The input is compared with *training data*, to an *alphabetic dictionary* of job titles (the '*Berufs- und Tätigkeitsverzeichnis*' that is part of the 2010 GCO; Bundesagentur für Arbeit (2011a)), or to an index of *search words* (created by the German Federal Employment Agency for operative purposes; Bundesagentur für Arbeit (2013)).
(d) The statistic dimension prescribes how to calculate category-specific scores θ_{lj} from large numbers of matching entries.

    The most basic statistic is the *code frequency*, i.e. the absolute frequency $\#\{\text{answer}, c_j\}$ of each code c_j that appears in the selected subset.

    '*Posterior expectation*' is the posterior expectation for some category c_j and '*posterior probability*' is the posterior probability that parameter$_j > 0.05$. The underlying Bayesian model consists of a subset-specific multinomial likelihood to model the observed code frequencies $\#\{\text{answer}, c_j\}$ and a Dirichlet prior that depends on relative code frequencies in the complete training data, Dirichlet$(0.5 \cdot \#\{c_1\}/N, \ldots, 0.5 \cdot \#\{c_J\}/N)$. The posterior is thus a Dirichlet$(\#\{\text{answer}, c_1\} + 0.5 \cdot \#\{c_1\}/N, \ldots, \#\{\text{answer}, c_J\} + 0.5 \cdot \#\{c_J\}/N)$ having posterior expectation $\theta_{lj}^{(3)} = \omega \cdot \#\{\text{answer}, c_j\}/\#\{\text{answer}\} + (1-\omega) \cdot \#\{c_j\}/N$, a weighted average with weights $\omega = \#\{\text{answer}\}/(\#\{\text{answer}\} + 0.5)$ that shrinks the relative code frequencies in the selected subset towards the prior expectations.

    For closed questions, we calculate the *proportions* $\hat{p}(\text{input}|c_j) = \#\{\text{answer}, c_j\}/\#\{c_j\}$, which is the subset-specific code frequency divided by the absolute frequency of each code in the complete training data. On the basis of the *naive Bayes* assumption, we esti-

Table 1. Overview of matching methods

m	Input	Comparison	Compared with	Statistic	Mean†
Open-ended questions					
1	Full text	Identical to	Training data	Code frequency	0.00261
2	Full text	Substring of	Training data	Code frequency	0.00439
3	Full text	Identical to	Training data	Posterior expectation	0.00009
4	Full text	Identical to	Training data	Posterior probability	0.00019
5	Full text	Word by word	Training data	Naive Bayes	0.00008
6	Full text	Identical to	Search words	Code frequency	0.00019
7	Full text	Substring of	Search words	Code frequency	0.00400
8	Full text	Substring of	Alphabetic dictionary	Code frequency	0.01264
9	Phrase	Identical to	Training data	Code frequency	0.00261
10	Phrase	Identical to	Training data	Posterior expectation	0.00009
11	Phrase	Identical to	Training data	Posterior probability	0.00019
12	Phrase	Word by word	Training data	Naive Bayes	0.00008
13	Phrase	Identical to	Search words	Code frequency	0.00049
14	Phrase	Substring of	Search words	Code frequency	0.14391
15	Phrase	Substring of	Alphabetic dictionary	Code frequency	0.02347
Closed questions					
16	Occupational status	Identical to	Training data	Proportion	0.11747
17	Differentiated occupational status	Identical to	Training data	Proportion	0.05755
18	Number of staffers	Identical to	Training data	Proportion	0.20734
19	Superviser	Identical to	Training data	Proportion	0.10926
20	Number of employees supervised	Identical to	Training data	Proportion	0.12083
21	Education required	Identical to	Training data	Proportion	0.04996
22	Industry	Identical to	Training data	Proportion	0.15182
23	Company size	Identical to	Training data	Proportion	0.05623
Other					
24	Multiply scores 5 and 16–23, and relative code frequency in training data				0.03127
25	Are full text and phrase identical?			Yes or no	0.41820
26	Number of suggested categories from all matching methods			Count	184.0031

†Mean $= (1/N)(1/J)\Sigma_{l=1}^{N}\Sigma_{j=1}^{J}\theta_{lj}^{(m)}$ is the mean score over all categories and respondents for the matching method m. It is based on a subset of 958 respondents where the algorithm finds possible categories.

mate probabilities to observe the observed text under any given job category by using the formula

$$\hat{p}(\text{input text}_l|c_j) \propto \prod_{v=1}^{V}\{0.95\hat{p}(T_v|c_j)+0.05\hat{p}(T_v)\}.$$

This is a product over all words T_v that appear in the input text. $\hat{p}(T_v|c_j)$ and $\hat{p}(T_v)$ are both relative frequencies as calculated from the training data. $\hat{p}(\text{input text}_l|c_j)$ is standardized to sum to 1. The proportion statistic and the naive Bayes statistic were originally developed as a by-product to estimate $p(c_j|l)$ as in $\theta_{lj}^{(24)} = \hat{p}(c_j|l) := \hat{p}(c_j)\,\hat{p}(l|c_j)/\hat{p}(l)$ with

$$\hat{p}(l|c_j) := \hat{p}(\text{input text}_l|c_j)\prod_{m=16}^{23}\theta_{lj}^{(m)}.$$

The final score $\theta_{lj}^{(26)}$ is a person-specific (equal to category-independent) variable that

counts how many different categories were found that have code frequency greater than 0 in one of the matching methods. If this number is high, many different categories appear possible, and the final probability of picking the correct category might be lower.

Schierholz (2014) explained the scores and the underlying reasoning in more detail.

3.3.2. Predict correctness probabilities

We have now 26 different scores that are expected to correlate with the true probability $p(c_j|l)$ and that can be interpreted, in the case of $\theta_{lj}^{(3)}$, $\theta_{lj}^{(10)}$ and $\theta_{lj}^{(24)}$, as an estimate for this probability. We write all scores from a single person in a data frame as depicted in Table 2 and concatenate the data frames from all training observations to form a single data frame. How can we *combine* the different scores to form a single more accurate prediction $\hat{p}(c_j|l)$? We need to estimate a function $\hat{g}: (\theta_{lj}^{(1)}, \ldots, \theta_{lj}^{(26)}) \mapsto \hat{p}(c_j|l)$. Although the outcome is multinomial, we regard it as a binary problem and aim to predict the probabilities $p\,(c_j \text{ correct}|l)$ instead. A similar problem of combining predictions ('stacking') was studied by Stone (1974), LeBlanc and Tibshirani (1996) and Breiman (1996) who restricted themselves to linear combinations g of the predictors $\theta_{lj}^{(m)}$. Stone (1974), LeBlanc and Tibshirani (1996) and Breiman (1996) assumed that predictors $\theta_{lj}^{(m)}$ are in themselves estimates of the outcome variable that may be obtained from any supervised learning model; however, we see no reason why their method should be limited to linear combinations of other models' predictions. The key problem is that estimated parameters would be biased if the same training observations were used twice: a first time to estimate the functions f_m and a second time to estimate g. To avoid double usage, we apply leave-one-out cross-validation, i.e. the first-stage predictions $\theta_{lj}^{(m)(-l)} = f_m$ are not based on the observed outcome of respondent l. The observed outcome from the training data is used only afterwards for estimation of g, together with the leave-one-out estimates $\theta_{lj}^{(m)(-l)}$. To estimate the function g, we train gradient-boosted trees as implemented by Hothorn *et al.* (2010), which is a more flexible tool than linear regression that allows for non-linearities and high order interactions. In doing so, a sequence of decision trees is trained iteratively, each iteration focusing on examples that the previous iteration got wrong. The final prediction is a sum over the different trees.

Training the gradient boosting model on a data set with $32\,887 \times 11\,194 \approx 368$ million rows is computer intensive and time consuming. Furthermore, computers need to have a very large random-access memory to load a boosted model if the training data consist of many observations. This is a shortcoming of our approach and three workarounds are needed to make this manageable.

(a) We keep only the rows in the data frame in which at least one score obtained via the verbal answer indicates that this category could be correct. If the text does not indicate

Table 2. Illustrative data frame for person l with correct job category $c_j = 01104101$

c_j	c_j *correct*	$Score_{lj}^{(1)}$	$Score_{lj}^{(2)}$	$\cdots$
$c_1 = 01104100$	False	$\theta_{l,1}^{(1)}$	$\theta_{l,1}^{(2)}$	$\cdots$
$c_2 = 01104101$	True	$\theta_{l,2}^{(1)}$	$\theta_{l,2}^{(2)}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_{11194} = 99998115$	False	$\theta_{l,11194}^{(1)}$	$\theta_{l,11194}^{(2)}$	$\cdots$

the possibility of correctness (as operationalized in $\theta_{lj}^{(26)}$), the row is removed. The resulting data set has 461 816 rows, many of them still highly unlikely to be correct.

(b) We randomly split the data by rows into 10 disjoint sets and estimate five separate boosting models, leaving the other five sets aside because of performance restrictions. To predict a new code for a given response l, we can then
   (i) predict the scores $\theta_{lj}^{(m)}$ by using the complete training data,
   (ii) predict probability vectors for 'c_j correct' with each of the five boosting models and
   (iii) average over the five predictions.

(c) To speed up model training, we use only 45 iterations, which are fewer than recommended. To reach close-to-optimal solutions after 45 iterations, we increase the step size. Tuning parameters (maximum tree size, 9; step size, 0.5) are chosen according to extensive exploratory bootstrap-type cross-validation.

Although we are confident that these design decisions do not negatively affect the algorithm much, more efficient solutions are clearly desirable.

3.3.3. Suggest job categories

At this point, the algorithm enables us to calculate a number of possible *Dokumentationskennziffer* categories and corresponding estimated correctness probabilities for new data. For most responses, dozens of categories are found—more than would be convenient to ask in a survey. We therefore restrict the maximum number of suggested categories to five, which is a suitable number for unordered response options in telephone surveys (Schnell (2012), page 94). It is desired to suggest job titles that cover a range of GCO categories. For this we select (up to) five *Dokumentationskennziffer* categories with the highest correctness probabilities under the condition that not more than two of the selected *Dokumentationskennziffer* categories may belong to the same GCO category. Only if we cannot fill the five available spaces according to this rule are additional *Dokumentationskennziffer* categories from the same GCO category added according to their correctness probability (highest first). Finally, the categories suggested are ordered by GCO code numbers and the answer option for 'other occupation' is added.

3.4. Quality analysis

We evaluate the quality of our new approach by comparing the machine-learning-assisted within-interview coded answers with professional post-survey coding, as is traditionally done in Germany. The analysis includes two steps:

(a) manual coding and
(b) double-checking of answers for which at least one code might be wrong.

For the first step, two professional coders were asked to code the verbal answers independently from each other and without knowledge about the interview-assigned codes. Both are experienced coders and offer this service on a paid basis. Their respective coding documentations show that both coders have different coding procedures and, for ambiguous answers, different decision rules are used. In addition, one of the coders provides a special indicator describing which verbal answers have multiple possible codes. To guarantee the anonymity of the coders, the differences cannot be described in more detail.

In a second step, all observations for which there was disagreement between any two of the three codes (from interview coding and both professional codings) was subject to additional examination. This also includes observations for which one of the professional coders expressed his uncertainty. Observations that were not interview coded are excluded. Two

student assistants checked the correctness of the different codes for each of the 368 observations. Both assistants worked independently of each other. They were provided with the same source material as the two coders (verbal answers and additional answers from the interview; see the on-line appendix) and with the codes from the professional coding and interview coding. Their task was to categorize each coding decision in one of the following three categories.

(a) Acceptable: there is a good argument for the coding decision to be considered correct. This is independent of the fact that other plausible arguments may lead to different coding decisions that may be considered correct as well.
(b) Wrong: it is obvious that the coding decision is erroneous and other codes are clearly more appropriate.
(c) Uncertain: this is the residual category to be assigned when a code is not obviously erroneous and at the same time there is no good argument for it to be correct. Three reasons are most common why a category is classified as uncertain:
    (i) the job title that is selected during the interview appears correct at a first glance, but a different category definition from the GCO, volume 2, describes the job activities more precisely;
    (ii) the interview-coded job category requires a level of skill that is contradictory to the answers from the interview (i.e. to the questions on the vocational training that is usually required or the differentiated occupational status);
    (iii) the answers from the interview suggest a different thematic focus, but at the same time the code is not entirely wrong.

The complete instructions including examples, which were given to the student assistants, are provided in the on-line appendix.

3.5. Interviewer behaviour

To evaluate the new approach further we coded the interviewer behaviour itself. This allows us to analyse the extent how often interviewers correctly applied standardized interviewing techniques that are prescribed for the new question on occupation (see Ongena and Dijkstra (2016) for an overview on behaviour coding). At the beginning of the interview, all respondents were asked for permission to record the conversation, obtaining an 87.5% rate of consent. Of the consenters who provided answers to the occupation questions and whose audio recordings did not contain personal identifiers, a total of 211 were randomly selected for behaviour coding. An independent coder from Kantar Public recorded whether the interviewer read the question text and each answer option as instructed, or if he or she diverged from the text or omitted suggested job categories. In addition, the coder noted what the respondent said as a first reaction. The complete coding instructions are provided in the on-line appendix. Two audio files were excluded, because the recordings only start after the occupational questions have been asked. In the course of the analysis, the first author listened to several recordings and felt reassured that the coder delivered high quality. Various interpretations of the interviewer–respondent interaction in the result section were also obtained from listening to the recordings with careful attention to the specified aspects.

4. Results and evaluation

This section starts with three key criteria to assess the tested system: productivity, interview du-

ration and quality, followed by two examples to explain some particularities of our instrument. We then report from the detailed analysis of the audio recordings to understand how interviewers and respondents interact, and discuss the strengths and weaknesses of the prediction algorithm. We close this section with an examination of errors resulting from the classification material. Throughout all descriptions, we highlight shortcomings in the tested system and mention possible modifications to obtain even better results in a future version of the instrument.

4.1. Productivity analysis

Table 3 provides an overview of the productivity of our system. Among the 1064 people who responded to the survey questions about occupation, the algorithm found possible categories for 90.0%, leaving only 10.0% for whom the algorithm did not suggest a single job category. This happens if the algorithm cannot relate the text that is entered by the interviewer to any previous input from the training data or from the job title databases. This is often due to misspelled job titles and could be reduced by using spell checking algorithms.

72.4% of the respondents selected a job title from the list generated. This number is highly important, because it shows that nearly three-quarters of the coding task could be carried out during the interview, which considerably reduces the work for post-interview coding.

13.6% of the respondents did not find an appropriate job title among those suggested by the algorithm and declared that they have a different occupation instead. This was expected, as the algorithm is optimized to suggest appropriate job titles, but it is impossible to guarantee that it will always propose correct job categories. In fact, the matching methods in our algorithm often find dozens or even hundreds of possible job titles. For usability, we restrict the maximal number of suggested job titles to five. When filtering out the five best-suited job titles, frequently relevant categories are missed, whereas irrelevant categories are suggested. The quality of the suggestions depends on the availability of training data and details in the algorithm. With additional training data and improved algorithms for prediction, we thus expect to decrease the proportion of answers for which no code is assigned and to increase the productivity of our system.

For respondents who answer that no job title is appropriate and who indicate that they have an 'other occupation' (applicable for 13.6%, as shown above), two additional lists are generated automatically and suggested to them. The first contains titles from the more general occupational subgroups (four-digit GCO). The respondent can then select a subgroup or terminate the procedure by saying that no subgroup is appropriate. When selecting a subgroup, *Dokumentationskennziffer* job titles only from the chosen subgroup are suggested to the respondent. This demanding follow-up process was implemented because the algorithm usually finds dozens of possible job titles and, although it is desired that respondents can navigate to the best fitting job title during the interview, it is impossible to suggest all of them within a single question. Contrary to our expectations, 79% of the eligible respondents did not select an occupation during

Table 3. Productivity of the coding system

Number of respondents who give a job description	1064	100.0%	
Algorithm provides no job suggestion	106	10.0%	
Algorithm finds possible categories: thereof,	958	90.0%	
... Respondent chooses a job title	770		72.4%
... Respondent chooses 'other occupation'	145		13.6%
... Item non-response	3		0.3%
... Other experimental conditions	40		3.8%

this process. In case they did, this interview-coded occupation is not in agreement with manual coding in 77% of cases. Fig. A1 and Table A1 in the on-line appendix provide additional details. We conclude that these follow-up questions yield unsatisfactory results and should be dropped. If respondents select 'other occupation', responses should be referred to manual coding. They are thus excluded from the subsequent analysis.

Table 3 also shows that three of the 1064 people did not respond to the new instrument. The remaining 3.8% were due to the following experimental artefact: if the algorithm finds only a single job title or more than 250 possible job titles, job titles were not suggested within the regular closed question on occupation, but different question wordings were tested instead. Results are shown in Tables A2 and A3 in the on-line appendix. Both experimental conditions were not worthwhile for our research because the number of observations falls below our expectations. Standard procedures, as if 2–250 categories were suggested, would probably have worked equally well.

4.2. Interview duration

If coding during the interview is to replace the present procedure that asks two or three open-ended questions about a respondent's job, it is of high relevance that the duration of the interview does not increase. Longer interviews are more expensive and tiresome for the respondent. For respondents who select an occupation during the interview, our additional question takes 37 s on average. As further open-ended questions can be avoided (a standard question in German surveys is 'Please describe this occupational activity precisely', which takes 44 s on average), the total interview duration is reduced for these respondents. Conversely, for respondents who do not select an occupation but instead choose the category 'other occupation', additional open-ended answers are still necessary for coding after the interview, increasing the total duration of the interview. The objective must therefore be to minimize the number of respondents who choose 'other occupation'.

4.3. Quality analysis

Nearly three-quarters of the respondents select a job title during the interview. Although this is auspicious, the quality of the interview-coded categories is even more relevant. Two specific aspects of quality are analysed: the agreement between and the evaluation of the different coding procedures. Both measures enable conclusions about the quality.

Table 4 (the first three rows) provides the intercoder reliabilities for the professional coders (coder 1 and coder 2) and their respective rates of agreement when compared with the codes from the interview. Agreement between five-digit categories from the 2010 GCO is highest with 66.23% when comparing coder 2 with interview coding. All agreement rates improve for broader classifications with fewer digits, but coder 2 and interview coding again have the highest rates of agreement. Agreement between both professional coders is lowest with almost 39% of disagreement, leaving room for improvement.

An explanation for the lower agreement between professional coders might be that it is easy to find a correct code for some job descriptions, whereas it is not for others (e.g. Cantor and Esposito (1992) and Conrad *et al.* (2016)). For example, if respondents find their previous verbal answer in the list of suggested job titles, they often select this job title, acting similarly to what professional coders do. We assume that people with more complex job descriptions are more hesitant to choose one of the suggested job titles, as the titles are less likely to be appropriate. Consequently, simpler job descriptions are more often interview coded. In contrast, professional coders are required to code all occupations, regardless of the selection process during the interview, including also the more complex job descriptions, on which professional

Table 4. Agreement rates between the two professional coders (coder 1 and coder 2) and interview coding (interview)†

Agreement between	*Number of codes‡*	*First . . . digits are in agreement (%)*				
		1	*2*	*3*	*4*	*5*
	(All available)					
Coder 1 and coder 2	1039	87.20	79.40	74.98	67.56	61.11
Coder 1 and interview	754	87.67	80.37	75.46	67.77	61.80
Coder 2 and interview	770	89.09	82.21	77.53	71.56	66.23
	(Subset)					
Coder 1 and coder 2	754	89.26	82.76	79.18	72.68	65.78
Coder 1 and interview	754	87.67	80.37	75.46	67.77	61.80
Coder 2 and interview	754	88.86	81.83	77.19	71.35	66.05

†The 2010 GCO consists of five-digit codes; aggregates for broader classifications with fewer digits are shown for convenience.
‡'Number of codes' shows how many codes are available for each comparison. Coder 1 provides codes for 1041 out of 1064 occupations. For three occupations, the 'qualification is unknown', one occupation is a worker without further specification, and for 19 occupations 'multiple codes [are considered] possible'. Coder 2 provides codes for 1062 out of 1064 occupations, whereas the other two occupations are 'not codable'. Interview coding provides codes for 770 occupations. The quotes stem from the respective coding documentation.

coders presumably agree less often. This argument is supported by the fact that agreement between coder 1 and coder 2 increases from 61.11% to 65.78% when this number is calculated only for the subset of the 754 occupations that were also coded during the interview.

To allow for more meaningful comparisons of the different coding procedures, the bottom part of Table 4 is based on a subset of respondents for whom we have codes that are available from all three procedures. Looking at agreement rates of 1–4 digits, agreement is highest between coder 1 and coder 2. For five-digit codes, this pattern changes with coder 2 and interview achieving the highest rate of agreement. Hypothesizing that interview-coded answers are more often miscoded, they would agree less often with accurate codes from professional coding. This could explain that agreement between coder 1 and coder 2 is highest for 1–4 digits, suggesting lower quality of data of interview coding. However, the differences between the coding procedures are small and even non-existent for five-digit codes. Furthermore, higher miscoding rates of interview coding are only one possible explanation for the observed pattern. Another explanation assumes that professional coders lack relevant information because it was not provided during the interview. In this case, they both would assign wrong codes, leading to agreement between professional coders but to disagreement with the respondent's own choice, who knows better. Taken together, the evidence provided so far is inconclusive about the validity of each coding procedure and suggests that differences between the coding procedures are only small, if existent at all. The second step of our analysis will elucidate this further.

For 402 out of 754 respondents (53.32%), the professional coders both agree with the respondents' own choices. For these cases we can be highly certain that interview coding yields a code with quality that is comparable with manual coding. In what follows, we assume that these codes are 'acceptable'. More problematic are the $770 - 402 = 368$ cases in which at least one human coder deviates from the code that was obtained via interview coding.

Table 5 shows the results from the students' evaluation of the quality of coding. For the

Table 5. Student assistants' evaluation of the coding process quality†

		Student 2			
		Acceptable	*Uncertain*	*Wrong*	Σ
Correctness for coder 1					
Student 1	Acceptable	(402 +) 232	6	19	257
	Uncertain	45	3	20	68
	Wrong	19	3	21	43
	Σ	296	12	60	368
Correctness for coder 2					
Student 1	Acceptable	(402 +) 194	8	23	225
	Uncertain	35	7	20	62
	Wrong	27	12	42	81
	Σ	256	27	85	368
Correctness for interview coding					
Student 1	Acceptable	(402 +) 189	13	13	215
	Uncertain	54	12	16	82
	Wrong	33	15	23	71
	Σ	276	40	52	368

†Cases are analysed only if at least one human coder deviates from the code that was obtained via interview coding.

majority of the 368 problematic codes, both student assistants agreed that the codes are acceptable. Viewed as a proportion of all 770 interview-coded answers, coder 1, coder 2 and interview coding are acceptable in 82.3%, 77.4% and 76.8% of the cases. Coder 1 is rated best: 232 (plus 402 acceptable by assumption) of his assignments are considered acceptable, which is significantly more than the 194 (plus 402) acceptable codes from professional coder 2 (H_0: equal proportions of acceptable codes from coder 1 and coder 2; $\chi^2 = 5.53$; $p = 0.019$) and more than from interview coding (189 + 402). Our aspired goal to increase data quality with interview coding was not achieved. Coder 1 also produced the lowest number of wrong codes (21) among the three different coding procedures. All other codes from coder 1 are somewhere in between acceptable and wrong, with little agreement between the student assistants. The frequent disagreement between students cautions not to overinterpret these results.

To conclude, we find it difficult to determine a clear winner concerning optimal quality of data. Both the analysis of agreement rates and the students' evaluation provide evidence that differences in quality of data among coding procedures are quite small and may be negligible for practical purposes. Furthermore, there is frequent disagreement on the correct code, reflecting the complexity of occupation coding (see part C in the on-line appendix for coding examples). The students' evaluation shows that it is possible to identify obvious errors in all coding procedures, but these errors account for only a small proportion of the overall disagreement. In the next section, we illustrate why disagreement may occur.

4.4. Illustrative examples

Occupation coding in general and interview coding in particular have several particularities that are worth discussing in detail. The following two examples help us to understand weaknesses and to consider ways of improvement. The example of the 'vice director and teacher', which was

introduced in Fig. 1, was chosen because it offers various insights. The second example ('truck salesman') was chosen because it is symptomatic for a type of error that we observed more than once in interview coding.

4.4.1. Vice director and teacher

In interview coding, the category 84114 'Teacher—elementary school' is selected, which is plausible given the last word from the text written down by the interviewer. Audiorecording confirms that this person is a vice principal and teacher at an elementary school. A professional coder would not have known that this person works at an elementary school because the interviewer failed to write down the complete information, making this answer a candidate for error.

Two professional coders were asked to code the textual answer. Both decided on the category 84194 'Managers in school of general education'. This category is the most appropriate from a post-interview coding perspective, for three reasons.

(a) Additional questions from the interview show that this person supervises 14 employees, indicating that managerial responsibilities may dominate his professional tasks as a teacher. This would favour the category 84194, because the main focus of activities performed in the job is, according to the 2010 GCO, the criterion to decide for the best-suited category.
(b) The alphabetic dictionary that is part of the 2010 GCO assigns 'vice principal' to code 84194. Note, however, that our algorithm does not recognize synonyms.
(c) The respondent answered 'vice director' before 'teacher'. Coding rules often determine that the first job title is coded if multiple titles are provided in the *verbatim* response and the other titles do not specify the first title.

The school manager category 84194, which both professional coders prefer, is missing in the list of suggested job titles. Only if the respondent had had the chance to choose this category, would one know whether he actually had preferred this category instead or still the category that he selected. The algorithm fails to find this or any other managerial category because the calculated phrase for text matching is 'TEACHER', which is not linked in any database to category 84194. The word 'director', however, could theoretically be linked via some databases to the desired category (and to many more managerial categories), but the text matching methods that we applied to those databases do not work if there are additional words besides the key term in the textual input. It is by no means an exception that relevant answer options are missing in the dialogue: the category which was selected by the professional coder 1 is missing for 36.0% of the eligible respondents. Upgraded algorithms and/or larger training data would be needed for improvement, although it still cannot be guaranteed that all appropriate categories will be suggested to the respondent.

Although one relevant category is missing in Fig. 1, other suggested job titles are less relevant: 'Sports teacher' is clearly implausible and the job titles 'Teacher—*Hauptschulen*' and 'Teacher—*Real-/Mittelschulen*' are repetitive. Both are associated with a single GCO category (84124), allowing respondents a detailed choice between two different school types. Yet, the GCO does not distinguish between both and it would be sufficient to ask for a single overarching category 'secondary school teacher' instead (non-existent in the *Dokumentationskennziffer*). As we restricted the number of suggested job titles to a maximum of five, such a reduction of answer options would create space for other possible categories.

4.4.2. Truck salesman

The second example illustrates a major mechanism of how interview coding leads to wrong and

uncertain codes. Consider a person who sells trucks. Our algorithm for coding during the interview is not sufficiently intelligent to suggest the correct job title 'motor vehicle seller', which would lead to the correct job code 62272. Instead, the respondent chooses the more general job title 'salesman' which appears correct to him. Unfortunately, this job title is associated with category 62102 titled 'Sales occupations in retail trade (without product specialization)', which is the wrong code for this person's actual job. The point here is that job titles from the *Dokumentationskennziffer* are not well suited to support coding during the interview. Textbooks recommend using the most specific, unambiguous terms for question design and avoiding an overlap between the answer options (e.g. Tourangeau *et al*. (2000) and Krosnick and Presser (2010)); yet, this is quite difficult to accomplish for a question about occupation. In the *Dokumentationskennziffer*, many general job titles, such as 'salesman', exist. This causes the risk that people might select a job title which appears to be correct but which leads, in fact, to a wrong code. To eliminate this type of error, one might try to reword or delete all general job titles in the *Dokumentationskennziffer* so that the meaning becomes clearer and the respondents will in no case prefer an incorrect answer option over the alternative 'other occupation'. In doing so, quality is likely to improve, but the proportion of interview-coded answers will probably decrease.

4.5. Interviewer behaviour

Our new technique was tested in a telephone survey. Compared with self-administered surveys in which the respondents can be confronted directly with the answer options suggested, the telephone survey has an extra level of interaction between respondents and interviewers. Interviewers are trained to follow the rules of standardized interviews, i.e. they are supposed to read the questions and answers exactly as worded and respondents are supposed to select the most appropriate answer without any help from the interviewer (e.g. Fowler and Mangione (1990) and Schaeffer *et al*. (2010)). This general training was not repeated for our particular survey. Since interviewers often spontaneously choose to violate these guidelines for the proposed question on occupation, it is relevant to describe how the interview-coded occupations are obtained.

Immediatly before the job titles are suggested, the algorithm needs a few seconds to calculate the most plausible job titles. Although interviewers are provided with a standardized text to explain the situation, interviewers may feel the need to keep the conversation running and to fill the gap by explaining what comes next in their own words. When the answer options pop up, it is often not necessary to read the exact question text ('Are you employed in one of the following occupations?') to proceed with the interview. In 177 out of 209 interviews (85%) that were subject to behaviour coding, the question text was not read.

Frequently, job titles are automatically suggested although they are definitely not appropriate. In the above-mentioned example (see Fig. 1), the interviewer knows from the preceding conversation that the list of suggestions contains only one job title that is appropriate in her view. Not reading out inappropriate suggestions saves time and prevents possibly confusing the respondent. This makes it attractive for interviewers to skip inappropriate job suggestions. In 97 out of 209 interviews (46%), at least one suggested job title was not read. In 10 cases (10%) this happened because the algorithm found a job title that is identical to the verbal answer that was previously provided by the respondent, in 35 cases (36%) because the job titles suggested are definitely inappropriate, and in 23 cases (24%) because of both reasons. Some interviewers steer respondents towards a specific answer: in 27 out of 209 interviews (13%), the interviewer reads out only a single job title, typically formulated in the form of a question (e.g. 'Here we have…. Is this correct?'), but sometimes also formulated as a statement, so that the respondent is not required to confirm this job title. In eight interviews (4%), the interviewers did not read out loud the suggestions at all but independently selected the most appropriate answer option.

It is also very common for interviewers to skip the answer option 'other occupation', which was read to 37 out of 209 respondents (18%) only. Reasons for this might be that the answer option is highlighted in the question text or because interviewers think that an appropriate job title had already been found.

Every question should usually be followed by an appropriate answer from the respondent. In a first reaction, 156 out of 209 respondents (75%) provided such an answer, either interrupting the interviewer (21 people) or naming it after the interviewer had finished reading out the entire question (135 people). Normally, this answer marks the end of the occupation coding process unless the respondent chooses 'other occupation' or the interviewer starts to reason with the respondent about a more appropriate category, as we have observed in a few interviews. Cases in which the respondents do not give an appropriate answer immediately are more problematic. If no job title is appropriate at first sight, respondents hesitate to answer. Because of this confusion, 17 respondents (8%) mentioned additional details about their jobs and, as a result, 'other occupation' was most often selected. Another 18 respondents (9%) were confused or asked the interviewer to explain or repeat the job title suggested. 14 out of the 18 respondents eventually agreed with one of the suggestions. In 18 additional interviews (9%), the respondents did not have a chance to speak because the interviewer was thinking or saying something without asking a question. It is then typically the interviewer who selects the most appropriate answer option.

In summary, our exercise in behaviour coding shows that many interviewers did not closely follow the rules for standardized interviews. It is the exception that an interviewer reads out the exact question text and all answer options, including the last option for 'other occupation'. When an interviewer skips a job title, decides all by himself or herself without asking the respondent, or starts a discussion with the respondent about the most appropriate answer option, one might worry that interviewer effects can be large for this question. However, these problems should not be exaggerated. Many skipped job titles are definitely inappropriate, typically respondents and not interviewers make the decision and it is not clear whether data quality is diminished when interviewers play an overly active part, as they often have a good understanding of the respondent's job. Instead, they often have good reasons for departures from the script. For future improvements of the instrument, the interplay between interviewer, question (length, number of categories, formulation) and respondent should be considered an important issue.

4.6. Algorithm analysis

Another element contributing to the overall success is the algorithm itself. The prediction algorithm should provide job category suggestions for as many respondents (i.e. verbal answers) as possible. Furthermore, these categories should be of high quality so that the respondents find their own jobs in the suggested list. In what follows, we analyse how well the algorithm currently performs regarding both objectives and search for possible ways of improvement.

Any algorithm must match the verbal responses given by respondents with some database containing possible categories. To find possible job categories for a maximal number of respondents, we apply three different databases: our training data consist of 14 912 unique entries, the search word catalogue has 153 588 entries and there are 24 000 entries in the alphabetic dictionary which is part of the 2010 GCO. However, a larger size of the database does not imply more matches. Matching respondents' answers with identical entries in the respective database provides job category suggestions for 45.7%, 46.5% and 40.8% of the 1064 respondents who answered the open-ended questions on employment. Despite the different sizes of the databases, these numbers are remarkably similar, probably because the alphabetic dictionary and the search word catalogue were not constructed for our purpose.

Many respondents reply to the open question with common and precise one-word job titles that can easily be matched with any database. These people are easy to code, either during or after the interview. In our sample, 33.6% of the respondents provided answers that enable identical matching with any database, showing that the different databases have an enormous overlap.

However, all databases fail to make suggestions via exact matching for at least half of the respondents. To overcome this limitation, two additional inexact matching methods were implemented. Results for all the different text matching methods and all databases are shown in Table 6. When the verbal answer is not required to be identical with a database record but only needs to be a substring of it, more matches are found (49.2% *versus* 45.7% and 51.8% *versus* 46.5%), but the gains are relatively small. This is because this matching technique is appropriate only for short answers. 349 respondents (32.8%), however, provided longer answers with at least three words (operationalized by two blank characters), of which only 45 can be matched with the above-mentioned identical and substring matching methods.

The second inexact matching method is more promising for longer answers: when searching for a meaningful subsequence of words (equal to a phrase as defined above) in the original verbal answer, which is then again matched to the different databases, the number of matches increases considerably, as can be seen in the lower half of Table 6.

Column (2), 'Percentage of respondents for whom at least one suggested category was also coded by at least one professional coder', confirms that we find suitable matches with all methods. For most respondents and any matching method, categories are suggested that are relevant in

Table 6. Descriptive results for various matching methods and databases†

m	*Matching method*	*(1) (%)*	*(2) (%)*	*(3)‡*		
				Median	*Mean*	*Maximum*
	Answer matches with training data					
1,3,4	Identical	45.7	39.9	2	4.2	45
2	Answer is substring	49.2	43.1	4	8.0	122
	Answer matches with file of search words					
6	Identical	46.5	39.7	2	3.8	66
7	Answer is substring	51.8	46.9	5	12.6	187
8	Answer matches with alphabetic dictionary	40.8	38.9	GCO/DKZ 2/23	GCO/DKZ 4.8/71.3	GCO/DKZ 69/1012
	Phrase matches with training data					
9–11	Identical	73.9	57.0	3	7.0	45
—§	Answer is substring	82.1	69.8	8	57.3	1479
	Phrase matches with file of search words					
13	Identical	71.4	52.3	3	6.8	82
14	Answer is substring	83.7	72.5	12	133.6	3878
15	Phrase matches with alphabetic dictionary	57.2	52.3	GCO/DKZ 2/30	GCO/DKZ 7.5/94.7	GCO/DKZ 96/1190

†Column (1), percentage of respondents for whom the matching method suggests at least one category. Column (2), percentage of respondents for whom at least one suggested category was also coded by at least one professional coder. Column (3), average number of categories, provided that at least one category is suggested.
‡GCO/DKZ: the alphabetic dictionary links job titles only to categories from the 2010 GCO. All *Dokumentationskennziffer* categories that are associated with the so-found GCO categories are possible candidates for suggestion. We thus provide the number of GCO suggestions first and the number of *Dokumentationskennziffer* suggestions second.
§This matching method was not included in the production software.

Table 7. Productivity of the coding system under various hypothetical situations†

Ask first inquiry only if . . .	*(1)*	*(2)*	*(2)/(1) (%)*	*(3)*	*(3)/{(1)−(2)} (%)*
Condition (a): . . . identical match with training data and match with alphabetic dictionary	386	12	3.1	312	83.4
Condition (b): . . . no shorter phrase is found	532	27	5.1	416	82.4
Condition (c): . . . no shorter phrase is found or phrase matches with alphabetic dictionary	712	60	8.4	511	78.4
Condition (d): always (actual condition in this study)	915	145	15.8	574	74.5

†Column (1), number of respondents who would be asked under the given condition. Column (2), number of respondents who answer 'other occupation' under the given condition. Column (2)/(1), column (2) divided by column (1). Column (3), number of respondents under the given condition who select a code that is in agreement with at least one professional coder. Column (3)/{(1)−(2)}, column (3) divided by the difference between columns (1) and (2).

the sense that professional coders usually select one of the suggested categories independently. This is not self-evident—especially in the case of the phrase matching methods it does happen that the phrase itself is meaningless for coding (e.g. words like 'in' or 'and') and matching such a phrase certainly brings no improvement.

The downside of inexact matching is summarized in column (3), 'Average number of categories if at least one category is suggested'. Identical matching methods usually suggest small numbers of possible categories and inexact matching methods find larger numbers. Obviously, not all suggested categories are always appropriate for a given occupation and it is also prohibitive to suggest dozens or hundreds of categories to a respondent during the interview. The overall performance of the system shows that these difficulties are well absorbed by the gradient boosting algorithm, which calculates correctness probabilities for all categories that are suggested by any matching method. Boosting thus integrates the different matching methods to a single prediction algorithm and allows finding the most probable categories.

These descriptions suggest a trade-off with each additional matching method. On the one hand, adding a matching method offers the possibility that additional categories can be suggested to the respondents. On the other hand, suggesting more categories can also mean suggesting more unsuitable categories, which may protract the interview, induce more people to choose 'other occupation', or lead to inaccurate coding. Therefore, system improvements might be expected if candidate job categories are not suggested to all possible respondents but only to a subgroup for which the matching methods meet specific criteria. Residual respondents would not come in contact with our proposed system. We searched for corresponding criteria and found three possible conditions to be particularly meaningful. Table 7 presents the hypothetical results for a modified algorithm, i.e. it shows what would have happened if these conditions had been applied in the field. The conditions are as follows.

(a) Answers have identical matches in both the training data and the alphabetic dictionary.
(b) No shorter phrase is found. This condition comprises all cases from the first condition with only two exceptions.
(c) The second condition holds or, alternatively, a phrase is found that must match with the alphabetic dictionary. A match with the alphabetic dictionary confirms that the phrase is a job title which makes this term especially relevant for coding.

Column (1) in Table 7 shows that the number of respondents who are presented with job category suggestions increases when the conditions are loosened, allowing more respondents

to code their occupation during the interview. At the same time, not only the absolute number (column (2)) but also the proportion (column (2)/(1)) of respondents who select 'other occupation' increases. This is detrimental to the original goal of keeping interview times in check because those respondents would be asked an additional open question. Furthermore, the proportion of respondents who select a code that is in agreement with at least one professional coder (column (3)/{(1) − (2)}) decreases when the conditions are loosened, suggesting that the quality of interview coding is also affected. The trade-off hypothesis is thus confirmed.

Which condition should be chosen to find an optimal balance between both objectives? In our opinion, condition (c) is best. (712−60)/1064 = 61.3% of the respondents would have chosen a job title during the interview under this condition, which is still a considerable proportion. At the same time, only 60/1064 = 5.6% of the population would have selected 'other occupation', which is a substantial improvement. It is not acceptable to have (145 − 60)/(915 − 712) = 41.9% of the respondents who do not fulfil condition (c) select 'other occupation', as it was implemented in the tested system.

This result also has implications for our algorithm. Job category suggestions are satisfactory when verbal answers are short and can be matched by identical or substring matching to any database. The predictions are still sufficiently accurate if the algorithm can extract a phrase from a multiworded verbal answer that is a job title from the alphabetic dictionary. The remaining verbal answers require more attention to improve the algorithm further. They may be characterized as follows. The algorithm finds a shorter phrase that is not listed in the alphabetic dictionary for 203 verbal answers. These answers contain at least two words—often more—but frequently lack a single job title that would be most relevant for coding. Algorithms that exploit interactions between words can prove useful here but were not employed so far. For 106 answers the algorithm does not find a single match in any database. These answers usually consist of a single word. Spelling errors and compound words are frequent reasons why matching is not possible. Future improvements of the algorithm should address these problems.

To motivate our algorithm, we claimed that better predictions can be achieved when the algorithm learns not only from training data but also from existing databases. Did we succeed? We compare our algorithm with predictions from multinomial regression with elastic net regularization as implemented by Friedman *et al*. (2010). We use the same training data with identical covariates as before to train the multinomial regression model. The full text, as obtained from the first verbal answer, is converted to a document term matrix that counts how often each word appears in the respondents' answers. As the software package requires that each category in the outcome variable occurs at least twice, we remove 680 cases from the training data whose *Dokumentationskennziffer* codes occur only once. All covariates are dummy coded. The problem is thus to predict one of 1600 *Dokumentationskennziffer* codes from a sparse predictor matrix of size 32 275 observations times 10 930 columns. We explore various tuning parameters to estimate the model and obtain best results when setting $\alpha = 0.05$ and $\lambda = 0.001$, implying only weak regularization. This model is used to predict job titles and associated correctness probabilities in the current study for all 1041 respondents for whom we have codes from coder 1 available. After running the same procedure as in our original algorithm to select five job titles, multinomial regression suggests at least one job title from the same GCO category that was chosen by coder 1 for 557 respondents (54%). Our own algorithm reaches a slightly better performance, suggesting at least one job title from the same GCO category for 578 respondents (56%). Although a sceptic may argue that this small improvement does not justify the complexity of our algorithm, we are more optimistic and suggest including the predictions that are obtained via multinomial regression as another covariate in the boosting procedure. This should improve the performance of our own algorithm even further.

4.7. Classification material

Two features of the classification material should be highlighted as contributing factors to coding errors, both in interview coding and in traditional post-survey coding efforts.

First, a classical strategy for automatic occupation coding is to search for a given job title in a database and to assign the associated category accordingly. We matched the first verbal answer from the interview to a database that we prepared from the alphabetic dictionary of 24 000 job titles that is part of the 2010 GCO. Although we matched job titles only if they were clearly associated with a single category, successful exact database matches were found for 418 out of 1064 verbal responses. For these people, it was then possible to compare the codes with those obtained from manual coding (coder 1 and coder 2), with the following results: all three codes are identical for 307 responses (73.4%), only one manual coder agreed with the code from the database for 88 responses (21.1%) and both disagreed with the database for 23 responses (5.6%). These numbers show that a substantial proportion of respondents mention job titles that can be coded automatically in some category with the alphabetic dictionary, though this does not mean that these categories are the only possible categories. Manual coders frequently disagree with those codes and base their decision on more information, which they retrieve from additional answers. Many job titles exist whose semantic content is vague and does not uniquely determine a single correct job category. If a coding technique relies on vague job titles—and the proposed system for coding during the interview does so excessively, like many other approaches—we cannot hope for an optimal quality of coding which guarantees that every respondent will be classified in the category that describes his or her occupational tasks and duties best.

Another source of error that leads to low intercoder reliabilities can be found in both manual and interview coding. Coders are usually required to select a single correct category; multiple categories are not permitted, even if appropriate. The decision for a single category can be difficult, either because information from the respondent to determine a precise category is missing or because categories from the job classification are not pairwise disjoint and, as a consequence, the occupational activity does not belong to a single category. The following numbers indicate that this issue requires further attention. When looking only at the subset of respondents for which both student assistants agreed that the assigned codes from coder 1 and coder 2 both are acceptable, we can have high confidence that both codes for this subset of 137 respondents are correct. However, for 52 respondents in this subset, both codes are different and it appears that more than one category may be considered correct.

5. Summary and conclusion

Traditional coding of occupations is costly and time consuming. In our study, two independent coders obtained a reliability of 61.11%: a number that is low but by no means an exception. We implemented and tested a technical solution with increased interaction during the interview to counter these challenges. After a verbal answer has been entered in the interview software, the computer automatically calculates a small set of possible job categories and suggests them to the respondent, who in turn can select the most appropriate. Our results show that this strategy for interactive coding during the interview is technically feasible.

Our system achieves high productivity: 72.4% of the respondents choose an occupation during the interview. The proportion for which manual coding is still necessary is thus reduced to 27.6%. This result is promising because coding costs can be saved and data are available directly after the interview.

The quality of interview coding was compared with that of two professional coders and was

found to be slightly lower than the quality of the first coder and comparable with the quality of the second. We also find frequent disagreement between both coders, which can be partly attributed to a lack of information provided by the respondents and to the fact that both coders observed different coding rules. Our desire to increase the quality of the coded occupations by collecting more information already during the interview was not fulfilled for several reasons: categories that are suggested by the algorithm are sometimes inappropriate, the two generated follow-up questions are unsuited to elicit more appropriate codings and respondents occasionally select overly general job titles, which lead to incorrect categories.

For respondents whose occupations are coded successfully during the interview, the duration of the interview is reduced by a few seconds; others who do not select one of the categories suggested will have to bear the burden of slightly longer interviews with an additional question. This is a major drawback of the tested system, affecting 13.6% of the population.

Our system was optimized to achieve high productivity. This may not be the best strategy because marginal gains at high levels of productivity imply larger costs in terms of the number of people who will have to endure longer interviews. We instead suggest a different strategy that finds an optimal balance between both objectives. For this, we identify four conditions that are easy to implement in the current algorithm. One condition, which would decrease the productivity rate from 72.4% to 61.3%, is recommended in particular because, under this condition, fewer respondents (5.6% compared with 13.6% now) would have to bear the burden of longer interviews.

These results are satisfactory for the first trial of a complex instrument. The key component of the system proposed—a machine learning algorithm that suggests possible answer options during the interview—works well. Other minor features were tested but their results are discouraging. Some obvious adaptions would be necessary for future application. In addition, it would be useful to estimate whether the instrument proposed leads to cost reductions in the coding process. At the very least, our results show that coding during the interview can become a viable technique that may partly replace traditional post-interview coding in the future.

Before implementing the new instrument in a production environment, we recommend further testing in more practical settings. Our study has some limitations and survey operators may want to ask the following questions for their own application. First, is it possible to achieve a similar or better performance if some occupations were not underrepresented, as we have reported in our study? Second, what would happen if the interviewer and the algorithm had less information to predict a person's job from occupation-related questions preceding the new tool? Third, what would have to be changed in the proposed instrument if the researcher was not interested in self-reported occupations from telephone interviews, but in other types of occupation (e.g. job aspirations of adolescents and occupations of spouses and parents) that might be collected via different modes of operation (e.g. Internet surveys or computer-assisted personal interviewing)?

Throughout this paper, we described the strengths and weaknesses of the proposed instrument. For future developments, we have identified the following factors how to improve the process.

A supervised learning algorithm was used to generate plausible job category suggestions for the respondents. With an improved algorithm and additional training data, it is likely that the productivity of the system can be further increased. In the frequent situation that a verbal answer comprises more than one word and does not contain a predefined job title, we suspect largest gains in productivity. Spelling correction and the splitting of compound words may also prove to be helpful.

When respondents choose one of the job titles suggested, it is too often not the most appropriate. Respondents frequently select general job titles that are not entirely wrong but link to

suboptimal GCO categories. These inappropriate job titles stem from the *Dokumentationskennziffer*, which is therefore not well suited for coding during the interview. To preclude the possibility that respondents select an incorrect category, we recommend the development of an auxiliary classification that describes answer options more precisely. All answer options from this auxiliary classification should map to a single category in both classifications, national (2010 GCO) and international (2008 ISCO), for simultaneous coding.

Interviewers frequently did not act according to the rules of standardized interviews at the question proposed but often preferred rewording the question text and skipping suggested answer options. Although this behaviour leads to concerns about interviewer effects, we must not forget the positive effect: respondents are not confused by strange answer options and the duration of the interview is reduced. For an improved instrument, one may even try to provide interviewers with a medium-sized number of answer options (say 10). Since respondents cannot intellectually process so many answer options in a telephone interview, one would also explicitly request interviewers to skip inappropriate job categories. This procedure could partly remedy the current problem that the algorithm finds many possible job titles, but the most appropriate job category is not suggested to about 36% of the respondents. Furthermore, extended interviewer training will be necessary to ensure that interviewers know when they must follow the script and to reduce the risk of omitting relevant answer options.

Some answers in reply to the first open-ended question about occupation are very general and one would need to suggest a huge number of possible categories. Instead, our vision is to recognize these general answers automatically. An additional open-ended question would then be asked to collect more details and this second answer could be used as input for coding during the interview. Additionally, future research should consider the possibility that more than one job category may be appropriate.

In summary, such a system for occupation coding during the interview promises an increase of quality of data while reducing costs of data collection.

Acknowledgements

Funding for this work has been provided by the German Institute for Employment Research and the Mannheim Centre for European Social Research, and by grant KR 2211/3-1 from the German Research Foundation to Frauke Kreuter. The idea for this study originates from the Master's thesis written by Malte Schierholz. Miriam Gensicke and Nikolai Tschersich contributed with valuable comments and supervised the implementation in the survey software, which was technically demanding. We thank Alexandra Schmucker for helping us to test the proposed technique in a survey commissioned by the Institute for Employment Research and operated by Kantar Public (formerly TNS Infratest Sozialforschung). This study would not have been possible without Ariane Wickler and Gerd Döring, who implemented the interface between the interview software and the predictive system that was developed by the first author. Valuable comments from Josef Hartmann led to improvements in the questionnaire. We sincerely thank the Associate Editor and the referees for their constructive comments. We further thank the three coders, our student assistants (Max Hansen and Sebastian Baur) for quality checking, Hannah Laumann for proofreading and colleagues at the German Institute for Employment Research and the University of Mannheim for helpful comments.

References

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2010) Arbeiten und Lernen im

Wandel ∗ Teil 1: Überblick über die Studie. *Methodenreport 05/2010*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
Belloni, M., Brugiavini, A., Meschi, E. and Tijdens, K. (2016) Measuring and detecting errors in occupational coding: an analysis of share data. *J. Off. Statist.*, **32**, 917–945.
vom Berge, P., König, M. and Seth, S. (2013) Sample of integrated labour market biographies (SIAB) 1975-2010. *Datenreport 01/2013*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
Biemer, P. and Caspar, R. (1994) Continuous quality improvement for survey operations: some general principles and applications. *J. Off. Statist.*, **10**, 307–326.
Biemer, P. and Lyberg, L. (2003) *Introduction to Survey Quality*. Hoboken: Wiley.
Billiet, J. and Loosveldt, G. (1988) Improvement of the quality of responses to factual survey questions by interviewer training. *Publ. Opin. Q.*, **52**, 190–211.
Bobbitt, L. G. and Carroll, C. D. (1993) Coding major field of study. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 177–182.
Bradburn, N. M. (1978) Respondent burden. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 35–40.
Breiman, L. (1996) Stacked regressions. *Mach. Learn.*, **24**, 49–64.
Bundesagentur für Arbeit (2011a) *Klassifikation der Berufe 2010*, vol. 1, *Systematischer und Alphabetischer Teil mit Erläuterungen*. Nuremberg: Bundesagentur für Arbeit.
Bundesagentur für Arbeit (2011b) *Klassifikation der Berufe 2010*, vol. 2, *Definitorischer und Beschreibender Teil*. Nuremberg: Bundesagentur für Arbeit.
Bundesagentur für Arbeit (2013) Index of search words. Bundesagentur für Arbeit, Nuremberg. (Available from `http://download-portal.arbeitsagentur.de/files/`.)
Bushnell, D. (1998) An evaluation of computer-assisted occupation coding. In *New Methods for Survey Research* (eds A. Westlake, J. Martin, M. Rigg and C. Skinner), pp. 23–36. Southampton: Association for Survey Computing.
Campanelli, P., Thomson, K., Moon, N. and Staples, T. (1997) The quality of occupational coding in the United Kingdom. In *Survey Measurement and Process Quality* (eds L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz and D. Trewin), pp. 437–453. New York: Wiley.
Cantor, D. and Esposito, J. (1992) Evaluating interviewer style for collecting industry and occupation information. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 661–666.
Conrad, F. G., Couper, M. P. and Sakshaug, J. W. (2016) Classifying open-ended reports: factors affecting the reliability of occupation codes. *J. Off. Statist.*, **32**, 75–92.
Conrad, F. G. and Schober, M. F. (2005) Promoting uniform question understanding in today's and tomorrow's surveys. *J. Off. Statist.*, **21**, 215–231.
Couper, M. and Zhang, C. (2016) Helping respondents provide good answers in web surveys. *Surv. Res. Meth.*, **10**, 49–64.
Creecy, R. H., Masand, B. M., Smith, S. J. and Waltz, D. L. (1992) Trading mips and memory for knowledge engineering. *Communs ACM*, **35**, 48–64.
Drasch, K., Matthes, B., Munz, M., Paulus, W. and Valentin, M.-A. (2012) Arbeiten und Lernen im Wandel ∗ Teil V: Die Codierung der offenen Angaben zur beruflichen Tätigkeit, Ausbildung und Branche. *Methodenreport 04/2012*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.
Elias, P. (1997) Occupational classification (ISCO-88): concepts, methods, reliability, validity and cross-national comparability. *Labour Market and Social Policy Occasional Paper 20*. Organisation for Economic Cooperation and Development Publishing, Paris. (Available from `http://dx.doi.org/10.1787/304441717388`.)
Elias, P., Birch, M. and Ellison, R. (2014) CASCOT international version 5 user guide. Institute for Employment Research, University of Warwick, Coventry. (Available from `http://www2.warwick.ac.uk/fac/soc/ier/software/cascot/internat/`.)
Fowler, F. J. and Mangione, T. W. (1990) *Standardized Survey Interviewing: Minimizing Interviewer-related Error*. Newbury Park: Sage.
Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *J. Statist. Softwr.*, **33**, 1–22.
Galesic, M. and Bosnjak, M. (2009) Effects of questionnaire length on participation and indicators of response quality in a web survey. *Publ. Opin. Q.*, **73**, 349–360.
Geis, A. (2011) Handbuch für die Berufsvercodung. *Coding Documentation*. GESIS–Leibniz-Institut für Sozialwissenschaften, Mannheim. (Available from `http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/handbuch_der_berufscodierung_110304.pdf`.)
Geis, A. and Hoffmeyer-Zlotnik, J. H. (2000) Stand der Berufsvercodung. *ZUMA-Nachr.*, **24**, 103–128.
Granquist, L. and Kovar, J. (1997) Editing of survey data: how much is enough? In *Survey Measurement and Process Quality* (eds L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz and D. Trewin), pp. 415–435. New York: Wiley.
Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M. and Steiner, S. (2017) Three methods for occupation coding based on statistical learning. *J. Off. Statist.*, **33**, 101–122.

Hacking, W., Michiels, J. and Janssen-Jansen, S. (2006) Computer assisted coding by interviewers. In *Proc. 10th Int. Blaise Users Conf.* (ed. J. Bethlehem), pp. 283–296. Arnhem: International Blaise User Group.

Hoffmann, E., Elias, P., Embury, B. and Thomas, R. (1995) What kind of work do you do?: Data collection and processing strategies when measuring "occupation" for statistical surveys and administrative records. *STAT Working Paper. 95–1*. Bureau of Statistics, International Labour Office, Geneva. (Available from `http://www.ilo.org/public/libdoc/ilo/1995/95B09_135_engl.pdf`.)

Hoffmeyer-Zlotnik, J. H., Hess, D. and Geis, A. J. (2006) Computerunterstützte Vercodung der International Standard Classification of Occupations (ISCO-88): Vorstellen eines Instruments. *ZUMA-Nachr.*, **30**, 101–113.

Hoffmeyer-Zlotnik, J. H. and Warner, U. (2012) *Harmonisierung Demographischer und Sozioökonomischer Variablen: Instrumente für die International Vergleichende Surveyforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Holbrook, A. L., Green, M. C. and Krosnick, J. A. (2003) Telephone versus face-to-face interviewing of national probability samples with long questionnaires: comparisons of respondent satisficing and social desirability response bias. *Publ. Opin. Q.*, **67**, 79–125.

Holland, J. L. and Christian, L. M. (2009) The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Socl Sci. Comput. Rev.*, **27**, 196–212.

Hothorn, T., Bühlmann, P., Kneib, T., Schmid, M. and Hofner, B. (2010) Model-based boosting 2.0. *J. Mach. Learn. Res.*, **11**, 2109–2113.

International Labour Office (2012) *International Standard Classification of Occupations: ISCO-08*. Geneva: International Labour Organisation.

Javed, F., Luo, Q., McNair, M., Jacob, F., Zhao, M. and Kang, T. S. K. (2015) Carotene: a job title classification system for the online recruitment domain. In *Proc. 1st Int. Conf. Big Data Computing Service and Applications, Redmond City*, pp. 286–293. New York: Institute of Electrical and Electronics Engineers.

Jung, Y., Yoo, J., Myaeng, S.-H. and Han, D.-C. (2008) A web-based automated system for industry and occupation coding. In *Web Information Systems Engineering* (eds J. Bailey, D. Maier, K.-D. Schewe, B. Thalheim and X. Wang), pp. 443–457. Berlin: Springer.

Krosnick, J. and Presser, S. (2010) Question and questionnaire design. In *Handbook of Survey Research* (eds P. V. Marsden and J. D. Wright), pp. 263–313. Bingley: Emerald.

LeBlanc, M. and Tibshirani, R. (1996) Combining estimates in regression and classification. *J. Am. Statist. Ass.*, **91**, 1641–1650.

Loos, C., Eisenmenger, M. and Bretschi, D. (2013) Das Verfahren der Berufskodierung im Zensus 2011. *Wirtsch. Statist.*, 173–184.

Maaz, K., Trautwein, U., Gresch, C., Lüdtke, O. and Watermann, R. (2009) Intercoder-Reliabilität bei der Berufscodierung nach der ISCO-88 und Validität des sozioökonomischen Status. *Zeits. Erziehungs.*, **12**, 281–301.

Mangione, T. W., Fowler, F. J. and Louis, T. A. (1992) Question characteristics and interviewer effects. *J. Off. Statist.*, **8**, 293–307.

Mannetje, A. T. and Kromhout, H. (2003) The use of occupation and industry classifications in general population studies. *Int. J. Epidem.*, **32**, 419–428.

Measure, A. (2014) Automated coding of worker injury narratives. *Proc. Gov. Statist. Sect. Am. Statist. Ass.*, 2124–2133.

Office for National Statistics (2003) Quality of data capture and coding: evaluation report. *Census 2001 Review and Evaluation Report*. Office for National Statistics, Titchfield. (Available from `http://www.ons.gov.uk/ons/guide-method/census/census-2001/design-and-conduct/review-and-evaluation/evaluation-reports/processing/quality-of-data-capture-and-coding—evaluation-report.pdf`.)

Ongena, Y. P. and Dijkstra, W. (2016) Methods of behavior coding of survey interviews. *J. Off. Statist.*, **22**, 419–451.

Paulus, W. and Matthes, B. (2013) Klassifikation der Berufe ∗ Struktur, Codierung und Umsteigeschlüssel. *Methodenreport 08/2013*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Roberts, C., Gillian, E., Allum, N. and Lynn, P. (2010) Data quality in telephone surveys and the effect of questionnaire length: a cross-national experiment. *Research Working Paper 2010-36*. Institute for Social and Economic Research, University of Essex, Colchester. (Available from `https://www.iser.essex.ac.uk/publications/working-papers/iser/2010-36`.)

Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P. and Singer, E. (2016) Evaluating active (opt-in) and passive (opt-out) consent bias in the transfer of federal contact data to a third-party survey agency. *J. Surv. Statist. Methodol.*, **4**, 382–416.

Schaeffer, N. C., Dykema, J. and Maynard, D. W. (2010) Interviewers and interviewing. In *Handbook of Survey Research* (eds P. V. Marsden and J. D. Wright), pp. 437–470. Bingley: Emerald.

Schierholz, M. (2014) Automating survey coding for occupation. *Methodenreport 10/2014*. Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Schnell, R. (2012) *Standardisierte Befragungen in den Sozialwissenschaften*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Speizer, H. and Buckley, P. (1998) Automated coding of survey data. In *Computer Assisted Survey Information Collection* (eds M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II and J. M. O'Reilly), pp. 223–243. New York: Wiley.

Stone, M. (1974) Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc.* B, **36**, 111–147.

Svensson, J. (2012) Quality control of coding of survey responses at Statistics Sweden. In *Proc. Eur. Conf. Quality in Official Statistics*. Athens: Hellenic Statistical Authority–Eurostat.

Thompson, M., Kornbau, M. E. and Vesely, J. (2014) Creating an automated industry and occupation coding process for the American Community Survey. *Federal Economic Statistics Advisory Committee Meet*. US Census Bureau, Suitland. (Available from `http://www.census.gov/about/adrm/fesac/meetings/june-13-2014-meeting.html`.)

Tijdens, K. (2014a) Reviewing the measurement and comparison of occupations across Europe. *Working Paper 149*. Amsterdam Institute for Advanced Labour Studies, University of Amsterdam, Amsterdam. (Available from `http://hdl.handle.net/11245/1.432281`.)

Tijdens, K. (2014b) Dropout rates and response times of an occupation search tree in a web survey. *J. Off. Statist.*, **30**, 23–43.

Tijdens, K. (2015) Self-identification of occupation in web surveys: requirements for search trees and look-up tables. In *Survey Insights: Methods from the Field* (eds H. Best, P. Farago, D. Joye, L. Kaczmirek, C. Vandenplas, M. Vettovaglia and C. Wolf). Lausanne: Swiss Foundation for Research in Social Sciences–GESIS–Leibniz Institute for the Social Sciences.

Tourangeau, R., Rips, L. J. and Rasinski, K. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

United Nations and International Labour Office (2010) *Measuring the Economically Active in Population Censuses: a Handbook*. New York: United Nations and International Labour Office.

de Waal, T., Pannekoek, J. and Scholtus, S. (2011) *Handbook of Statistical Data Editing and Imputation*. Hoboken: Wiley.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Appendix to: Occupation coding during the interview'.

Appendix to: *Occupation Coding During the Interview*

Malte Schierholz
Mannheim Centre for European Social Research,
University of Mannheim, Germany
Institute for Employment Research, Nuremberg, Germany

Miriam Gensicke
TNS Infratest Sozialforschung, Munich, Germany

Nikolai Tschersich
TNS Infratest Sozialforschung, Munich, Germany

Frauke Kreuter
University of Maryland, College Park, USA, University of Mannheim
and Institute for Employment Research, Nuremberg, Germany

Contents

1 Part A: Additional Results **2**

2 Part B: Questionnaire **6**
2.1 English Translation 6
2.2 German Original 22

3 Part C: Instructions for Validation **39**

4 Part D: Behavior Coding Manual **46**

5 References **48**

Part A: Additional Results

Results for the second and third inquiry

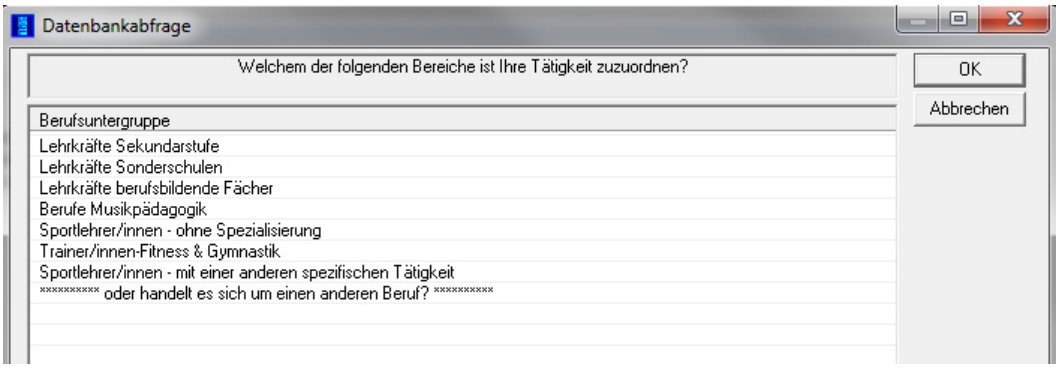


Fig. A 1: Follow-up screenshot related to figure 1 in the text. This question appears after when a respondent answers "vice principal and teacher" and selects a "different occupation" in figure 1. It shows possible 4-digit occupation sub-groups from the KldB 2010.

Table A 1: Results for the second and third inquiry, when the respondent chooses "other occupation" in the first inquiry (cf. table 1)

Number of respondents	145	100.0%
Item nonresponse in second inquiry	8	5.5%
No occupation sub-group in second inquiry selected, break-off	77	53.1%
Occupation sub-group chosen, item nonresponse in last inquiry	1	0.7%
Occupation sub-group chosen, but no job chosen in third inquiry	28	19.3%
Occupation sub-group chosen, but selected job category from third inquiry is not in agreement with coder 1	24	16.6%
Occupation sub-group chosen and selected job category from third inquiry is in agreement with coder 1	7	4.8%

Results in case the algorithm suggests only a single job category

If the algorithm suggests only a single job title, not this job title but a full job description was read to the respondent, who was asked to agree, partly agree, or disagree. This applies to 30 respondents. Job descriptions were not tailored for usage in a survey and respondents often did not understand them correctly, which causes frequent errors.

Table A 2: Results in case the algorithm suggests only a single job category. This triggers a special question "Is the following description correct for your occupation?" (Question 6.23a)

	Yes	No	In parts
Manual coding by coder 1 in agreement	12	3	1
Manal coding by coder 1 in disagreement	5	6	3

Results in case the algorithm finds more than 250 possible categories

If more than 250 job titles are suggested by the algorithm (applicable for 10 respondents), the first list with job titles usually shown to the interviewer is skipped and occupational sub-group titles from the KldB 2010 are asked instead. This is the same question that standard respondents get if they answer "other occupation" (Question 6.24a, Figure A1). After selecting an occupational sub-group category, respondents are provided with a detailed follow-up question in which they can select a job title (Question 6.24b).

Table A 3: Results in case the algorithm suggests at least 250 possible categories

Number of respondents	10	100%
Selected job title is in agreement with coder 1	4	40%
No job title was selected by the respondent	6	60%

Additional answer option: similar occupation

In setting up this study, we were concerned that people would too often choose an occupation that is not the perfect choice. A small experiment was included in the survey to encourage people not to select a job title from the first inquiry but to move forward to the second inquiry regarding 4-digit occupation sub-groups, hoping that respondents are more successful to find the correct job title in the second and third enquiry. For 77 respondents, who were selected at random, the first inquiry was slightly changed and an additional answer option "or do you work in a similar occupation?" was added. When this option is selected, the interview proceeds exactly as if the answer option "different occupation" were chosen. In the rest of this paper we do not distinguish between both options and use the general term "other occupation" for both cases. Table A4 compares both experimental conditions.

Table A 4: Experiment comparison

	Single answer option: "Different occupation"	*Additional answer option: "Similar occupation"*
No of eligible respondents	839	76
No of respondents who select "different occupation"	139	4
No of respondents who select "similar occupation'"	/	2
Proportion of eligible respondents who select "other occupation"	16.6%	7.9%
No of respondents who select a code that either agrees with coder 1 or with coder 2	520 (74% of interview-coded answers)	54 (77% of interview-coded answers)

The additional answer option was apparently not appealing and, contrary to our intention, makes it even less attractive to select "other occupation". We calculate the odds ratio OR = (700/139)/(70/6) = 0.43 and test the null hypothesis that no differences exist between one and two answer options, H0: OR = 1. The p-value from Fisher's Exact Test (two-sided) is significant with p = 0.04882. This result is contrary to our initial expectations and suggests that providing an additional answer option decreases the probability for respondents to select "other occupation".

The p-value is close to 0.05 and despite its significance it is still possible that chance alone can explain it. Two arguments exist in particular, why we do not believe that there is a real difference between both conditions: (1) If it made a difference and respondents were in fact more inclined to select one of the suggested

job titles whenever the additional option "similar occupation" is offered, respondents would have to choose inaccurate job titles more often and the quality would thus decrease. The last row in table A4 contrasts this logic, showing that agreement with professional coders is even higher for the respondents who get the additional answer option. (2) We also know from the analysis of interviewer behavior that, under the standard condition, the single answer option "different occupation" was only read aloud in 18% of the interviews. It is hard to explain why this answer option is more often selected although interviewers take so little note of it.

Can we detect from the interview if the respondent is correct?

Respondents may have trouble to select a job title when no suggested job title is entirely correct. We hypothesized that respondents who run into difficulties answering the question are less likely to provide an accurate answer. To uncover these problematic respondents, interviewers were asked if the respondent found it difficult to answer the question (no. 6.26 in appendix B). Table A5 contains the results. Short hesitation appears to be an indication of erroneous interview coding. However, the difference to standard behavior is not pronounced enough to use this characteristic for discrimination between accurate and questionable codes.

Table A 5: Analysis of respondent's behavior in interview coding

Behavior	*Freq. Behavior*	*Freq. Correct*	*Proportion Correct*
No anomalies	657	496	75.5%
Short hesitation	86	56	65.1%
Thinking for seconds	15	12	80.0%
Thinking aloud resp. asking queries	11	9	81.8%
Not applicable	1	1	100.0%

The second column shows frequencies for various behaviors. The third column shows how often the interview-coded category is in agreement with at least one professional coder. The last column gives the proportion how often the interview-coded category is in agreement, given the specified behaviour (i.e., Freq. Correct / Freq Behavior).

Part B: Questionnaire

Only the relevant parts from our questionnaire are documented here. The English translation is provided first, the original German questionnaire follows below. Question wordings and translations are heavily influenced and often identical to the questionnaire from the panel "Arbeitsmarkt und soziale Sicherung" (Trappmann et al., 2013).

The coding process starts by asking one open-ended question about the occupation ("Please tell me your occupational activity", number 6.3 in the questionnaire below). In very few cases (4.3% of the respondents), when the answer appears in a predefined list from TNS Infratest Sozialforschung containing overly general job titles (e.g., "salesman", "clerk"), another open-ended question is asked (no. 6.4). We also consider it helpful, although by no means necessary, for the productivity of our system to ask additional closed questions that are predictive for a person's job and common in many surveys (no. 6.2, 6.7-6.10, 6.13 - 6.18). Based on all these answers, the algorithm suggests possible job categories and the respondent can then select the most adequate one (no. 6.23b by default, but no. 6.23a, 6.24a, and 6.24b are related).

To compare the interview-coded category with professional manual coding, additional questions are asked in between (no. 4.2, 4.3, 6.5, 6.6, 6.11, 6.12). In particular, we follow the recommendations from the demographic standards (Statistisches Bundesamt, 2010) with some minor modifications and ask three open-ended questions (no. 6.3, 6.5, 6.6) to collect as many details as possible about the respondents' job for manual coding. Note that most questions mentioned above are primarily asked to evaluate the quality of interview coding in comparison with manual coding. If our suggested technique were applied in practice, it would be sufficient to ask questions 6.3 and 6.23b only and skip the additional questions, saving valuable interview time. Only if an occupation cannot be coded during the interview, questions 6.5 and 6.6 would still need to be asked for ex-post manual coding.

Interviews were carried out with the NIPO fieldwork system (NIPO Software, 2014) that writes relevant data to a MySQL database (Oracle Corporation, 2014). At the same time, NIPO launches the statistical programming language R (R Core Team, 2012) that processes the data and writes the suggested categories for the respondent back to the database. NIPO, in turn, reads the new database entries and shows the suggested categories to the interviewer. Because the algorithm requires more RAM than what is available on typical interviewer computers, calculations were carried out on an external server. For data handling in R, we employed the packages Rserve (Urbanek, 2013), foreign (R Core Team, 2014), data.table (Dowle et al., 2014), tm (Feinerer et al., 2014), stringr (Wickham 2012), and ODBC (Ripley and Lapsley, 2013).

English Translation

[4.2.] What is your highest degree of vocational training?

INT: Read all answer options. For degrees from foreign countries, let the respondent decide: What would be an equivalent degree in Germany?

1: Apprenticeship or vocational training in a company

2: Training at full-time vocational school (Berufsfachschule / Handelsschule/ school for health care professionals)

3: A master craftsman qualification, a technician qualification or a comparable advanced secondary vocational qualification

4: Degree from a University of Applied Sciences (Fachhochschule)

5: A college or university degree

6: A different degree, namely _____ [open]

*** 95 no degree from vocational training -> continue with 4.4

*** 96 currently in vocational training -> continue with 4.4

*** 98 DK -> continue with 4.4

*** 99 REF -> continue with 4.4

[numeric, 2-digits]

[4.3.] Occupation of highest degree from vocational training

<Variant 1: 4.2 = 1| 2 | 3 | 6?>

What was the exact occupation you were trained in?

Interviewer: *Ask for the exact job title. For example not "mechanist" but "precision or car mechanist"; not "teacher" but "teacher at a 'gymnasium'".* In case of several equivalent degrees, write down information about the most recent degree.

<Variant 2: 4.2 = 4 | 5>

What was the main subject you studied?

Interviewer: Ask for exact subject of study

_____ [in case there are two main subjects:] _____

*** 8 DK

*** 9 REF

[open]

[Other questions not relevant for occupation coding]

Employment Biography since 01.01.2010:

[6.1] Current and earlier jobs

Retrospective collection of up to 3 activities where the 2nd activity must begin prior to the 1st and the 3rd activity must begin prior to the second. End dates of activities are not validated.

At most 3 loops

<1. loop: **We would now like to talk with you about your job history in the last years. For example, we would like to know if you have been gainfully employed, doing an apprenticeship, registered as unemployed, or doing something else. It is important that you indicate every single activity, even if it lasted only for a short while. If you were involved in several activities at the same time, please refer to your main occupation.**

Beginning with today: what of the following is correct? Are you currently …

<2. or following loops: **What did you do prior to <start date from [6.27], [6.61], or [6.71] of the preceding loop>, that is before**

<*if [6.1] in the preceding loop = 1:* **the phase of paid employment**

if [6.1] in the preceding loop = 2: **the phase of unemployment**

if [6.1] in the preceding loop = 3: **your school visit**

if [6.1] in the preceding loop = 4: **your vocational training, the apprenticeship, or your studies**

if [6.70] in the preceding loop = 1: **your military service, your community service (Zivildienst oder Bundesfreiwilligendienst), or similar**

if [6.70] in the preceding loop = 2: **your workings as a <***female:* **housewife** *male:* **househusband>**

if [6.70] in the preceding loop = 3: **<**only for females: **your maternity leave> or your parental leave (Erziehungsurlaub und Elternzeit)**

if [6.70] in the preceding loop = 4: **your retirement, your pension, or your early retirement**

if [6.70] in the preceding loop = 5: **the phase>**

that we just talked about? What of the following is correct? Were you before that …

*** INT.: *If there were parallel phases, the phase of gainful employment dominates. For parallel employments, the employment with more working hours dominates. Under the gainfully employed category, we also aggregate part-time jobs from students, housewives, retired persons, and mini jobs (also called 400 resp. 450€-Jobs).*

1: gainfully employed, including part-time work and mini jobs ->continue with 6.2

2: registered as unemployed. We would like you to include as well times in which you participated in a measure or a programme of the 'Arbeitsagentur' or a job centre
-> continue with [6.61]

3: Student at a school -> continue with [6.61]

4: doing vocational training, an apprenticeship or a course at university or college
-> continue with [6.61]

5: or <*1. loop:* are, 2. loop: were> you doing something different?-> continue with [6.70]

***8: DK -> continue with 8

***9: REF -> continue with 8

[numeric, 1-digit]

[6.2] Job position

<Filter: Spell is current Employment (first loop)>

And what is your current job position? Are you …

<Filter: Spell is not the current employment >

And what was your job position at the end of your employment? Were you …

Interviewer: Read all answer options.

<Show answer options depending on sex>

1 Blue-collar worker

2 White-collar worker

3 Professional soldier or short-term career soldier

4 Judge or civil servant

5 Self-employed in an independent profession, that is doctor, lawyer or architect

6 Self-employed in trade or craft, commerce, industry, services

7 Self-employed farmer

8 Freelancer

9 Family member working for a self-employed relative or

10: <*1. loop:* is *2. loop or following:* was> it a mini job, also called 400€-job earler, 450€-job now? ->continue with [6.18]

97: target person cannot decide between blue-collar worker and white-collar worker.

98: DK

99: REF

[numeric, 2-digits]

[6.3] And please tell me the occupational activity you <1. loop: do 2. loop: did> .

Filter: only for the latest phase of gainful employment

*Filter: The following questions [6.3]-[6.26] are only asked **a single time**, in particular for the latest phase of employment. This means, only if the respondent is employed currently (6.1 == 1) or if he is not employed currently, then for his most recent phase of employment. After asking them a single time, they are not asked again. For all others, the next question is [6.27] (begin spell).*

Interviewer: Mind your orthography; let the respondent spell the answer if needed. Ask for the exact job title. For example not "mechanist" but "precision or car mechanist"; not "teacher" but "teacher at a 'gymnasium'".

When the job is about temporary employment, ask for the predominant occupational activity.

Programmer: Check in list of overly general job titles if question 6.4 needs to be asked

*** 99 no answer

[Open-ended]

Programmer: Save current time

[6.4] Please specify the exact job title. For example not <„clerk" but „forwarding merchant">, not <„worker" but „machine fitter">

Filter: Only for the latest phase of gainful employment and only if the answer in [6.3] was too general. Overly general statements are are defined via a list from TNS Infratest Sozialforschung, e.g. „salesman", „department head", „clerk". 4.3% of the respondents from our survey did give an answer in [6.3] that was too general.

_____

*** 99 no answer

[Open-ended]

Programmer: Save current time

Programmer:

Show male/female job titles depending on sex: <kaufmännische Angestellte, sondern: Speditionskauffrau>, <Arbeiterin, sondern: Maschinenschlosserin>

Programmer:

If this question is not asked, the answer from [6.3] should be saved at the place of [6.4]. For the algorithm call in R and later question instructions, the text from [6.4] should be used.

[6.5] Please describe your occupational activity precisely.

Filter: only for the latest phase of gainful employment

*** 99 no answer

[Open-ended]

Programmer: Save current time

[6.6] Is there a specific name for this job?

Filter: only for the latest phase of gainful employment

Filter: Ask this question only if question [6.4] was not asked

Yes, namely: ----------

No

*** 99 no answer

[Open-ended]

Programmer: Save current time

[6.7] Job position: Blue-collar worker

Filter: [6.2] = 1

<*1. loop:* **And what are you exactly? Are you . . .**

2. loop or following: **And what were you exactly? Were you . . .>?**

<Show answer options depending on sex>

1 An unskilled worker

2 A semi-skilled worker

3 A skilled worker

4 A foreman

5 A master craftsman, site foreman (Polier/in), work team leader (Brigadier/in)?

8: DK

9: REF

[numeric, 1-digit]

Filter: -> continue with [6.14]

Programmer: Save current time

[6.8] Job position: White-collar Workers

Filter: [6.2] = 2 | [6.2] = 97

<*1. loop:* **And what are you precisely? Are you an employee with...**

2. loop or following: **And what were you precisely? Were you an employee with...>?**

<Show answer options depending on sex>

<If the respondent cannot decide if he is a blue-collar or a white-collar ([6.2] = 97): Do not mention examples>

1 simple duties performed in accordance with general instruction (e.g., salesperson, secretarial assistant, nursing assistant)

2 under close supervision carrying out complex tasks independently (e.g., accounting clerk, bookkeeper, technical draftsman)

3 carrying out responsible tasks independently or with limited responsibility for others (e.g. researcher, authorized officer, department head or white-collar master craftsman)

4 with wide managerial responsibilities and decision making powers (e.g., director or executive, board member)

*** 8: DK

*** 9: REF

[numeric, 1-digit]

Filter: -> continue with [6.14]

Programmer: Save current time

[6.9] Job Position: Professional soldiers or short-term career soldiers

Filter: [6.2] = 3

<1*. loop:* **And what is your rank?**

2. loop or following: **And what was your rank?>**

<Show answer options depending on sex>

1 Enlisted personnel, other than non-commissioned officer

2 Enlisted personnel, non-commissioned officer without Portepee (Unteroffizier, Stabsunteroffizier)

3 Enlisted personnel, non-commissioned officer with Portepee (Feldwebel, Oberfeldwebel)
4 Commissioned officer, captain or lower rank (Leutnant, Hauptmann)
5 Commissioned officer, major or higher rank
*** 8: DK
*** 9: REF
[numeric, 1-digit]
Filter: -> continue with [6.14]
Programmer: Save current time

[6.10] Job position: Civil servants and judges
Filter: [6.2] = 4

<1. loop: **Are**
2. loop or following: **Were>you a civil servant with …**

1 simple administrative duties (einfacher Dienst), -> Continue with [6.14]
2 mid-level administrative duties (mittlerer Dienst)
3 senior administrative duties (gehobener Dienst)
4 executive duties (höherer Dienst)
*** 8: DK
*** 9: REF
[numeric, 1-digit]
Programmer: Save current time

[6.11] Technical service
Filter: [6.10] = 2, 3, 4, 8, 9:
<1. loop: **Are**
2. loop or following: **Were> you in technical service (technischer Dienst)?**
1: Yes
2: No
*** 8: DK
*** 9: REF
[numeric, 1-digit]
Filter: -> continue with [6.14]
Programmer: Save current time

[6.12] Job position: Self-employed farmers
Filter: [6.2] = 7
And how many hectares of agricultural land

<1. loop: **do**
2. loop or following: **did> you cultivate?**

Hectares: _ _ _ _ _ (5-digit)
99998: DK
99999: REF
[numeric, 5-digit]
Filter: -> continue with [6.13]
Programmer: Save current time

[6.13] Job position: Self-employed
Filter: [6.2] = 5 | 6 | 7
And how many employees
<1. loop: **do**
2. loop or following: **did> you have?**

Employees _ _ _ (3-digit)
996 More than 99 employees
*** 998: DK
*** 999: REF
[numeric, 3-digits]
Filter: -> continue with [6.16]
Programmer: Save current time

[6.14] Supervising responsibilities
Filter: [6.2] = 1|2|3|4|8|9|97|98|99 (not self-employed)
<1. loop: **Do**
*2. loop or following:***Did>**
your job responsibilities include supervising the work of other employees or telling them what they have to do?
INT: This refers to other employees only. If a teacher supervises pupils or a kindergarten teacher supervises children, they will not be included. But if a school headmaster supervises other teaching staff, they will of course be included.

1: Yes
2: No ->Continue with [6.16]
8: DK -> 6.19
9: REF -> 6.19
Programmer: Save current time

[6.15] Number of supervised persons:

Filter: [6.14] = 1

How many other employees

<1. loop: **do**

*2. loop or following:***did> you supervise directly?**

INT: This refers to employees only. If a teacher supervises children or a kindergarten teacher supervises children, they will not be included. But if a school headmaster supervises other teachers they will of course be included.

Number: _ _ _ _ (4-digit)

9998: DK

9999: REF

[numeric, 4-digits]

Programmer: Save current time

[6.16] Vocational training usually required

As a general rule, which kind of vocational training is for your job as <show job activity from [6.4]> required? Is no vocational training required, or is it necessary to complete semi-skilled training, to complete vocational training, training at full-time vocational school, a master craftsman qualification, or a qualification as a technician, or a degree from a University of Applied Sciences (Fachhochschule) or university?

1: no vocational training

2: semi-skilled training

3: completed vocational training

4: completed training at full-time vocational school

5: a master craftsman qualification, or a qualification as a technician

6: a degree from a University of Applied Sciences (Fachhochschule) or universiy

8: weiß nicht

9: keine Angabe

[numeric, 1-digit]

Programmer: Save current time

[6.17] Industry

Filter: [6.2] != 3|7|10 (not soldier or self-employed farmer)

<Variant 1: [6.2] = 1|2|4|8|97|98|99>

<1. loop: **Does the company you are working with belong to**

2. loop or following: **Did the company you were working with belong to>…**

INT: Read all answer options

1 the public service,

2 the industry,

3 handcraftship,
4 trade,
5 to other services,
6 to agriculture, forestry, or fishery,
7 to a different sector,
8 or is it a private household?
98: DK
99: REF
<Variant 2: Respondent is self-employed, freelancer, or working for family ([6.2] = 5|6|9)>
<1. loop: **Does your company belong to**
2. loop or following: **Did your company belong to>...**
INT: Read all answer options
2 the industry,
3 handcraftship,
4 trade,
5 to other services,
6 to agriculture, forestry, or fishery,
7 to a different sector,
8 or is it a private household?
98: DK
99: REF
[numeric, 2-digits]
Programmer: Save current time

[6.18] business employees:
<Variant 1: Public service ([6.17] = 1) or ([6.2] = 3)>
How many persons <*1. loop:***are** *2. loop or following:* **were> employed in the local office in which you work?**
<Variant 2: Not public service and not self-employed ([6.17] != 1 and [6.2 != 5,6,7])>
How many persons <*1. Schleife:***are** *2. loop or following:***were> employed in the company in which you work?**
What is meant here is the number of employed persons in the local office, that is excluding branch offices, which the company may have elsewhere?
Filter: Only for blue- and white-collar workers, soldiers, self-employed farmers, freelancers, helping family members, and DK, REF at 6.2, i.e. 6.2 = 1,2,3,4,8,9,97,98,99
Interviewer: For temporary workers this question is about the temporary employment company (lender).

Number: _ _ _ _ _ _ (6-digit)
999998: DK

999999: REF

[numeric, 6-digits]

Programmer: Save current time

[6.19] Employment-Tool

Programmer: Write the following variables in a database Berufe2r containing these column names to submit the data to R:

- interviewernummer (assigned by NIPO)
- beruflicheTaetigkeit1 (question 6.3): 250 characters
- beruflicheTaetigkeit2 (question 6.4, we expect that the answer from 6.3 is usually saved in this variable (see reference for question 6.4): 250 characters
- beruflicheTaetigkeit3 (question 6.5): 250 characters
- beruflicheTaetigkeit4 (question 6.6): 250 characters
- beruflicheStellung (question 6.2): integer 2-digits
- differenzierteBeruflicheStellung (comprises the following questions, no seperate column needed)
    - beruflicheStellungArbeiter (question 6.7): integer 1- digit
    - beruflicheStellungAngestellte (question 6.8): integer 1- digit
    - beruflicheStellungSoldat (question 6.9): integer 1- digit
    - beruflicheStellungBeamte (question 6.10): integer 1- digit
    - beschaeftigteSelbststaendige (question 6.13): integer 3- digits
- technischerDienst (question 6.11): integer 1- digit
- fuehrungsaufgaben (question 6.14): integer 1- digit
- anzahlArbeitskraefte (question 6.15): integer 4-digits
- ueblicherweiseErforderlicheAusbildung (question 6.16): integer 1-digit
- branche (question 6.17): interger 2-digits
- beschaeftigteImBetrieb (question 6.18): integer 6- digits
- hoechsterAusbildungsabschluss (question 4.2): integer 2- digits
- hoechsterAusbildungsabschlussFreitext (question 4.3): 250 characters

The R-algorithm calculates from this data a list of suggested occupations and their corresponding correctness probabilities. Three databases are needed to return these suggestions:

Berufe2n contains the following columns:

- interviewnummer (same as above)
- KldB2010: 5 digits, e.g., 92133 or 27312 (0 possible at first position)
- DKZ: 8 digits, e.g. 92133100 or … (0 possible at first position)
- correctnessProb: floating-point number (at least 5 post decimal positions), e.g. 0.4329982
- ALWAfrequencies: integer, e.g.: 0 oder 500
- Berufsbenennungen: 250 characters, unique job titles for each DKZ, e.g.: „Account-Manager/in“, „POS-Manager/in“ or „Helfer/in – Gartenbau“
- Berufsuntergruppe: 200 characters, unique 4-digit KldB-title, e.g.: „Aufsichts- & Führungskr.-Theater-, Film- & Fernsehproduktion“ or „Berufe Textilreinigung“
- jobDescription: 2-600 characters, Job description from the BerufeNET in one sentence or more.

Special cases:

- If no job title is predicted, a single line is returned with special code KldB2010 = “-004“ and DKZ = “-0000004“
- If too many job titles are predicted, a single line is returned with special code KldB2010 = “99996“ und DKZ “99999996“

Berufe2n2 contains the following columns:

- KldB42010: digits
- interviewnummer
- Berufsuntergruppe
- correctnessProb

Special case:

- If no job title is predicted, a single line is returned with special code KldB2010 = “-004“ and DKZ = “-0000004“

Berufe2n3 enthält die folgenden Spalten:

- KldB42010
- interviewnummer
- DKZ
- Berufsbenennungen
- correctnessProb

Special case:

- If no job title is predicted, a single line is returned with special code KldB2010 = “-004“ and DKZ = “-0000004“

While R searches for possible job titles and calculates their correctness probabilities (this takes a few seconds), the interviewer initiates further questioning about the occupation:

We try now to classify your occupation. A database query is made for this purpose. This can take a short moment.

Filter: *Check return values for errors. At least one entry must be in the database for this interviewnummer and the KldB2010 must not be “-4”. If this happens:*

Some error occurred. We continue with a different question.

-> Save error indicator and continue with question 6.27

Programmer: Save current time

Depending on the number of suggested DKZ categories, three different entry points are possible:

<Variant 1: Only one category suggested -> question 6.23a>

<Variant 2: Several categories suggested (2-250 job titles) -> question 6.23b>

<Variant 3: Large number of categories suggested (> 250) -> question 6.24>

[6.20] deleted

[6.21] deleted

[6.22] deleted

[6.23a] One category suggested

To be implemented with the database "Berufe2n"

Filter 1: *If there is only one DKZ is in the database "Berufe2n" (equivalently: there are two records in the database, one valid KldB and the code 99996, but never the code 99995)*

Filter 2: *For this DKZ is jobDescription != NA" (or NULL)*

Ist he following job description correct for your occupation?

<read suggested jobDescription >

Filter 2: *For this DKZ is jobDescription == „NA" (or NULL)*

<*1. loop:***Are** *2. loop or following:* **Were> *you employed as <suggested job title>?>***

*** 1: Yes -> continue with 6.26 (Employment-Tool completed)

*** 2: No -> continue with 6.26

*** 3: Partly (Ask for explanation) -> continue with 6.26 (Employment-Tool completed)

*** 6 Further information _____

[Int.: Please insert when the target person provides further information.

*** 8 DK -> continue with 6.27

*** 9 REF -> continue with 6.27

Programmer: Save the suggested DKZ in some variable

Programmer: Save current time

6.23b [6.23b] Several categories suggested

To be implemented with the database "Berufe2n"

Filter*: Skip 6.23b, if database contains only a single line with KldB2010 = 99994*

Reason: If more than 250 DKZ categories are in the database and the second most probable DKZ has correctnessProb > 0.05, skip this question and continue with question 6.24a. Berufe2n contains in such cases only a single line with KldB2010 = 99994

<*1. loop:***Are** *2. loop or following:* **Were> you employed in one of the following occupations?**

Interviewer: Please read all answers. Use filler words if necessary.

Consideration: Out of the suggested job titles we want to have a high bandwidth of probable occupations within a few answer options. Therefore, the following rules hold for the subsequent answer options:

- *(up to) 5 job titles with highest correctnessProb are listed.*
- *but not more than 2 job titles are allowed to have the same KldB2010*
    - *with the exception that almost all job titles have the same KldB2010 and otherwise no 5 occupations are available (Example: 3 job titles have KldB 91384 and the final job title has KldB 94512. All job titles are shown.)*
- *Occupations with identical KldB are shown next to each other.*

1: <job title 1> -> continue with 6.26

2: <job title 2> -> continue with 6.26

3: <job title 3> -> continue with 6.26

4: <job title 4> -> continue with 6.26

5: <job title 5> -> continue with 6.26

99995: or do you work in a similar occupation? (10% probability that R appends this answer option, otherwise not shown) -> continue with 6.24a

99996: or do you work in a different occupation? -> continue with 6.24a

*** 8 DK -> continue with 6.27

*** 9 REF -> continue with 6.27

Programmer: At most 5 "Berufsbenennungen" from the database "Berufe2n" are shown, ordered by the variable "correctnessProb" decreasingly. Save the corresponding 8-digit "DKZ".

Programmer: Save current time

[6.24a] Large number of categories suggested/General question

To be implemented with the database "Berufe2n2"

***Filter**: (If 6.23b = 99996 OR 6.23b = 99995) AND (More than 5 job titles are in the database "Berufe2n" (first table))*

OR

First table [6.23b] was skipped with code 99994

Which of the following fields matches your occupational activity?

Interviewer: Please read all answers. Use filler words if necessary.

Consideration: For the suggested job titles (without excluded ones) are (up to) 5 occupational sub-groups to be shown. Numerous job titles belong to each occupational sub-group (4 digits from the KldB). Order the display by correctness probability in decreasing order. Correctness probabilities are to be calculated as a sum over all corresponding job titles.

1 <official occupational sub-group title> -> continue with 6.24b

2 <official occupational sub-group title> -> continue with 6.24b

3 <official occupational sub-group title> -> continue with 6.24b

4 <official occupational sub-group title> -> continue with 6.24b

5 <official occupational sub-group title> -> continue with 6.24b

6 <official occupational sub-group title> -> continue with 6.24b

7 <official occupational sub-group title> -> continue with 6.24b

9996: or do you work in a different occupation? -> continue with 6.25

*** 8 DK -> continue with 6.27

*** 9 REF -> continue with 6.27

Programmer: At most 7 “Berufsuntergruppe”n from the database “Berufe2n2” are shown, ordered by the variable “correctnessProb” decreasingly. Save the corresponding 4-digit “KldB42010”.

Programmer: Save current time

6.24b [6.24b] Detailed question on the selected occupational sub-group

To be implemented with the database “Berufe2n3“

***Filter**: 6.24a != 9996 (In 6.24a an occupational sub-group was chosen and a 4-digit code is saved in the database) AND at least 2 occupational sub-groups (equivalently: 3 lines) are in the database Berufe2n2 (2. table)*

Please choose the most adequate occupation.

Interviewer: Please read all answers. Use filler words if necessary.

Consideration: Out of the suggested job titles from the chosen occupational sub-group we want to identify the correct KldB, not necessarily the correct DKZ. The following rules hold for the subsequent job titles:

- *(Up to) 5 job titles from the chosen occupational sub-group (KldB, 4-digits) are listed. Hereby*
    - *Occupations from all occupational types (5-digits KldB) should be listed.*
    - *From each occupational type are only the most probable job titles shown*

1: <job title 1> -> continue with 6.26

2: <job title 2> -> continue with 6.26

3: <job title 3> -> continue with 6.26

4: <job title 4> -> continue with 6.26

5: <job title 5> -> continue with 6.26

99996: or do you work in a different occupation? -> continue with 6.26

*** 8 DK -> continue with 6.27

*** 9 REF -> continue with 6.27

Programmer: job titles are only shown if the KldB42010 is identical with the answer from 6.24a, ordered decreasingly by correctnessProb. Save DKZ.

Programmer: Save current time

[6.25] deleted

[6.26] Question to the interviewer: Was it difficult for the respondent to decide for an occupation?

1: No abnormality

2: Slight hesitation

3: Secondlong thinking

4: Loud thinking or inquiries

8: not applicable

Programmer: Save current time

[Occupation coding is completed. The following questions serve to capture further activities retrospectively. This is relevant for occupation coding if the respondent has no job currently and the last occupation is to be coded.]

[6.27] Begin

<Variant 1 (gainfully employed): 6.1 = 1>

And since when have you been active

<*only if [6.4] is not empty:* **as [6.4]>**

without interruption

<*if [6.2] = 1,2,3,4,10,97:* **for the same employer**

if [6.2] = 5,6,7: **self-employed**

*if [6.2] = 8,9:***as <[6.2]>>?**

Interviewer: For temporary workers this question is about the temporary employment company (lender).

INT: If the target person remembers the seasons only, please insert the following numbers in the field „month“:

21: Beginning of year / winter

24: Spring / Easter

27: Midyear / Summer

30: Autumn / Fall

32: End of year

A: Month: _ _ (98 DK, 99 REF)

[numeric, 2-digits]

B: Jahr: _ _ _ _ (9998 DK, 9999 REF)

[numeric, 4-digits]

[other questions about this job and other part time jobs]

[6.80] Test: Current spell started before January 2010

If current spell started before January 2010(=[6.27] or [6.61] or [6.71])

Continue with 7

If current spell started after December 2009 (=[6.27] or [6.61] or [6.71]) and number of loops < 3:

Continue with [6.1] next loop

if number of loops = 3: continue with 7

[More questions follow that are not job-related]

German Questionnaire

[4.2.] Welchen höchsten beruflichen Ausbildungsabschluss haben Sie?

INT: Antwortvorgaben vorlesen. Bei Abschlüssen, die im Ausland erworben wurden, einordnen lassen: Was hätte diesem Abschluss in Deutschland ungefähr entsprochen?

1: Abschluss einer Lehre

2: Abschluss einer Berufsfachschule/ Handelsschule/ Schule des Gesundheitswesens

3: Meister-, Techniker-, Fachwirt

4: Fachhochschulabschluss

5: Hochschulabschluss

6: Anderer Ausbildungsabschluss und zwar _____ [offen]

*** 95 kein beruflicher Ausbildungsabschluss-> weiter mit 4.4

*** 96 derzeit noch in beruflicher Ausbildung-> weiter mit 4.4

*** 98 weiß nicht-> weiter mit 4.4

*** 99 keine Angabe-> weiter mit 4.4

[numerisch, zweistellig]

[4.3.] Beruf des höchsten Ausbildungsabschlusses

<Variante 1: 4.2 = 1| 2 | 3 | 6?>

In welchem Beruf genau haben Sie diese Ausbildung gemacht?

Interviewer: Bitte genauere Berufsbezeichnung nachfragen; nicht Mechaniker, sondern Fein- oder Kfz-Mechaniker; nicht Lehrer, sondern z.B. Gymnasiallehrer für Geschichte.
Falls mehrere gleichwertige Abschlüsse, bitte den letzten Abschluss erfassen.

<Variante 2: 4.2 = 4 | 5>

Welches Hauptfach haben Sie studiert?

Interviewer: Bitte Studienfach nennen lassen, genau erfassen!

_____ [falls zwei Hauptfächer, zweites ebenfalls erfassen] _____

*** 8 weiß nicht

*** 9 keine Angabe

[offen]

[Other questions not relevant for occupation coding]

Erwerbsbiografie seit 01.01.2010:

[6.1] Aktueller und frühere Jobs

Retrospektive Erfassung von bis zu 3 Aktivitäten, wobei Beginn der zweiten Aktivität vor der ersten und Beginn der dritten Aktivität vor der zweiten liegen muss. Die Enden der Aktivitäten werden nicht gegeneinander geprüft.

Maximal 3 Schleifendurchgänge

<*1. Schleife:* **Im Folgenden möchte ich gerne mit Ihnen über Ihren beruflichen Werdegang in den letzten Jahren sprechen. Wir möchten z.B. wissen, ob Sie erwerbstätig, in Ausbildung oder arbeitslos gemeldet waren oder etwas anderes gemacht haben. Es ist wichtig, dass Sie jede Aktivität einzeln angeben, auch wenn sie nur kurz gedauert hat. Falls mehrere Aktivitäten zeitgleich stattgefunden haben, nennen Sie bitte Ihre Haupt-Erwerbstätigkeit.**

Beginnen wir mit heute: Was von dem Folgenden trifft zu? Sind Sie derzeit…

2. oder folgende Schleifen: **Was haben Sie vor <Datum Beginn aus [6.27, [6.61] oder [6.71] der vorigen Schleife]>, also vor**

<*wenn [6.1] im vorherigen Spell = 1:* **der Erwerbstätigkeit, über die**

*wenn [6.1] im vorherigen Spell = 2:***der Arbeitslosigkeit, über die**

wenn [6.1] im vorherigen Spell = 3: **dem Schulbesuch, über den**

wenn [6.1] im vorherigen Spell = 4: **der beruflichen Aus- oder Weiterbildung, Lehre oder dem Studium, worüber**

*wenn [6.70] im vorherigen Spell = 1:***dem Wehr- oder Zivildienst, Bundesfreiwilligendienst oder Ähnlichem, worüber**

wenn [6.70] im vorherigen Spell = 2: **Ihrer Tätigkeit als <***weiblich:* **Hausfrau** *männlich:* **Hausmann>, worüber**

*wenn [6.70] im vorherigen Spell= 3:***<***nur wenn weiblich:***dem Mutterschutz,>dem Erziehungsurlaub oder der Elternzeit, worüber**

*wenn [6.70] im vorherigen Spell= 4:***der Rente, Pension oder Vorruhestand, worüber**

*wenn [6.70] im vorherigen Spell= 5:***der Phase, über die>**

wir gerade gesprochen haben, gemacht?

Was von dem Folgenden trifft zu? Waren Sie davor…>

****INT.: Wenn es parallele Phasen gab, sticht die Erwerbstätigkeit. Bei mehreren Erwerbstätigkeiten sticht diejenige mit den meisten Arbeitsstunden. Unter Erwerbstätigkeiten wollen wir auch Nebentätigkeiten von Schülern, Hausfrauen und Rentnern erfassen. Auch sogenannte Mini-Jobs (auch 400 bzw. 450€-Jobs genannt) zählen dazu.*

1: erwerbstätig, damit meinen wir auch Nebentätigkeiten oder Mini-Jobs ->weiter mit 6.2

2: arbeitslos gemeldet, damit meinen wir auch Zeiten, in denen Sie an einer Maßnahme oder einem Programm der Arbeitsagentur oder des Jobcenters <*1. Schleife:* teilnehmen *2. oder folgende Schleifen:* teilgenommen haben>, ->weiter mit [6.61]

3: Schüler/in ->weiter mit [6.61]

4: in einer beruflichen Aus- oder Weiterbildung, einer Lehre oder einem Studium ->weiter mit [6.61]

5: <*1. Schleife:* oder machen Sie etwas anderes *2. oder folgende Schleifen:* oder haben Sie etwas anderes gemacht>? ->weiter mit [6.70]

***8: weiß nicht ->weiter mit 8

***9: keine Angabe ->weiter mit 8

[numerisch, einstellig]

[6.2] Stellung im Beruf

<Filter: Spell ist aktuelle Erwerbstätigkeit (erster Durchlauf)>

Und wie ist da Ihre derzeitige berufliche Stellung? Sind Sie ...

<Filter: Spell ist nicht die aktuelle Erwerbstätigkeit >

Und wie war da zuletzt Ihre berufliche Stellung? Waren Sie ...

Interviewer: Antwortvorgaben bitte vollständig vorlesen.

<Antwortvorgaben geschlechtsspezifisch einblenden>

1: Arbeiter/in

2: Angestellte/r

3: Berufssoldat/in oder Zeitsoldat/in

4: Beamte/r oder Richter/in

5: Selbständige/r in einem freien Beruf also
z.B. Arzt/Ärztin, Rechtsanwalt/Rechtsanwältin oder Architekt/in

6: Selbständige/r in Handel, Gewerbe, Industrie, Dienstleistung

7: Selbständige/r Landwirt/in

8: Freier Mitarbeiter/Freie Mitarbeiterin

9: Mithelfende/r Familienangehörige/r oder

10: <*1. Schleife:* handelt *2. Schleife oder folgende:* handelte> es sich dabei um einen Mini-Job, früher 400€-Job, mittlerweile 450€-Job genannt? weiter mit [6.18]

97: ZP kann sich nicht zwischen „Arbeiter“ und „Angestellter“ entscheiden

98: weiß nicht

99: keine Angabe

[numerisch, zweistellig]

[6.3] Und sagen Sie mir bitte, welche berufliche Tätigkeit Sie da <1. Schleife: ausüben 2. Schleife oder folgende: ausübten>?

Filter: nur zeitlich letzte Erwerbstätigkeit

*Filter: Die folgenden Fragen [6.3]-[6.26] werden **nur einmal** gestellt, **und zwar nur für die zeitlich letzte Erwerbstätigkeit**. D.h. wenn der Befragte aktuell erwerbstätig ist (6.1 == 1) oder, wenn nicht aktuell erwerbstätig, dann, wenn es sich um die zeitlich letzte Erwerbstätigkeit handelt. Wenn sie einmal gestellt wurden, werden sie nicht mehr erhoben.*

Für alle anderen kommt Frage [6.27] (Beginn des Spells) als nächstes.

Interviewer: Bitte auf Rechtschreibung achten, ggf. buchstabieren lassen! Genaue Berufsbezeichnung nachfragen. Bitte z.B. nicht „Mechaniker“, sondern „Fein- oder Kfz-Mechaniker“; nicht „Lehrer“, sondern „Gymnasiallehrer“.

Falls es sich um Zeitarbeit handelt, nach der überwiegenden beruflichen Tätigkeit fragen!

Programmierer: Liste mit den allgemeinen Begriffen hinterlegen und in Abhängigkeit davon steuern, 6.4 gestellt werden muss.

*** 99 k.A.

[offen]

Programmierer: Bitte Zeitmessung vornehmen.

[6.4] Geben Sie mir bitte die genaue Tätigkeitsbezeichnung an. Also z. B. nicht <kaufmännischer Angestellter, sondern: Speditionskaufmann>, nicht <Arbeiter, sondern: Maschinenschlosser>.

Filter: nur zeitlich letzte Erwerbstätigkeit und wenn zu allgemeine Angabe in [6.3]. Zu allgemeine Angaben sind solche, die in einer Liste des Umfrageinstituts entsprechend markiert sind, z.B. Verkäufer, Abteilungsleiter, Sachbearbeiter. 4.3% der beschäftigten Personen aus unserer Umfrage haben bei 6.3. eine zu allgemeine Angabe gemacht.

Programmierer:

Steuerung in Abhängigkeit vom Geschlecht: <kaufmännische Angestellte, sondern: Speditionskauffrau>, <Arbeiterin, sondern: Maschinenschlosserin>

Programmierer:

Wenn diese zusätzliche Nachfrage nicht gestellt wird, soll trotzdem die Antwort aus [6.3] als Antwort von [6.4] gespeichert werden. Für die Übergabe in R und spätere Einblendungen kann und muss dann die Angabe in [6.4] verwendet werden.

*** 99 k.A.
[offen]
Programmierer: Bitte Zeitmessung vornehmen.

[6.5] Bitte beschreiben Sie mir diese berufliche Tätigkeit genau.
Filter: nur zeitlich letzte Erwerbstätigkeit

*** 99 k.A.
[offen]
Programmierer: Bitte Zeitmessung vornehmen.

[6.6] Hat dieser Beruf noch einen besonderen Namen?
Filter: nur zeitlich letzte Erwerbstätigkeit
Filter: Frage nur stellen, falls Frage [6.4] nicht gestellt wurde.

Ja, und zwar: ----------
Nein
*** 99 k.A.
[offen]
Programmierer: Bitte Zeitmessung vornehmen.

[6.7] Stellung Beruf: Arbeiter/in
Filter: [6.2] = 1
<*1. Schleife:* **Und was sind Sie genau? Sind Sie . . .**
2. Schleife oder folgende: **Und was waren Sie genau? Waren Sie . . .>?**

<Antwortvorgaben geschlechtsspezifisch einblenden>
1 ungelernte/r Arbeiter/in
2 angelernte/r Arbeiter/in
3 Facharbeiter/in
4 Vorarbeiter/in, Kolonnenführer/in oder
5 Meister/in, Polier/in, Brigadier/in?
8: weiß nicht
9: keine Angabe
[numerisch, einstellig]
Filteranweisung: -> Weiter mit [6.14]
Programmierer: Bitte Zeitmessung vornehmen.

[6.8] Stellung Beruf: Angestellte/r

Filter: [6.2] = 2 | [6.2] = 97

Und was sind Sie genau? Sind Sie Angestellte/r mit . . .

<*1. Schleife:* **Und was sind Sie genau? Sind Sie Angestellte/r mit. . .**

2. Schleife oder folgende: **Und was waren Sie genau? Waren Sie Angestellte/r mit. . .>?**

<Antwortvorgaben geschlechtsspezifisch einblenden>

<Wenn der Befragte sich nicht zwischen Arbeiter und Angestellter entscheiden kann ([6.2] = 97): keine Beispiele nennen>

1 ausführender Tätigkeit nach allgemeiner Anweisung (z.B. Verkäufer/in, Datentypist/-in, Sekretariatsassistent/-in, Pflegehelfer/-in)

2 qualifizierter Tätigkeit, die Sie nach Anweisung erledigen (z.B. Sachbearbeiter/-in, Buchhalter/in, technische/r Zeichner/in)

3 eigenständiger Leistung in verantwortlicher Tätigkeit oder mit Fachverantwortung für Personal (z.B. wissenschaftliche/r Mitarbeiter/in, Prokurist/in, Abteilungsleiter/in oder Meister/in im Angestelltenverhältnis)

4 umfassenden Führungsaufgaben und Entscheidungsbefugnissen(z.B. Direktor/in oder Geschäftsführer/in, Mitglied des Vorstands)

*** 8: weiß nicht

*** 9: keine Angabe

[numerisch, einstellig]

Filteranweisung: -> Weiter mit [6.14]

Programmierer: Bitte Zeitmessung vornehmen.

[6.9] Stellung Beruf: Berufssoldat/in oder Zeitsoldat/in

Filter: [6.2] = 3

<1*. Schleife:* **Sind Sie …**

2. Schleife oder folgende: **Waren Sie …>**

<Antwortvorgaben geschlechtsspezifisch einblenden>

1 Träger eines Mannschaftsdienstgrades

2 Unteroffizier ohne Portepee (Unteroffizier, Stabsunteroffizier)

3 Unteroffizier mit Portepee (Feldwebel, Oberfeldwebel usw.)

4 Offizier (Leutnant, Hauptmann)

5 Stabsoffizier (ab Major)

8: weiß nicht

9: keine Angabe

[numerisch, einstellig]

Filteranweisung: -> Weiter mit [6.14]
Programmierer: Bitte Zeitmessung vornehmen.

[6.10] Stellung Beruf: Beamter/Beamtin
Filter: [6.2] = 4

<1. Schleife: **Sind**
2. Schleife oder folgende: **Waren>Sie Beamter<r> im einfachen, mittleren, gehobenen oder höheren Dienst?**

1 im einfachen Dienst -> Weiter mit [6.14]
2 im mittleren Dienst
3 im gehobenen Dienst
4 im höheren Dienst
8: weiß nicht
9: keine Angabe
[numerisch, einstellig]
Programmierer: Bitte Zeitmessung vornehmen.

[6.11] Technischer Dienst
Filter: [6.10] = 2, 3, 4, 8, 9:
<1. Schleife: **Sind**
2. Schleife oder folgende: **Waren>Sie im technischen Dienst?**
1: Ja
2: Nein
8: weiß nicht
9: keine Angabe
[numerisch, einstellig]
Filteranweisung: -> Weiter mit [6.14]
Programmierer: Bitte Zeitmessung vornehmen.

[6.12] Stellung Beruf: Selbständige/r Landwirt/in
Filter: [6.2] = 7
Und wie viele Hektar hat die landwirtschaftliche Fläche, die Sie
<1. Schleife: **bewirtschaften**
2. Schleife oder folgende: **bewirtschafteten>?**

Hektar: _ _ _ _ _ (5-stellig)
99998: weiß nicht

99999: keine Angabe

[numerisch, fünfstellig]

Filteranweisung: -> Weiter mit [6.13]

Programmierer: Bitte Zeitmessung vornehmen.

[6.13] Stellung Beruf: Selbständige/r

Filter: [6.2] = 5 | 6 | 7

Und wie viele Mitarbeiter

<1. Schleife: **haben**

2. Schleife oder folgende: **hatten> Sie?**

Mitarbeiter _ _ _ (3-stellig)

996 Mehr als 99 Mitarbeiter

998: weiß nicht

999: keine Angabe

[numerisch, dreistellig]

Filteranweisung: -> Weiter mit [6.16]

Programmierer: Bitte Zeitmessung vornehmen.

[6.14] Führungsaufgaben

Filter: [6.2] = 1|2|3|4|8|9|97|98|99 (nicht selbstständig)

<1. Schleife: **Gehört**

*2. Schleife oder folgende:***Gehörte>**

es zu Ihren beruflichen Aufgaben, die Arbeit anderer Arbeitskräfte zu beaufsichtigen oder ihnen zu sagen, was sie tun müssen?

INT: Hier sind nur Arbeitskräfte gemeint. Wenn ein Lehrer Schüler beaufsichtigt oder eine Kindergärtnerin Kinder, dann zählen diese nicht dazu. Wenn aber ein Schuldirektor andere Lehrer beaufsichtigt, zählt das natürlich schon.

1: Ja

2: Nein ->weiter mit [6.16]

8: weiß nicht-> 6.19

9: keine Angabe-> 6.19

Programmierer: Bitte Zeitmessung vornehmen.

[6.15] Anzahl Arbeitskräfte:

Filter: [6.14] = 1

Wie viele andere Arbeitskräfte

<1. Schleife: **beaufsichtigen**

*2. Schleife oder folgende:***beaufsichtigten>Sie direkt?**

INT: Hier sind nur Arbeitskräfte gemeint. Wenn ein Lehrer Schüler beaufsichtigt oder eine Kindergärtnerin Kinder, dann zählen diese nicht dazu. Wenn aber ein Schuldirektor andere Lehrer beaufsichtigt, zählt das natürlich schon.

Anzahl: _ _ _ _ (4-stellig)

9998: weiß nicht

9999: keine Angabe

[numerisch, vierstellig]

Programmierer: Bitte Zeitmessung vornehmen.

[6.16] Üblicherweise erforderliche Ausbildung

Welche Art von Ausbildung ist für die Ausübung Ihrer Tätigkeit als <Tätigkeit aus [6.4] einblenden> in der Regel erforderlich? Ist keine Ausbildung erforderlich, ist eine Anlernausbildung, eine abgeschlossene berufliche Ausbildung, eine abgeschlossene Fachschulausbildung, ein Meister- oder Technikerabschluss oder ist ein abgeschlossenes Fachhochschul- oder Hochschulstudium erforderlich?

1: es ist keine Ausbildung erforderlich

2: eine Anlernausbildung

3: eine abgeschlossene berufliche Ausbildung

4: eine abgeschlossene Fachschulausbildung

5: ein Meister- oder Technikerabschluss

6: ein abgeschlossenes Fachhochschul- oder Hochschulstudium

8: weiß nicht

9: keine Angabe

[numerisch, einstellig]

Programmierer: Bitte Zeitmessung vornehmen.

[6.17] Branche

Filter: [6.2] != 3|7|10 (nicht Berufssoldat oder selbstständiger Landwirt)

<Variante 1: Befragter ist Arbeiter, Angestellter, Beamter, freier Mitarbeiter oder kann sich nicht entscheiden ([6.2] = 1|2|4|8|97|98|99)>

<1. Schleife: **Gehört der Betrieb, in dem Sie derzeit arbeiten,**

*2. Schleife oder folgende:***Gehörte der Betrieb, in dem Sie arbeiteten,>...**

INT: Antwortvorgaben bitte vorlesen.

1 zum öffentlichen Dienst,

2 zur Industrie,

3 zum Handwerk,

4 zum Handel,

5 zu sonstigen Dienstleistungen,

6 zur Land- und Forstwirtschaft oder Fischerei,

7 zu einem anderen Bereich oder

8 ist das ein Privathaushalt?

98: weiß nicht

99: keine Angabe

<Variante 2: Spell ist zeitlich aktuellste Erwerbstätigkeit & Befragter ist selbstständig, freiberuflich tätig oder mithelfender Familienangehöriger ([6.2] = 5|6|9)>

<1. Schleife: **Gehört Ihr Betrieb**

2. Schleife oder folgende: **Gehörte Ihr Betrieb>...**

INT: Antwortvorgaben bitte vorlesen.

2 zur Industrie,

3 zum Handwerk,

4 zum Handel,

5 zu sonstigen Dienstleistungen oder

6 zur Land- und Forstwirtschaft oder Fischerei,

7 zu einem anderen Bereich oder

8 ist das ein Privathaushalt?

98: weiß nicht

99: keine Angabe

[numerisch, zweistellig]

Programmierer: Bitte Zeitmessung vornehmen.

[6.18] Beschäftigte im Betrieb:

<Variante 1: Spell ist zeitlich aktuellste Erwerbstätigkeit und (öffentlicher Dienst ([6.17] = 1) oder [6.2] = 3)>

Wie viele Personen <*1. Schleife:***sind in der örtlichen Dienststelle, in der Sie derzeit arbeiten,** *2. Schleife oder folgende:* **waren in der örtlichen Dienststelle, in der Sie arbeiteten,> beschäftigt?**

<Variante 2: Spell ist zeitlich aktuellste Erwerbstätigkeit & nicht öffentlicher Dienst ([6.17] != 1 und [6.2 != 5,6,7] (d.h. nicht an Selbstständige in freien Berufen oder in Handel, Gewerbe, Industrie, Dienstleistung)>

Wie viele Personen <*1. Schleife:***sind in dem Betrieb, in dem Sie derzeit arbeiten,** *2. Schleife oder folgende:***waren in dem Betrieb, in dem Sie arbeiteten,> beschäftigt?**

Gemeint ist hier die Beschäftigtenzahl an der örtlichen Arbeitsstelle, also ohne Zweigstellen usw., die Ihr Betrieb vielleicht noch woanders <*1. Schleife:***hat** *2. Schleife oder folgende:* **hatte>?**

Filter: Nur an abhängig Beschäftigte, Soldaten, selbstständige Landwirte, freie Mitarbeiter, mithelfende Familienangehörige und w.n., k.A. bei 6.2 d.h. 6.2 = 1,2,3,4,8,9, ,97,98,99

Interviewer: Bei Leiharbeit ist hier die Zeitarbeitsfirma (Verleiher) gemeint.

Anzahl: _ _ _ _ _ _ (6-stellig)

999998: weiß nicht

999999: keine Angabe

[numerisch, sechsstellig]

Programmierer: Bitte Zeitmessung vornehmen.

[6.19] Berufe-Tool

Programmierer: Schreibe die folgenden Variablen in eine Datenbank Berufe2r mit den folgenden Spalten zur Übergabe an R:

- interviewernummer (von NIPO zugewiesen)
- beruflicheTaetigkeit1 (Frage 6.3): 250 Zeichen
- beruflicheTaetigkeit2 (Frage 6.4, es wird erwartet, dass in den meisten Fällen die Antwort von 6.3 bereits in 6.4 übertragen wurde (siehe Programmieranweisung dort): 250 Zeichen
- beruflicheTaetigkeit3 (Frage 6.5): 250 Zeichen
- beruflicheTaetigkeit4 (Frage 6.6): 250 Zeichen
- beruflicheStellung (Frage 6.2): integer 2-stellig
- differenzierteBeruflicheStellung (umfasst die folgenden Fragen, aber keine eigene Variable nötig)
    - beruflicheStellungArbeiter (Frage 6.7): integer 1-stellig
    - beruflicheStellungAngestellte (Frage 6.8): integer 1-stellig
    - beruflicheStellungSoldat (Frage 6.9): integer 1-stellig
    - beruflicheStellungBeamte (Frage 6.10): integer 1-stellig
    - beschaeftigteSelbststaendige (Frage 6.13): integer 3-stellig
- technischerDienst (Frage 6.11): integer 1-stellig
- fuehrungsaufgaben (Frage 6.14): integer 1-stellig
- anzahlArbeitskraefte (Frage 6.15): integer 4-stellig
- ueblicherweiseErforderlicheAusbildung (Frage 6.16): integer 1-stellig
- branche (Frage 6.17): integer 2-stellig
- beschaeftigteImBetrieb (Frage 6.18): integer 6-stellig
- hoechsterAusbildungsabschluss (Frage 4.2): integer 2-stellig
- hoechsterAusbildungsabschlussFreitext (Frage 4.3): 250 Zeichen

Mit diesen Angaben berechnet der Algorithmus in R eine Liste möglicher Berufe zusammen mit Korrektheitswahrscheinlichkeiten. Die Rückgabe erfolgt in drei Datenbanken:

Berufe2n enthält die folgenden Spalten:

- interviewnummer (die gleiche Nummer wie oben)
- KldB2010: 5 Zeichen (Ziffern), z.B. 92133 oder 27312 (0 an erster Stelle möglich)
- DKZ: 8 Zeichen (Ziffern), z.B. 92133100 oder … (0 an erster Stelle möglich)
- correctnessProb: Gleitkommazahl (mindestens 5 Nachkommastellen), z.B. 0.4329982
- ALWAfrequencies: integer, z.B.: 0 oder 500
- Berufsbenennungen: 250 characters, eindeutige Berufsbezeichnungen zu jeder DKZ, z.B.: „Account-Manager/in“, „POS-Manager/in“ oder „Helfer/in – Gartenbau“

- Berufsuntergruppe: 200 characters, eindeutige Bezeichnung der KldB auf 4-Steller-Ebene, z.B.: „Aufsichts- & Führungskr.-Theater-, Film- & Fernsehproduktion“ oder „Berufe Textilreinigung“
- jobDescription: 2-600 characters, Beschreibung eines Berufs aus dem BerufeNET in einem oder in mehreren Sätzen

Sonderfälle:

- Wenn kein einziger Beruf gefunden wird, wird eine Zeile zurückgegeben mit dem Sondercode KldB2010 = „-004“ und DKZ = „-0000004“
- Wenn sehr (zu) viele Berufe gefunden wurden, wird eine Zeile zurückgegeben mit dem Sondercode KldB2010 = „99996“ und DKZ „99999996“

Berufe2n2 enthält die folgenden Spalten:

- KldB42010: 4 Zeichen (Ziffern)
- interviewnummer
- Berufsuntergruppe
- correctnessProb

Sonderfall:

- Wenn kein einziger Beruf gefunden wird, wird eine Zeile zurückgegeben mit dem Sondercode KldB2010 = „-004“ und DKZ = „-0000004“

Berufe2n3 enthält die folgenden Spalten:

- KldB42010
- interviewnummer
- DKZ
- Berufsbenennungen
- correctnessProb

Sonderfall:

- Wenn kein einziger Beruf gefunden wird, wird eine Zeile zurückgegeben mit dem Sondercode KldB2010 = „-004“ und DKZ = „-0000004“

Während R mögliche Berufe und Korrektheitswahrscheinlichkeiten berechnet (dauert einige Sekunden), leitet der Interviewer die weiteren Nachfragen zum Beruf ein:

Wir versuchen nun Ihren Beruf genauer einzuordnen. Zu diesem Zweck erfolgt eine Datenbankabfrage. Dies kann einen kurzen Moment dauern.

Filter: *Rückgabe auf Fehler prüfen. Zur Interviewnummer muss mindestens ein Eintrag in der Datenbank stehen und die KldB2010 davon darf nicht „-4“ sein. Falls dies auftritt:*

Da ist uns wohl ein Fehler unterlaufen. Wir machen dann mit einer anderen Frage weiter.

-> Eintreten des Fehlers abspeichern und weiter mit Frage 6.27

Programmierer: Bitte Zeitmessung vornehmen.

Je nach Anzahl der vorgeschlagenen Berufe, sind drei Einstiege möglich:

<Variante 1: Nur ein Beruf vorgeschlagen -> Frage 6.23a>

<Variante 2: Mehrere Berufe vorgeschlagen (2-250 Berufe) -> Frage 6.23b>

<Variante 3: Sehr viele Berufe vorgeschlagen (> 250) -> Frage 6.24>

[6.20] gestrichen

[6.21] gestrichen

[6.22] gestrichen

[6.23a] Ein Beruf vorgeschlagen

Umsetzung erfolgt mittels der Datenbank „Berufe2n“

Filter 1: *Wenn nur ein einziger Beruf in der Datenbank „Berufe2n“ steht (äquivalent: es stehen zwei Zeilen in der Datenbank, eine gültige KldB und der Code 99996, aber nie der Code 99995)*

Filter 2: *für diesen Beruf ist jobDescription != „NA“ (bzw NULL)*

Trifft die folgende Beschreibung für Ihren Beruf zu?

<jobDescription vorlesen>

Filter 2: *für diesen Beruf ist jobDescription == „NA“ (bzw NULL)*

<*1. Schleife:***Sind** *2. Schleife oder folgende:* **Waren> *Sie als <Berufsbenennungen> tätig?>***

*** 1: Ja -> weiter mit 6.26 (Berufe-Tool abgeschlossen)

*** 2: Nein -> weiter mit 6.26

*** 3: Teilweise (Bitte erläutern lassen) -> weiter mit 6.26 (Berufe-Tool abgeschlossen)

*** 6 Weitere Informationen _____

[Int.: Bitte eintragen, wenn die ZP von sich aus weitere Informationen gibt.

*** 8 w.n. -> weiter mit 6.27

*** 9 k.A. -> weiter mit 6.27

Programmierer: In einer weiteren Variablen bitte die abgefragte DKZ abspeichern.

Programmierer: Bitte Zeitmessung vornehmen.

6.23b [6.23b] Mehrere sinnvolle Berufe vorgeschlagen

Umsetzung erfolgt mittels der Datenbank „Berufe2n“

Filter*: Überspringe 6.23b, wenn Datenbank nur eine Zeile enthält mit KldB2010 = 99994*

Überlegung dabei: Wenn mehr als 250 Berufe in der Datenbank stehen und der zweitwahrscheinlichste Beruf correctnessProb > 0.05 hat, diese Frage überspringen und direkt weitermachen mit Frage 6.24a. In diesen Fällen enthält Berufe2n nur eine Zeile mit KldB2010 = 99994

<*1. Schleife:***Sind** *2. Schleife oder folgende:* **Waren> Sie in einem der folgenden Berufe tätig?**

Interviewer: Bitte alle Antworten vorlesen. Ggf. Füllwörter zur Beschreibung nutzen.

Überlegung: Von den vorgeschlagenen Berufen wollen wir in wenigen Antwortmöglichkeiten eine möglichst große Bandbreite an wahrscheinlichen Berufen auflisten. Deshalb gelten die folgenden Regeln für die nachfolgend gelisteten Berufe:

- *es werden die (bis zu) 5 Berufe mit größter correctnessProb angezeigt,*
- *aber nicht mehr als 2 Berufe dürfen die gleiche KldB2010 haben*
    - *außer, wenn die Berufe fast alle die gleiche KldB haben und ansonsten keine 5 Berufe vorhanden sind (Beispiel: 3 Berufe haben KldB 91384 und der letzte Beruf hat KldB 94512. Alle Berufe werden angezeigt.)*
- *Berufe mit der gleichen KldB werden untereinander angezeigt*

1: <Berufsbenennungen 1> -> weiter mit 6.26

2: <Berufsbenennungen 2> -> weiter mit 6.26

3: <Berufsbenennungen 3> -> weiter mit 6.26

4: <Berufsbenennungen 4> -> weiter mit 6.26

5: <Berufsbenennungen 5> -> weiter mit 6.26

99995: oder handelt es sich um einen ähnlichen Beruf? (experimentell für 10% der Fälle in R umgesetzt, wird nicht immer angezeigt) -> weiter mit 6.24a

99996: oder handelt es sich um einen anderen Beruf? -> weiter mit 6.24a

*** 8 w.n. -> weiter mit 6.27

*** 9 k.A. -> weiter mit 6.27

Programmierer: Angezeigt werden die maximal 5 „Berufsbenennungen“ aus der Datenbank „Berufe2n“ absteigend sortiert nach der Variablen „correctnessProb“. Abgespeichert wird die zugehörige 8-stellige „DKZ“.

Programmierer: Bitte Zeitmessung vornehmen.

[6.24a] Sehr viele Berufe vorgeschlagen/Grobe Abfrage

Umsetzung erfolgt mittels der Datenbank „Berufe2n2“

***Filter**: (Wenn 6.23b = 99996 ODER 6.23b = 99995) UND (Es stehen mehr als 5 Berufe in der Datenbank „Berufe2n“ (erste Tabelle))*

ODER

Erste Tabelle [6.23b] wurde mit Code 99994 übersprungen

Welchem der folgenden Bereiche <*1. Schleife:* **ist** *2. Schleife oder folgende:* **war> Ihre Tätigkeit zuzuordnen?**

Interviewer: Bitte alle Antworten vorlesen. Ggf. Füllwörter zur Beschreibung nutzen.

Überlegung: Von den vorgeschlagenen Berufen (ohne ausgeschlossene) sollen (bis zu) 5 Berufsuntergruppen angezeigt werden. Zu einer Berufsuntergruppe (die ersten 4 Ziffern der KldB) gehören immer zahlreiche Berufe. Die Anzeige soll geordnet werden mit der wahrscheinlichsten Berufsuntergruppe zuerst und der fünftwahrscheinlichsten zuletzt. Die Wahrscheinlichkeit berechnet sich als Summe über die correctnessProb der zugehörigen Berufe.

1 <offizielle Bezeichnung Berufsuntergruppe> -> weiter mit 6.24b

2 <offizielle Bezeichnung Berufsuntergruppe> -> weiter mit 6.24b

3 <offizielle Bezeichnung Berufsuntergruppe> -> weiter mit 6.24b

4 <offizielle Bezeichnung Berufsuntergruppe> -> weiter mit 6.24b

5 <offizielle Bezeichnung Berufsuntergruppe> -> weiter mit 6.24b

9996 oder handelt es sich um einen anderen Beruf? -> weiter mit 6.25

*** 8 w.n. -> weiter mit 6.27

*** 9 k.A. -> weiter mit 6.27

Programmierer: Angezeigt werden die maximal 7 „Berufsuntergruppe“n aus der Datenbank „Berufe2n2“ absteigend sortiert nach der Variablen „correctnessProb“. Abgespeichert wird die zugehörige 4-stellige „KldB42010“.

Programmierer: Bitte Zeitmessung vornehmen.

6.24b [6.24b] Detailabfrage zur ausgewählten Berufsuntergruppe

Umsetzung erfolgt mittels der Datenbank „Berufe2n3“

***Filter**: 6.24a != 9996 (In 6.24a wurde eine Berufsuntergruppe ausgewählt und ein 4-stelliger Code in der Datenbank abgespeichert) UND es stehen mindestens 2 Berufsuntergruppen, also 3 Zeilen, in der Datenbank Berufe2n2 (2. Tabelle)*

Wählen Sie bitte den passenden Beruf aus.

Interviewer: Bitte alle Antworten vorlesen. Ggf. Füllwörter zur Beschreibung nutzen.

Überlegung: Von den vorgeschlagenen Berufen aus der gewählten Berufsuntergruppe wollen wir die korrekte KldB identifizieren, nicht notwendigerweise die korrekte DKZ. Deshalb gelten die folgenden Regeln für die nachfolgend gelisteten Berufe:

- *Es werden (bis zu) 5 Berufe aus der gewählten Berufsuntergruppe (KldB, 4-stellig) gelistet. Dabei*
    - *Sollen Berufe aus allen Berufsgattungen (5-stellig) gleichmäßig vertreten sein*
    - *Angezeigt werden aus jeder Berufsgattung jeweils nur die wahrscheinlichsten Berufe*

1: <Berufsbenennungen 1> -> weiter mit 6.26

2: <Berufsbenennungen 2> -> weiter mit 6.26

3: <Berufsbenennungen 3> -> weiter mit 6.26

4: <Berufsbenennungen 4> -> weiter mit 6.26

5: <Berufsbenennungen 5> -> weiter mit 6.26

99996: oder handelt es sich um einen anderen Beruf? -> weiter mit 6.26

*** 8 w.n. -> weiter mit 6.27

*** 9 k.A. -> weiter mit 6.27

Programmierer: Angezeigt werden die Berufsbenennungen, für die die KldB42010 identisch mit der Antwort aus 6.24a ist, absteigend geordnet nach correctnessProb. Abgespeichert wird die DKZ.

Programmierer: Bitte Zeitmessung vornehmen.

[6.25] gestrichen

[6.26] Frage an den Interviewer: Ist es dem Befragten schwer gefallen, sich für einen Beruf zu entscheiden?

1: Nein bzw. keine Auffälligkeiten

2: Leichtes Zögern

3: Sekundenlanges Nachdenken

4: Lautes Überlegen bzw. Nachfragen

8: trifft nicht zu

Programmierer: Bitte Zeitmessung vornehmen.

[Ende Berufskodierung. Die folgenden Fragen dienen der retrospektiven Erfassung weiterer Aktivitäten. Dies ist für die Berufskodierung dann relevant, wenn die befragte Person aktuell keinen Job hat und die zuletzt ausgeübte Tätigkeit kodiert werden soll.]

[6.27] Beginn

<Variante 1 (erwerbstätig): 6.1 = 1>

Und seit wann

<*1. Schleife:***sind**

*2. Schleife oder folgende:***waren> Sie**

<*nur wenn [6.4] nicht leer:* **als [6.4]>**

ohne Unterbrechung

<*wenn [6.2] = 1,2,3,4,10,97:* **beim selben Arbeitgeber beschäftigt**

wenn [6.2] = 5,6,7: **selbständig tätig**

*wenn [6.2] = 8,9:***als <[6.2]> tätig>?**

Interviewer: Bei Leiharbeit ist hier die Zeitarbeitsfirma (Verleiher) gemeint.

INT: Falls die Zielperson sich nur an Jahreszeiten erinnert, bitte nach der ungefähren Jahreszeit fragen und ggf. eine der folgenden Nummern in das Feld „Monat“ eingeben:

21: Jahresanfang/Winter

24: Frühjahr/Ostern

27: Jahresmitte/Sommer

30: Herbst

32: Jahresende

A: Monat: _ _ (98 weiß nicht, 99 keine Angabe)

[numerisch, zweistellig]

B: Jahr: _ _ _ _ (9998 weiß nicht, 9999 keine Angabe)

[numerisch, vierstellig]

[more questions about this job and other part time jobs]

[6.80] Prüfung: Beginn aktueller Spell startete bis 01.2010

Wenn Beginn aktueller Spell (=[6.27] oder [6.61] oder [6.71]) bis einschließlich 01.2010

Weiter mit 7

Wenn Beginn aktueller Spell (=[6.27] oder [6.61] oder [6.71]) nach 01.2010 & Anzahl Durchläufe < 3:

Weiter mit [6.1] nächste Schleife

Wenn Anzahl Schleifendurchläufe = 3: Weiter 7

[More questions follow that are not job-related]

Part C: Instructions for Validation

In occupation coding, respondents describe their job with their own words. The coders task is then to infer the underlying bundle of job activities that is frequently not well described. The next step is to match these activities to one of several categories from some occupational classification, here from the German Classification of Occupations 2010 (KldB 2010). Within this process, errors may happen: Either the occupational activities remain unclear from the job description and/or it is not possible to find a single category that matches the described activities exactly. Thus, coders should take into account these ambiguities for the following task.

In our study, a small set of possible job categories was suggested to the respondent and respondents were asked during the interview to select the most adequate one. The aim is now to assess the quality from this process.

In a first step, two independent coders assigned answers to KldB-categories, each according to personal coding instructions. From interview coding and from both human coders we thus have three independent codes. In a second step, all three codes are given to a human coder for validation. The coder's task is now to determine the quality from all three codes. For this, each of the three codes is to be coded to one of the following categories:

- Acceptable
- Uncertain
- Wrong

Category definitions from the KldB, volume 2, are the most important resource for coders to understand the content of KldB-categories. When the category descriptions provided there are not sufficient for a coding decision, coders should resort to the skill level dimension as described in volume 1. Coders should always take into account all verbatim answers about job tasks and duties. Additional job related answers and the job titles as selected by the respondent during the interview may be looked up when considered helpful.

This research was motivated by the expectation that coding quality improves when the respondent himself codes the answers. Coders are therefore asked to decide if there is evidence that the interview-coded answer is an improvement over manual coding. For detailed coding instructions see below.

We next give definitions and examples to clarify the meaning for the three groups "Acceptable", "Uncertain", and "Wrong". In all examples we provide the relevant answers from the interview process and the 5-digit KldB-codes as they were chosen by the respondents and by the professional coders respectively. All KldB-codes are further described in parenthesis containing the exact category name from the KldB and also the selected job title for interview-coded answers. The specified job titles for interview-coded categories are always those that the respondent chose during the interview. In addition, we provide arguments for illustrative

purposes only why we feel that a specific example should be considered acceptable, uncertain, or wrong:

Acceptable

Definition:
A good argument exists why the category may be considered correct. This is independent from the fact that other plausible arguments may lead to different categories that may be considered correct as well.

Examples:
Interview answers: "Lehrkraft" - "Ich bin Lehrkraft an einer Akademie"
Possible Categories:
Interview-Coded: 84214 (Selected job title: "Lehrer/in - berufliche Schulen", Category title: Lehrkräfte für berufsbildende Fächer - hoch komplexe Tätigkeiten)
Professional Code 1: 84304 (Berufe in der Hochschullehre und -forschung - hoch komplexe Tätigkeiten)
Professional Code 2: 84124 (Lehrkräfte in der Sekundarstufe)
Argument: 84214 and 84304 are both acceptable: Many vocational schools call themselves academy. Also, there are academies of fine arts that are at university level.

Interview answers: "Elektriker" - "Instandhaltung von der technischen Einrichtung, Beleuchtung, Klima"
Possible Categories:
Interview-Coded: 25132 ("Serviceelektriker/in", Technische Servicekräfte in der Wartung und Instandhaltung - fachlich ausgerichtete Tätigkeiten)
Professional Codes: 26212 (Berufe in der Bauelektrik - fachlich ausgerichtete Tätigkeiten)
Arguments: 25132 is acceptable because 1. A main focus is on maintenance and 2. The KldB assigns "Serviceelektriker" to category 25132". 26212 is also acceptable because that category contains most electrical tasks.

Interview answers: "kaufmännischer Angestellter im Außendienst-ich berate,betreue u. verkaufe"
DBS: Angestellter mit eigenständiger Leistung in verantwortlicher Tätigkeit oder mit Fachverantwortung für Personal
übl. erf. Ausbildung: abgeschlossene berufliche Ausbildung
Possible Categories:
Interview-Coded: 61123 ("Außendienstmitarbeiter/in" - Berufe im Vertrieb (außer IuK-Technologien) - komplexe Spezialistentätigkeiten)
Professional Code 1: 61122 (Berufe im Vertrieb - fachlich ausgerichtete Tätigkeiten)
Professional Code 2: 61123 (Berufe im Vertrieb (außer IuK-Technologien) - komplexe Spezialistentätigkeiten)

Arguments: 61123 is acceptable while 61122 is uncertain. 1. The skill level remains unclear from the respondent's answer; 2. the description from KldB, vol. 2, for code 61122 is about indoor service while only 61123 extends to field service.

Interview answers: "Fachinformatiker für Systemintegration"
Possible Categories:
Interview-Coded: 43102 ("Fachinformatiker/in - Systemintegration", Berufe in der Informatik - fachlich ausgerichtete Tätigkeiten)
Professional Code 1: 43104 (Berufe in der Informatik - hoch komplexe Tätigkeiten)
Professional Code 2: 43102 (Berufe in der Informatik - fachlich ausgerichtete Tätigkeiten)
Argument: The KldB assigns "Fachinformatiker" to the category 43102 and we have no reason to believe that the skill level is sufficient for 43104.

Interview answers: "stellvertretener Direktor und Lehrer"
Possible Categories:
Interview-Coded: 84114 ("Lehrer/in - Grundschulen (Primarstufe)", Lehrkräfte in der Primarstufe - hoch komplexe Tätigkeiten)
Professional Codes: 84194 (Führungskräfte an allgemeinbildenden Schulen - hoch komplexe Tätigkeiten)
Argument: 84114 and 84194 are both acceptable: Possible are both situations: Predominant parts of this persons' work may be in teaching at a primary school or in leading the school.

Wrong

Definition:
It is obvious that the category is erroneous and other codes are clearly more adequate. See the examples below for possible arguments when this is the case.

Examples:
Interview answers: "Sachbearbeiterin in einer Zahnarztpraxis, im Verwaltungsbereich tätig"
Interview-Coded: 71402 ("Bürokraft/Kaufmännische Fachkraft", Büro- und Sekretariatskräfte (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten)
Professional Codes: 73222 ("Zahnarztsekretärin")
Argument: 71402 is wrong because a more specialized code is available.

Interview answers: "Psychologin" - "im psychologischen Dienst einer Lebenshilfeeinrichtung berate und betreue Menschen mit seelischer und geistiger Behinderung"
Interview-Coded: 81614 ("Psychologin Wirtschaftspsychologie", Nicht klinische Psychologie)

Professional Codes: 81624 (Klinische Psychologie)
Argument: 81614 is wrong because the KldB, vol. 2, clearly states that counseling to mentally handicapped persons belongs to clinical psychology.

Interview answers: "Bauleiter" - "Bauleitung der Elektromontage auf dem Bau mit Verantwortung für Personal", ohne Meistertitel, beaufsichtigt 2 Arbeitskräfte
Interview-Coded: 33393 ("Bauleiter Ausbau", Aufsichtskräfte - Aus- und Trockenbau, Isolierung, Zimmerei, Glaserei, Rollladen- und Jalousiebau)
Professional Code 1: 26212 ("Bauelektriker")
Professional Code 2: 31194 ("Bauleiter"/"Baustellenleiter")
Argument: Although this respondent claims to be a "Bauleiter", he does not have the required training and supervises not enough workers. 33393 and 31194 are clearly wrong.

Interview answers: "Kundenbetreuer Softwareintegration" - "Beratung, Vertrieb, Support"
Interview-Coded: 62183 ("Kundendienstberater", Berufe im Verkauf (ohne Produktspezialisierung) (sonstige spezifische Tätigkeitsangabe) - komplexe Spezialistentätigkeiten)
Professional Code 1: 43224 ("Softwareberater", Berufe in der IT-Anwendungsberatung - hoch komplexe Tätigkeiten)
Professional Code 2: 43223 ("IT-Kundenbetreuer", Berufe in der IT-Anwendungsberatung - komplexe Spezialistentätigkeiten)
Argument: 62183 is wrong because this person has a specialization in the product offered.

Uncertain

Definition:
This is the residual category to be assigned when a code is not obviously erroneous and at the same time there exist no good argument why this code should be correct. Three reasons are most common why a category is classified as uncertain:

1. The job title selected during the interview appears correct at a first glance, but a different category definition from the KldB, volume 2, describes the job activities more precisely.

2. The interview-coded job category requires a skill level that is contradictory to answers from the interview (i.e., to the questions on the vocational training usually required or the differentiated occupational status)

3. The answers from the interview suggest a different thematic focus, but at the same time the code is not entirely wrong.

Examples:
Interview answers: "Verkäuferin für Lebensmittel" - "beim Discounter Kassentätigkeit ausgeführt"
Possible Categories:
Interview-Coded: 62302 ("Fachverkäufer/in - Nahrungsmittel", Berufe im Verkauf von Lebensmitteln (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten)
Professional Code 1: 62112 (Kassierer/innen und Kartenverkäufer/innen - fachlich ausgerichtete Tätigkeiten)
Professional Code 2: 62102 (Verkauf (ohne Produktspezialisierung) - fachlich ausgerichtete Tätigkeiten)
Argument: 62112 is the most precise description (acceptable) although 62302 is still possible (uncertain). On the contrary, 62102 is wrong because a product specialization exists.

Interview answers: "Verkäuferin in einer Metzgerei an Heißen Theke" - "Verkauf von selbsthergestellten Gerichten"
Berufsausbildung als Fotolaborantin, angestellt seit 2013, angelernte Arbeiterin, beaufsichtigt 2 Arbeitskräfte
Interview-Coded: 62322 ("Fachverkäufer/in Lebensmittelhandwerk (Fleischerei)", Berufe im Verkauf von Fleischwaren - fachlich ausgerichtete Tätigkeiten)
Professional Code 1: 62301 (Berufe im Verkauf von Lebensmitteln (ohne Spezialisierung) - Helfer- und Anlerntätigkeiten)
Professional Code 2: 62101 (Verkauf (ohne Produktspezialisierung) - Helfer- und Anlerntätigkeiten)
Argument: All categories are uncertain because 1.) the skill level is questionable for category 62322, 2.) presumably the person is not a standard butcher salesperson although 3.) some specialization still exists.

Interview answers: "Projektleiter"- "Softwareentwicklung Betreuung eines Softwareprojekts"
Interview-Coded: 43394 ("IT-Projektleiter", Führungskräfte - IT-Netzwerktechnik, IT-Koordination, IT-Administration und IT-Organisation)
Professional Code 1: 43494 ("Führungskräfte - Softwareentwicklung und Programmierung")
Professional Code 2: 71393 ("Projektleiter", Aufsichts- und Führungskräfte - Unternehmensorganisation und Strategie - Komplexe Spezialistentätigkeiten)
Argument: The verbatim answer favors the code 43494 (acceptable), but at first glance the category 43394 is quite reasonable (uncertain). 71393 is obviously wrong because the person leads a computer project.

Interview-Coded is better than manual coding

Definition:
Select yes if the following three conditions are fulfilled:

- the interview-coded job title is plausible,
- it contains additional job details we would otherwise not know about,
- and it leads to a different (hopefully better) code.

If any condition is not fulfilled, select no.

Examples (see the last three examples from the uncertain category for details):
The two interview-coded job titles "Beschichtungsmaschinenführer Kunststoff und Kautschuk" and "Auslieferungsfahrer" fulfill all the required conditions.

Interview answers: "Maschinenführer" - "Einstellen von Maschinen, Tourarbeiten"
Interview-Coded: 22102 ("Beschichtungsmaschinenführer Kunststoff und Kautschuk", Berufe in der Kunststoff- und Kautschukherstellung (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten)
Professional Code 1: 24202 (Berufe in der Metallbearbeitung (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten)
Professional Code 2: 25122 (Maschinen- und Anlagenführer/innen - fachlich ausgerichtete Tätigkeiten)

Interview answers: Kraftfahrer für die Deutsche Post" - "Transport von Paketen"
Interview-Coded: 52182 ("Auslieferungsfahrer", Fahrzeugführer/innen im Straßenverkehr (sonstige spezifische Tätigkeitsangabe) - fachlich ausgerichtete Tätigkeiten)
Professional Codes: 52122 (Berufskraftfahrer/innen (Güterverkehr/LKW) - fachlich ausgerichtete Tätigkeiten)

Conversely, this is not the case for the "IT-Projektleiter". This job title is not plausible because this person appears to be a project leader in software development and not in computer technology. Nor does this title contain additional job details because we know already that this person is a project leader in the computer industry.

Because the student assistants frequently do not agree if interview coding is better than manual coding, we do not analyze the results in the main article. For reasons of completeness we provide the contingency table here:

Table C 1: Contingency table how the two student assistants evaluate the statement “Interview coding is better than manual coding”, cross-tabled over rows and columns.

	No	*Yes*	$\sum$
No	290	17	307
Yes	37	24	61
$\sum$	327	41	368

Part D: Behavior Coding Manual

The following is a translation from the German original. In addition to this manual, the coder was provided with the audio files and an excel file containing IDs, starting times, respondents' verbatim answers about the occupation, and suggested answer options.

B1 Question text was read literally as prescribed.

Question text: *„Sind/Waren Sie in einem der folgenden Berufe tätig?" („Sind" und „waren" ist beides richtig.)* - "Are/were you employed in one of the following occupations?" ("Are" and "were" are both correct.)

1 Yes
2 No
9 Unclear (if none of the categories above is correct)

B2 How are answer options read?

Specify for each answer option separately. Answer options are separated by "||"

1 Read aloud as displayed (needs not to be literally but must not distort the meaning. Example: This applies if "Lager- u. Transportarbeiter/in" is read as "Lager- oder Transportmitarbeiterin")
2 Read aloud but distorting the meaning
3 Not read
4 Abbreviated read aloud
5 No answer option provided by the database
9 Unclear (if none of the categories above is correct)

B3 Why were answer options skipped?

Both answer options can apply. Not obligatory to answer.

1 Options were obviously inadequate
2 Respondent mentioned that job title before that is one suggested answer option

B4 Respondent's behavior

Only the respondent's first reaction is of interest. If the respondent says a word/a sentence and the interviewer replies, the further dialogue is irrelevant for coding purposes.

1 Interruption with answer (Respondent interrupts before all answer options are read and selects a suggested answer option. This option does not apply if the interviewer stops by himself to read the different answers.)
2 Interviewer talks (After (possibly incomplete) reading of possible job categories the respondent is expected to answer. If the interviewer keeps talking instead, select this option.)
3 Respondent decides (The intended normal case: The respondent selects an answer option after the interviewer has finished reading.)
4 puzzled/query (Respondent is a bit puzzled or asks a question)
5 Respondent gives additional information about the job
9 Unclear (if none of the categories above is correct)

B5 Other/Unclear

1 Applicable (Indicate special cases)
2 Not applicable

Note
When the first author analyzed the coded data, it became clear that two issues required further attention: (1) The coding instructions were ambiguous regarding the answer option “other occupation” in cases where less than five job titles were provided to the interviewer. The code for “other occupation” was then written down in different variables. It was possible to discover the resultant coding errors by checking if the assigned codes are logically permitted given the number of answer options that were provided in each case. These errors were patched, which required listening to several interviews. (2) Some interviews were coded as “Unclear” and “Other” in B4 and/or B5. To understand the reasons it was again necessary to listen to these interviews.

References

Dowle, M., Short, T. and Lianoglou, S. (2012). *data.table: Extension of data.frame for Fast Indexing, Fast Ordered Joins, Fast Assignment, Fast Grouping and List Columns*. R package version 1.8.6.
URL: *https://cran.r-project.org/package=data.table*

Feinerer, I., Hornik, K. and Meyer, D. (2008). Text mining infrastructure in R, *Journal of Statistical Software* **25**(1): 1–54.

NIPO Software (2014). *NIPO Fieldwork System*, NIPO Software, Amsterdam.

Oracle Corporation (2014). *MySQL*, Oracle Corporation, Redwood City.

R Core Team (2012). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.

R Core Team (2014). *foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...* R package version 0.8-66.
URL: *https://CRAN.R-project.org/package=foreign*

Ripley, B. and Lapsley, M. (2013). *RODBC: ODBC Database Access*. R package version 1.3-10.
URL: *https://CRAN.R-project.org/package=RODBC*

Statistisches Bundesamt (2010). *Demographische Standards*, Statistisches Bundesamt, Wiesbaden.

Trappmann, M., Beste, J., Bethmann, A. and Müller, G. (2013). The pass panel survey after six waves, *Journal for Labour Market Research* **46**(4): 275–281.

Urbanek, Simon (2013). *Rserve: Binary R server*, Urbanek, Simon. R package version 1.7-3.
URL: *http://CRAN.R-project.org/package=Rserve*

Wickham, H. (2015). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.0.0.
URL: *https://CRAN.R-project.org/package=stringr*

Institut für Arbeitsmarkt- und Berufsforschung
Die Forschungseinrichtung der Bundesagentur für Arbeit

IAB-Discussion Paper 13/2018

Beiträge zum wissenschaftlichen Dialog aus dem Institut für Arbeitsmarkt- und Berufsforschung

Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung

Leitgedanken und Dokumentation

Malte Schierholz
Lorraine Brenner
Lea Cohausz
Lisa Damminger
Lisa Fast
Ann-Kathrin Hörig
Anna-Lena Huber
Theresa Ludwig
Annabell Petry
Laura Tschischka

ISSN 2195-2663

Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung

Leitgedanken und Dokumentation

Malte Schierholz (IAB und Mannheimer Zentrum für Europäische Sozialforschung (MZES), Universität Mannheim)
Lorraine Brenner (MZES, Universität Mannheim)
Lea Cohausz (MZES, Universität Mannheim)
Lisa Damminger (MZES, Universität Mannheim)
Lisa Fast (MZES, Universität Mannheim)
Ann-Kathrin Hörig (MZES, Universität Mannheim)
Anna-Lena Huber (MZES, Universität Mannheim)
Theresa Ludwig (MZES, Universität Mannheim)
Annabell Petry (MZES, Universität Mannheim)
Laura Tschischka (MZES, Universität Mannheim)

Mit der Reihe „IAB-Discussion Paper“ will das Forschungsinstitut der Bundesagentur für Arbeit den Dialog mit der externen Wissenschaft intensivieren. Durch die rasche Verbreitung von Forschungsergebnissen über das Internet soll noch vor Drucklegung Kritik angeregt und Qualität gesichert werden.

The “IAB-Discussion Paper” is published by the research institute of the German Federal Employment Agency in order to intensify the dialogue with the scientific community. The prompt publication of the latest research results via the internet intends to stimulate criticism and to ensure research quality at an early stage before printing.

Inhalt

Zusammenfassung 4
Abstract 4
1 Einleitung 6
2 Hintergrund: „Beruf“ in der Statistik 10
2.1 Beruf, ausgeübte Tätigkeit und Berufsbenennungen 10
2.2 Die deutsche und die internationale Berufsklassifikation 13
3 Leitgedanken zur Hilfsklassifikation 23
3.1 Defizite bisheriger Verfahren 24
3.1.1 Simultane Kodierung nach KldB 2010 und ISCO-08 24
3.1.2 Unpräzise Berufsbenennungen 25
3.1.3 Übersichtliche Darstellung 28
3.2 Vorgehen bei der Entwicklung 29
3.2.1 Disjunkte Hilfskategorien und eindeutige Zuordnung 29
3.2.2 Verständliche Darstellung der Hilfskategorien 31
3.3 Spezialfälle 33
3.3.1 Militärberufe 34
3.3.2 Aufsichts- und Führungskräfte 34
3.3.3 Berufe ohne Spezialisierung 36
3.3.4 Residualkategorien für sonstige spezifische Berufe 38
4 Diskussion 39
Literatur 41

Zusammenfassung

Berufsklassifikationen sind anhand der ausgeübten Tätigkeit gegliedert und entsprechend wird auch in Umfragen zur Erfassung des Berufs nach der „beruflichen Tätigkeit“ gefragt. Obwohl sich diese Abfrage auf die Klassifikation bezieht, wird bei der Kodierung von Antworten nur selten auf die Tätigkeitsbeschreibungen aus der Berufsklassifikation zurückgegriffen. Stattdessen erfolgt die Kodierung meist indirekt, indem Kodierer Berufsbenennungen aus einem Kodier-Index auswählen. Da viele Berufsbenennungen aber unpräzise sind und nur unzureichend die zugrundeliegende Tätigkeit beschreiben, kann es dabei zu fehlerhaften Kodierungen kommen.

Als alternative Vorgehensweise entwickeln wir eine tätigkeitsorientierte Hilfsklassifikation zur Verwendung in computergestützten Vorschlagssystemen. Dies unterstützt Kodierer, die passendste Tätigkeit ohne den Umweg über Berufsbenennungen auszuwählen. Die neue Hilfsklassifikation basiert auf der deutschen *Klassifikation der Berufe 2010* und *der internationalen Standardklassifikation der Berufe 2008* und soll eine simultane Kodierung in beide Klassifikationen ermöglichen. Da zur Nutzung der Hilfsklassifikation Detailkenntnisse über die ausgeübte Tätigkeit des Befragten nötig sind, erwarten wir den größten Nutzen beim Einsatz während des Interviews, wenn Befragte die für sie passendste Tätigkeit selbst auswählen.

Abstract

Occupational classifications are structured by the type of work that employees perform. Consequently, German surveys ask employees about the work they perform in order to collect information about occupation. Although the question is tailored to this classification principle, category definitions describing the work are infrequently used for coding. Instead, coding is more indirect as coders often select job titles from a separate coding index. Since many job titles are imprecise and do not sufficiently describe the work actually performed, incorrect assignments may occur.

As an alternative, we develop an auxiliary classification describing work activities, useful for computer-assisted coding. It allows coders to select the most appropriate work activity without using imprecise job titles for coding. The new auxiliary classification is based on both the *2010 German Classification of Occupations* and the *2008 International Standard Classification of Occupations* and allows simultaneous coding to both classifications. The greatest benefits are realized if detailed knowledge about the respondents’ work activities is available. Therefore, the auxiliary classification is most useful if respondents themselves can select the most appropriate work activity from it.

JEL-Klassifikation: C830, J400

Keywords: Occupation, ISCO-08, KldB 2010, Coding, Survey methodology

Danksagung: Die vorliegende Arbeit wurde von der Deutschen Forschungsgemeinschaft im Rahmen der Sachbeihilfe KR 2211/3-1 gefördert. Idee und Konzeption der neuen Hilfsklassifikation stammen von Malte Schierholz. Die weiteren Autorinnen waren für die Formulierung der einzelnen Hilfskategorien verantwortlich. Wir danken den Kolleginnen und Kollegen am Institut für Arbeitsmarkt- und Berufsforschung und an der Universität Mannheim für wertvolle Hinweise und Diskussionen.

1 Einleitung

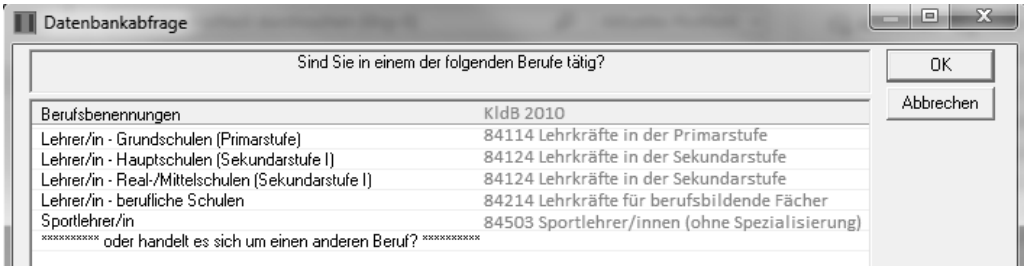
"Und was machen Sie beruflich?" So einfach diese Frage in alltäglichen Unterhaltungen zu beantworten sein mag, so schwierig ist die fehlerfreie Erfassung des Berufs in wissenschaftlichen Befragungen. Die Relevanz ist unbestritten: Seit der ersten „Berufs- und Betriebszählung“ im Deutschen Reich im Jahr 1882 (Rauchberg 1888) wurde die Bevölkerung vielfach in weiteren Volkszählungen und in wissenschaftlichen Befragungen zu ihrem Beruf befragt. Dabei ist die Erfassung des Berufs nach wie vor schwierig, aufwändig und fehleranfällig (Elias 1997; Conrad/Couper/Sakshaug 2016; Schierholz et al. 2018).

Zur Erfassung des Berufs verwenden deutsche Behörden die *Klassifikation der Berufe 2010* (KldB 2010, Bundesagentur für Arbeit 2011). In dieser Klassifikation sind knapp 28.000 Berufsbenennungen aufgelistet, die in 1.286 gesondert dokumentierten *Berufskategorien* zusammengefasst sind. Für internationale Vergleiche wird hingegen zumeist die *International Standard Classification of Occupations 2008* (ISCO-08, International Labour Office 2012) verwendet, die mit 7.018 Berufsbenennungen und 436 ausführlich dokumentierten Berufskategorien jedoch weniger umfangreich ist. Unsere nachfolgend vorgestellte Hilfsklassifikation wurde auf der Grundlage beider Klassifikationen entwickelt und soll eine simultane Kodierung in beide Klassifikationen ermöglichen.

Mit dem vorliegenden Beitrag erfolgt eine Weiterentwicklung eines von Schierholz et al. (2018) vorgestellten Instruments zur Erfassung des Berufs. Dort gelang es, 72,4 Prozent von 1064 zufällig ausgewählten Personen direkt während des Interviews einer Berufskategorie zuzuordnen. Dem Vorgehen lag die Idee zugrunde, den Befragten eine offene Frage nach ihrer beruflichen Tätigkeit zu stellen und ihnen, darauf basierend, automatisiert einige mögliche Antwortoptionen zur Auswahl vorzuschlagen.

Jedoch war die so erzielte Datenqualität nicht vollends zufriedenstellend, weshalb wir nun weitere Verbesserungen anstreben. Schierholz et al. (2018) verwendeten Berufsbenennungen als Antwortoptionen (siehe Abbildung 1), was aber einige Nachteile mit sich bringt, die wir im Folgenden näher erläutern. Zur Verbesserung des Verfahrens schlagen wir nun eine Hilfsklassifikation vor (siehe Abbildung 2), die die charakteristischen Tätigkeiten von Berufen beschreibt. Diese Tätigkeitsbeschreibungen sollen eine genauere Erfassung des Berufs ermöglichen als dies anhand von Berufsbenennungen möglich ist. Dieses Discussion Paper dokumentiert die Entwicklung der Hilfsklassifikation und die zugrundeliegenden Überlegungen.

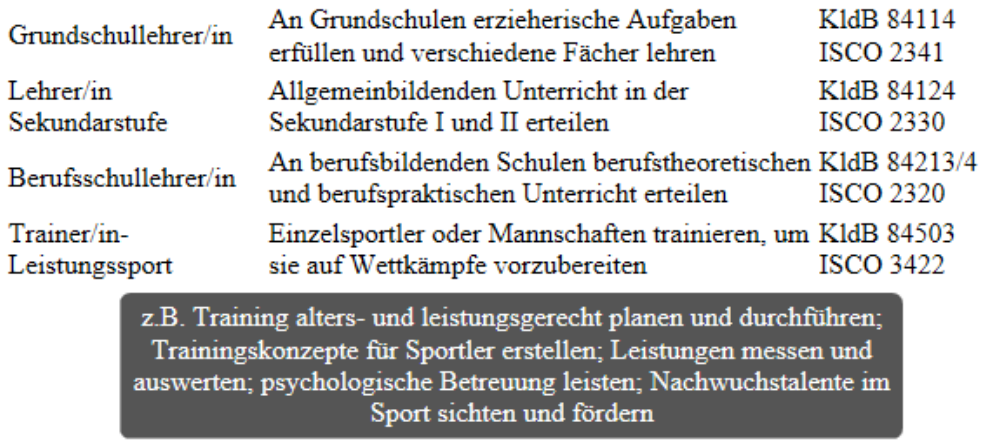
Abbildung 1
Vorgeschlagene Berufsbenennungen



Nachdem ein Befragter „Stellvertretener [sic] Direktor und Lehrer“ geantwortet hat, werden die dargestellten Berufsbenennungen zur weiteren Auswahl vorgeschlagen. Den Berufsbenennungen sind Berufskategorien aus der offiziellen deutschen Berufsklassifikation zugeordnet (grauer Text, der im Interview nicht angezeigt wurde). Schierholz et al. (2018) diskutieren das Beispiel ausführlich.

Quelle: Schierholz et al. (2018)

Abbildung 2
Vorgeschlagene Antwortoptionen aus der neuen Hilfsklassifikation



Exemplarische Darstellung derselben Berufskategorien wie in Abbildung 1 mithilfe von Texten aus der Hilfsklassifikation. Berufsbezeichnungen (Links) geben einen ersten Eindruck und die Beschreibungen (Mitte) machen explizit, was die charakteristische Tätigkeit in den zugeordneten Berufskategorien ist. Der eingeblendete Tooltip (dunkler Hintergrund) enthält bei Bedarf weitere Details zu den einzelnen Bestandteilen der Tätigkeit.

Quelle: eigene Darstellung

Die Nutzung von Tätigkeitsbeschreibungen stellt eine Abkehr von konventionellen Verfahren der Berufskodierung dar. Üblicherweise sollen Befragte zwei bis drei offene Fragen zu ihrer „beruflichen Tätigkeit“ beantworten (Schönbach 1979; International Labour Office 2012; Statistisches Bundesamt 2016). Die Angaben aus den verschiedenen Fragen können dabei im Widerspruch zueinander stehen (Cantor/Esposito 1992; Conrad/Couper/Sakshaug 2016), was den Kodierern die Arbeit erschwert und die Transparenz des Kodierprozesses beeinträchtigt. Zur Kodierung empfehlen Praktiker in unterschiedlicher Intensität die Verwendung von Kodier-Indices, aus denen die Kodierer eine passende Berufsbenennung auswählen sollen (vgl. Geis/Hoffmeyer-Zlotnik 2000; Meier 2003; Geis 2011; International Labour Office 2012; Paulus/Matthes 2013; Loos/Eisenmenger/Bretschi. 2013; Prigge et al. 2014). Auf diese Weise wird die vom Befragten beschriebene berufliche Tätigkeit, die mehrere Aspekte umfassen kann, zu einem einzigen Wort zusammengefasst. In den Kodier-Indices

haben Klassifikationsexperten zu jeder Berufsbenennung die zugehörige Berufskategorie vermerkt, sodass bei der Auswahl einer Berufsbenennung auch gleichzeitig eine Berufskategorie zugeordnet wird. Demselben Prinzip folgend entwickelte auch Tijdens (2010) eine aus Berufsbenennungen bestehende Datenbank, die für den gleichen Zweck wie unsere nachfolgend vorgestellte Hilfsklassifikation konzipiert wurde, nämlich um den befragten Personen mögliche Berufe automatisch vorzuschlagen (Tijdens 2014). Derartige Kodierverfahren, die eine Verwendung von Berufsbenennungen voraussetzen (berufliche Tätigkeit → Berufsbenennung → Berufskategorie), sollen im Folgenden als *paradigmatischer Kodierablauf* bezeichnet werden. In der Praxis wird der paradigmatische Kodierablauf oft nicht in Reinform verwendet. In prinzipieller Hinsicht dient uns der paradigmatische Kodierablauf zur Kontrastierung und soll verdeutlichen, wie sich die Berufskodierung durch den Einsatz der Hilfsklassifikation verändern könnte.

Berufsbenennungen sind grundlegend für den paradigmatischen Kodierablauf, doch zugleich sind viele Benennungen allgemein, unpräzise oder mehrdeutig. Da die Bedeutung von Berufsbenennungen zwischen unterschiedlichen Ländern variiert, wurde die International Standard Classification of Occupations geschaffen. Diese Klassifikation soll die internationale Vergleichbarkeit von statistischen Daten fördern, indem jede dort enthaltene Berufskategorie anhand ihrer charakteristischen Tätigkeiten definiert wird. Es wird unsere Aufgabe sein darzulegen, dass auch auf nationaler Ebene Berufsbenennungen nicht ausreichen, um berufliche Tätigkeiten präzise zu beschreiben.

Sowohl die KldB 2010 als auch die ISCO-08 enthalten umfassende Definitionen der einzelnen Berufskategorien. Diese Definitionen bilden eine Alternative zum paradigmatischen Kodierablauf, denn Beschäftigte lassen sich direkt anhand der von ihnen ausgeübten beruflichen Tätigkeit, ohne den Umweg über Berufsbenennungen, derjenigen Berufskategorie zuordnen, die ihrer Definition nach am besten passt (berufliche Tätigkeit → Berufskategorie). Gegenüber dem paradigmatischen Kodierablauf hat dies den Vorteil, dass Befragte nicht anhand der Ähnlichkeit ihrer sprachlichen Äußerungen, sondern anhand der Ähnlichkeit ihrer tatsächlich ausgeübten Tätigkeiten in Kategorien zusammengefasst werden. Weiterhin ist es für spätere Datennutzer transparenter, wenn sie die Bedeutung der Berufskategorien in den Definitionen nachschlagen können und sie sich nicht auf die Korrektheit der Zuordnungen im Kodier-Index verlassen müssen.

Unsere Hilfsklassifikation soll eine Hilfestellung bieten, mit der diesem Ideal einer direkten Klassifizierung der beruflichen Tätigkeit anhand der Definitionen besser entsprochen werden kann. Zu diesem Zweck stellt die Hilfsklassifikation kurze Zusammenfassungen (Tätigkeitsbeschreibungen) der langen Definitionen bereit, sodass bei einer Kodierung in die Hilfsklassifikation auch zugleich Berufskategorien aus der KldB 2010 und aus der ISCO-08 zugeordnet werden (berufliche Tätigkeit → Tätigkeitsbeschreibung → Berufskategorie). Technisch gesehen übernehmen Tätigkeitsbeschreibungen aus der Hilfsklassifikation dabei bloß die Rolle der Berufsbenennungen beim

paradigmatischen Kodierablauf (vgl. Abbildung 3); ein wichtiger Unterschied liegt aber darin begründet, dass die Hilfsklassifikation präzise beschriebene Kategorien enthält, die unter Bezugnahme auf die Definitionen der offiziellen Klassifikationen speziell für unseren Zweck entwickelt wurden.

Abbildung 3
Wichtige Konzepte für die Hilfsklassifikation

Befragte haben eine berufliche Tätigkeit, die sie mit ihren Freitextantworten, oft nur ungenau, umschreiben. Ziel ist es, die berufliche Tätigkeit (und nicht die unpräzise Freitextantwort) möglichst passend den offiziellen Klassifikationen zuzuordnen. Dabei soll die Hilfsklassifikation eine Kodierung der beruflichen Tätigkeit anhand von Tätigkeitsbeschreibungen ermöglichen und so den paradigmatischen Kodierablauf (Kodierung der Freitextantworten anhand von Berufsbenennungen) ersetzen.
Quelle: Eigene Darstellung

Zur effizienten Nutzung wird man die Hilfsklassifikation in entsprechende Computerprogramme einbinden müssen. Ähnliche computergestützte Vorschlagssysteme („computer-assisted coding“) werden zur Berufskodierung bereits vielfach eingesetzt (vgl. Bushnell 1998; Paulus/Matthes 2013; Elias/Birch/Ellison 2014; Bundesagentur für Arbeit 2017). Auch die in Abbildung 1 dargestellte Lösung von Schierholz et al. (2018) ist ein solches System. Diese Programme orientieren sich jedoch noch am paradigmatischen Kodierablauf und schlagen zur Auswahl Berufsbenennungen vor, wie sie ursprünglich zur Verwendung in gedruckten, alphabetisch sortierten Nachschlagewerken zusammengestellt wurden. Demgegenüber erlauben computergestützte Vorschlagssysteme neue, interaktive Darstellungsformen und eröffnen so weitergehende Möglichkeiten, auf welche Weise Antwortoptionen übersichtlich und zugleich in der nötigen Ausführlichkeit dargestellt werden können. Derartige Computerprogramme gilt es in einem nächsten Schritt zu programmieren, wofür wir mit der Hilfsklassifikation eine inhaltliche Grundlage bereitstellen.

Unser zentrales Ziel ist es, die Validität der Berufsmessung zu erhöhen, indem wir eine Alternative entwickeln, wie die Kommunikation über Berufe auch ohne mehrdeutige und irreführende Berufsbenennungen möglich wird. Unser Ansatz basiert auf der Annahme, dass sich einzelne Berufskategorien jeweils über ihre Kerntätigkeit definieren lassen. In einer kurzen Beschreibung wird diese Kerntätigkeit möglichst präzise benannt, sodass derjenige Beruf (bzw. diejenige Kerntätigkeit) ausgewählt werden kann, der für die ausgeübte Tätigkeit des Befragten am passendsten erscheint. Dieses Alternativprodukt nutzt Tätigkeitsbeschreibungen und ist damit ein Kompromiss

zwischen umfassend dokumentierten Kategorien aus Berufsklassifikationen und vergleichsweise einfachen Berufsbenennungen aus Kodier-Indices. Ob die vorgelegten Tätigkeitsbeschreibungen tatsächlich besser geeignet sind als die bisher verwendeten Berufsbenennungen, muss sich in der Praxis erst noch zeigen. Da zur Auswahl der passendsten Tätigkeitsbeschreibung detaillierte Informationen über die berufliche Tätigkeit des Befragten bekannt sein sollten, erwarten wir den größten Nutzen, wenn die Auswahl durch den Berufstätigen selbst geschieht und nicht erst nach dem Interview durch einen Kodierer.

Dieser Beitrag ist wie folgt strukturiert: Abschnitt 2 erläutert, was unter „Beruf“ zu verstehen ist und anhand welcher Prinzipien Berufsklassifikationen aufgebaut sind. Abschnitt 3 beschreibt die Defizite bisheriger Verfahren, die uns zur Entwicklung der Hilfsklassifikation motivierten. Weiterhin werden grundlegende Prinzipien der Hilfsklassifikation vorgestellt, die bei der Entwicklung berücksichtigt wurden. Abschnitt 4 schließt mit einer Diskussion. Die Hilfsklassifikation und eine umfassende Dokumentation stehen als Anhang zu diesem Artikel auf der Homepage des IAB zum Download bereit.

2 Hintergrund: „Beruf“ in der Statistik

2.1 Beruf, ausgeübte Tätigkeit und Berufsbenennungen

Im Zuge der Arbeitsteilung (Smith 1776) haben sich Gruppierungen („Berufe“) gebildet, deren Mitglieder bestimmte Aufgabenfelder erfüllen. Die Arbeit in einem Beruf erfolgt mit berufstypischen Gegenständen und es werden spezifische Arbeitsverfahren und -techniken verwendet. Viele Berufe sind an einen bestimmten Arbeitsort gebunden und nur in einer einzigen Branche anzutreffen. Zur Ausübung der Tätigkeit werden Wissen, Fähigkeiten und Kompetenzen vorausgesetzt, die in standardisierten Berufsausbildungen erlernt und durch Bildungszertifikate nachgewiesen werden (vgl. Damelang/Schulz/Vicari 2015). Menschen finden aufgrund ihres Berufs Anerkennung in der Gesellschaft, sie identifizieren sich mit ihrem Beruf, machen in ihrem Beruf Karriere und richten ihre Lebensplanung danach aus. Durch die berufliche Sozialisation bilden sich Berufsmilieus („Klassen“, vgl. Weeden/Grusky 2005), die von homogenen Einstellungen und Lebensstilen geprägt sind. Ansprüche an das Gehalt, an betriebliche Rechte sowie an Versorgungsleistungen nach dem Ausscheiden werden durch den Beruf begründet. Zur Vertretung ihrer Interessen schließen sich Angehörige eines Berufs in Gewerkschaften und in berufsständischen Körperschaften (Kammern, Berufsverbände) zusammen. Diese Ausführungen zeigen, dass Beruflichkeit ein multidimensionales Konzept ist, welches sich erst in der Kombination von Ausbildung (fachliches Wissen, Zertifikate), Tätigkeit (Aufgabe, Arbeitsmaterial, Kompetenzen), Einbettung in ein soziales Umfeld (Arbeitgeber, berufsständische Vertretung, Identifikation, Milieu) und durch die Gesellschaft zugeschriebene Eigenschaften (gesetzliche Vorgaben und Privilegien, zugedachte Rolle, Berufsprestige) ausdrückt. Freilich ist eine derart umfassend verstandene Beruflichkeit nicht für alle Tätigkeiten gleich stark ausgeprägt und vielleicht am ehesten in ärztlichen und juristischen Beru-

fen sowie im Handwerk vorhanden. Der Beruf und die damit verbundenen Arbeitsaufgaben dienen aber nach wie vor als ein strukturgebendes Prinzip in der Gesellschaft, welches für viele Berufstätige von Bedeutung ist (vgl. Dostal/Stooß/Troll 1998; Demszky v.d. Hagen/Voß 2010; Watson 2012).

All die genannten Aspekte des Berufs statistisch in einer einzigen Variablen zu erfassen, ist nicht möglich. Stattdessen hat sich die internationale Berufsstatistik bereits früh darauf geeinigt, dass „Beruf" ein arbeitsplatzbezogenes Merkmal des Beschäftigten sein soll, welches die dort ausgeübte Tätigkeit („work performed", „material worked in or handled", „process performed" in den Worten des International Labour Office (1923)) in den Mittelpunkt rückt. Insbesondere wurde argumentiert, dass der individuelle Beruf vom Wirtschaftszweig (Branche) des Arbeitgebers zu unterscheiden ist, einem weiteren Merkmal der Erwerbsstatistik. Tatsächlich wurden in den ersten Berufszählungen im Deutschen Reich noch beide Konzepte miteinander vermischt und erst ab 1925 wurde mit einer systematischen Trennung von Wirtschaftszweig und Beruf begonnen. Die Schwierigkeiten der Trennung lassen sich unmittelbar vor Augen führen, wenn man an die vorindustriellen Handwerksberufe denkt, beispielsweise an den Schuhmacher, der seine eigene Werkstatt führt und Schuhe per Hand fertigt, denn dort waren Wirtschaftszweig und Beruf noch gleichbedeutend. Erst die Erkenntnis, dass beispielsweise ein Tischler nicht nur in der Möbeltischlerei tätig sein kann, sondern auch in der Maschinenindustrie und in anderen Branchen, machte es erforderlich, den Beruf als eigenständiges Merkmal statistisch zu erheben (International Labour Office 1923; Meerwarth 1925; Fürst 1929; Willms 1983).

Die Statistik folgt damit weitestgehend einer Definition von Max Weber (1921: 80):

> *„Beruf soll jene Spezifizierung, Spezialisierung und Kombination von Leistungen einer Person heißen, welche für sie Grundlage einer kontinuierlichen Versorgungs- oder Erwerbschance ist."*

Aufgrund der Multidimensionalität des Berufskonzeptes ist es aber schwierig, die ausgeübte Tätigkeit in Bevölkerungsbefragungen fehlerfrei zu erfassen. Viele Befragte antworten auf die Frage nach ihrer „beruflichen Tätigkeit" mit einer Berufsbenennung, die für die Statistik weiter verarbeitet werden muss. Allerdings nimmt Stooß zufolge die Anschaulichkeit derartiger Berufsbenennungen immer weiter ab, sodass die Benennungen „immer aussageärmer und damit letztlich völlig unverständlich werden" (1977: 73).

Zum Beleg nennt Stooß (1977) einige Berufsbenennungen aus dem Jahr 1568 – Buchdrucker, Goldschmied, Koch, Müller, Bäcker –, die anschaulich das Produkt und das Herstellungsverfahren beschreiben. Aus verschiedenen Gründen ist heutzutage eine derartige Anschaulichkeit nicht mehr für alle Benennungen gegeben, sodass auch die Grenzen zwischen einzelnen Berufen verschwimmen:

a) Im Zuge der Industrialisierung haben sich die ehemaligen Handwerksberufe immer weiter aufgefächert und eine Vielzahl neuer Berufe/Berufsbenennungen ist

entstanden. Stooß nennt beispielhaft für neu entstandene Spezialtätigkeiten den Industriebuchbinder, den Metalldrücker, den Maschinenglasbläser sowie als Beispiele für hunderte weitere Bezeichnungen im „Schneider“-Beruf die Näher, Stepper (benannt nach der Maschinenfunktion), Krawattennäher (benannt nach dem hergestellten Produkt), Hefter und Garnierer (benannt nach Spezialtätigkeit).

b) Neue unterstützende Aufgaben entstehen im Zusammenhang mit der Professionalisierung von Berufen (z.B. Medizinisch-technischer Assistent, Ingenieurassistent, Technische Zeichner, Agrarlaborant).

c) Berufsbenennungen sollen nicht abwertend wirken, sondern attraktiv und interessant klingen oder ein gewisses Image vermitteln (z.B. Raumpfleger anstelle von Putzhilfe, Sekretär anstelle von Schreibkraft, Sozialarbeiter anstelle von Fürsorger, Klinikreferent anstelle von Handelsvertreter der Pharmaindustrie). Aus demselben Grund werden oft auch englischsprachige Benennungen verwendet (z.B. Texter, Visualizer, Campaigner, Manager).

d) Seit jeher sind Berufsbenennungen im Bereich Handel, Büro und Verwaltung wenig ausdifferenziert. Noch heute antworten in diesem Bereich Beschäftigte häufig mit allgemeinen, wenig anschaulichen Bezeichnungen (z.B. Bürokaufmann, Sachbearbeiter, kaufmännischer Angestellter) oder eine nähere Beschreibung erfolgt unter Bezugnahme auf den Wirtschaftszweig (z.B. Bankkaufleute).

Die aufgeführten Beispiele zeigen bereits die sprachliche Vielfalt von Berufsbenennungen. Stooß und Saterdag (1979: 44) nennen neun Merkmale, dargestellt in Tabelle 1, die bei der Kommunikation über und bei der Systematisierung von Berufen verwendet werden.

Tabelle 1
Verschiedene Dimensionen von Berufsbenennungen

Werkstoff, Material, Produkt	Objekt der Tätigkeit	*Metall*arbeiter
Arbeitsverfahren, -techniken	Aktivitätstyp, -kombination	*Melk*er
Arbeitsgerät (Maschinen, Werkzeuge	Instrumentierung	*Fräsmaschine*nbediener
Betrieblicher Einsatzbereich	Funktionsbereich	*Innendienst*leiter
Arbeitsmilieu, -ort, -platz	„Allokation der Arbeitskraft“	*Schleusen*wärter
Wirtschaftszweig, Branche	wirtschaftsfachliche Zuordnung	*Versicherungs*kaufmann
Hierarchische Einordnung in den Betrieb	Stellung im Betrieb	Technischer *Assistent*
Stellung im Beruf	Status	Finanz*beamter*
Üblicher Zugang, erforderliche Ausbildung	Qualifikation	*Jurist*

Zur Veranschaulichung haben wir beispielhafte Berufsbenennungen aus der KldB 2010 hinzugefügt.
Quelle: Stooß und Saterdag (1979: 44).

Wenn uns also der Beruf im Sinne der ausgeübten Tätigkeit interessiert, dann lässt sich dies sprachlich in mindestens neun Dimensionen beschreiben. Die einzelnen Dimensionen hängen eng miteinander zusammen. In Benennungen ist häufig bereits mehr als eine Dimension enthalten, oder es lassen sich bei Kenntnis einer Dimension Informationen zu weiteren Dimensionen ableiten. Auf diese Weise entstehen aus einfachen Berufsbenennungen komplexe Bilder und Vorstellungen zum jeweiligen Beruf; sie erzeugen ein Stereotyp, das von der Realität weit entfernt sein kann. Stooß und

Saterdag (1979) bezeichnen Berufsbenennungen daher als „Bündel von Informationen“, die eine umgangssprachliche Verständigung über Berufe erst möglich machen, aber nicht auf eine wissenschaftliche Erfassung des Berufs ausgerichtet sind.

2.2 Die deutsche und die internationale Berufsklassifikation

Für die Statistik und die Arbeitsverwaltung stellt sich nun die Aufgabe, die berufliche Vielfalt zu ordnen und zu systematisieren. Berufsklassifikationen erfüllen diesen Zweck, indem sie ähnliche Berufe/Berufsbenennungen zu möglichst homogenen systematischen Einheiten zusammenfassen und in einer monohierarchischen Struktur anordnen. Darüber hinaus werden Berufsklassifikationen in Bevölkerungsbefragungen verwendet, um Befragte mit hinreichend ähnlicher Tätigkeit in künstlich erschaffenen systematischen Einheiten zusammenzufassen. Unser Produkt baut auf bestehenden Berufsklassifikationen auf, weshalb es erforderlich ist, zunächst die dort zugrundeliegenden Prinzipien zu beschreiben.

Mit der deutschen *Klassifikation der Berufe 2010* (KldB 2010) steht eine aktuelle Klassifikation bereit, welche die Besonderheiten des deutschen Arbeitsmarktes berücksichtigt und zugleich eine möglichst hohe Kompatibilität zur *International Classification of Occupations 2008* (ISCO-08) aufweist. Die ISCO selbst wurde entwickelt, um die internationale Vergleichbarkeit von berufsstatistischen Daten zu erleichtern und um Länder bei der Entwicklung und Überarbeitung ihrer eigenen Klassifikationen zu unterstützen. Die KldB 2010 und die ISCO-08 entstanden, da sich Berufe im ständigen Wandel befinden und daher Aktualisierungen der veralteten Vorläuferversionen von 1988/1992 notwendig wurden.

Wie bereits die *Klassifizierung der Berufe von 1961* anmerkt (Statistisches Bundesamt 1961), besteht ein für unsere Zwecke wesentlicher Unterschied zwischen beiden Klassifikationen:

a) Die deutschen Systematiken zielten traditionell darauf ab, die vorkommenden Berufsbenennungen – in der aktuellen Fassung enthält die KldB 2010 knapp 28.000 Nennungen – zu ordnen und zu diesem Zweck in systematischen Einheiten zusammenzufassen. Auch ungeschulten Kräften wurde es auf diese Weise ermöglicht, Berufsbenennungen aus Befragungen zweifelsfrei den systematischen Einheiten zuzuordnen (Stooß 1977). Die Entwicklung der deutschen Berufsklassifikationen und der paradigmatische Kodierablauf sind daher eng miteinander verwoben. Für die Beibehaltung dieses Vorgehens sprechen bis heute pragmatische Gründe, da viele Befragte auf die Frage nach ihrer „beruflichen Tätigkeit“ mit einer Berufsbenennung antworten.

b) Für die internationale Vergleichbarkeit besteht hingegen die Herausforderung, dass sich Bedeutung und Verwendung von beruflichen Begrifflichkeiten in einzelnen Sprachen stark unterscheiden können. Aus diesem Grund wurde bei der ISCO besonders auf klare, qualitätsvolle Definitionen der ausgeübten Tätigkeiten in den systematischen Einheiten geachtet und die wenigen zugeordneten Berufsbenennungen sind bloß als zusätzliches Hilfsmittel vorgesehen. Da die ISCO also auf den Inhalt der Berufstätigkeit abzielt, wird im deutschen Kontext empfohlen,

im Interview explizit nach der „beruflichen Tätigkeit“ zu fragen und nicht etwa nach einer Berufsbenennung, was dem Prinzip der deutschen Klassifikationen entsprechen würde (Geis/Hoffmeyer-Zlotnik 2000).

Tabelle 2 stellt den monohierarchischen Aufbau von Klassifikationen dar und verdeutlicht zugleich den beschriebenen Unterschied zwischen den deutschen und den internationalen Systematiken. Dazu ordnen wir die Berufsbenennung „Krawattennäher“ in die verschiedenen Klassifikationen ein und listen exemplarisch einige benachbarte und die übergeordneten Kategorien auf.

a) In den deutschen Klassifikationen ist die Zuordnung eindeutig, da berufskundliche Experten bereits über die zutreffende systematische Einheit entschieden haben. Dem paradigmatischen Kodierablauf folgend ist eine Zuordnung anhand des Kodier-Index allein aufgrund der Wortgleichheit in die Kategorie 3561 (KldB 1988) bzw. 28222 (KldB 2010) möglich. Dabei wird aber vernachlässigt, dass aus dem Interview oft weitere Informationen vorliegen, die eine derart sprachlich bedingte Zuordnung in Frage stellen können. Falls etwa bekannt wäre, dass diese Person auch noch andere Textilien näht, wäre in der KldB 1988 die Kategorie 3560 ebenfalls möglich.

b) Die Einordnung nach der ISCO-08 kann hingegen nicht nach sprachlichen Kriterien erfolgen, da für die ISCO-08 kein Kodier-Index vorliegt, der diese Berufsbenennung enthält. Stattdessen sind Kenntnisse über Tätigkeitsinhalte erforderlich oder müssen unterstellt werden, damit eine Entscheidung für die bestmöglich passende systematische Einheit getroffen werden kann. Wenn die Krawatten maßgeschneidert für einzelne Kunden hergestellt werden, passt die Kategorie 7531 ihrer Definition nach am besten. Wenn hauptsächlich per Hand mit Nadel und Faden gearbeitet wird, trifft Kategorie 7533 zu. Die Kategorie 8153 sollte zugeordnet werden, wenn die Bedienung entsprechender Maschinen charakteristisch für die ausgeübte Tätigkeit ist. Das soeben beschriebene Verfahren einer direkten Zuordnung von Befragten auf Grundlage ihrer Tätigkeit ohne Umweg über die Berufsbenennung stellt unsere Vorstellung vom Idealfall dar; im Gegensatz dazu empfiehlt die ISCO-08 die Entwicklung nationaler Kodier-Indices und die Verwendung des paradigmatischen Kodierablaufs.

Für die Entwicklung der Hilfsklassifikation sind die Definitionen von Berufskategorien maßgeblich, was eher einer ISCO-Denkweise entspricht. Das bis heute in der KldB vorrangig verfolgte Ziel, Berufsbenennungen zu systematisieren, ist für unsere Zwecke hingegen zweitrangig und führte bei unserer Arbeit zu kleineren Komplikationen. Tabelle 3 stellt weitere Merkmale beider Klassifikationen gegenüber. In weiten Teilen basieren beide Klassifikationen trotz der unterschiedlichen Zielsetzung auf ähnlichen Konzepten. Da unser Produkt Konzepte aus der KldB 2010 und aus der ISCO-08 übernimmt, sollen die entsprechenden Konzepte zunächst beschrieben werden.

Tabelle 2
Auszüge aus verschiedenen Klassifikationen zur Einordnung des Krawattennähers. Bei systematischen Einheiten der untersten Ebene sind zugeordnete Berufsbenennungen aufgelistet.

Quelle: Klassifizierung der Berufe von 1988

3	Fertigungsberufe			
	35	Textilverarbeiter		
		351	Schneider	
		352	Oberbekleidungsnäher	
		…		
		356	Näher, anderweitig nicht genannt	
			3560	Näher, ohne nähere Angabe
			3561	Krawattennäherinnen Bindernäher, Krawattennäherin, …
			3562	Gardinennäher
			3563	Segelmacher
			…	
		357	Sonstige Textilverarbeiter	

Quelle: Klassifikation der Berufe von 2010

2	Rohstoffgewinnung, Produktion und Fertigung				
	28	Textil- und Lederberufe			
		281	Textiltechnik und -produktion		
		282	Textilverarbeitung		
			2821	Berufe im Modedesign	
			2822	Berufe in der Bekleidungs-, Hut- und Mützenherstellung	
				28221	Helfer- und Anlerntätigkeiten Bekleidungshelfer/in, Schneiderhelfer/in, Textilverarbeiterhelfer/in, …
				28222	Fachlich ausgerichtete Tätigkeiten Änderungsschneider/in, Bekleidungstechnische/r Assistent/in, Damenschneider/in, Krawattennäher/in, …
				28223	Komplexe Spezialistentätigkeiten Bekleidungstechniker/in, Techniker/in – Bekleidungstechnik, Bekleidungsgestalter/in – Damenbekleidung, …
				28224	Hoch komplexe Tätigkeiten Bekleidungsingenieur/in
			2823	Technische Konfektionäre/Konfektionärinnen, Segelmacher/innen	

Quelle: International Standard Classification of Occupations, 2008

7	Craft and Related Trades Workers			
	75	Food Processing, Woodworking, Garment and Other Craft […] Workers		
		753	Garment and Related Trades Workers	
			7531	Tailors, Dressmakers, Furriers and Hatters Dressmaker, Fur grader, Furrier, Hatter, Milliner, Tailor, …
			7532	Garment and Related Patternmakers and Cutters Fur patternmaker, Garment cutter, Garment patternmaker, …
			7533	Sewing, Embroidery and Related Workers Embroiderer, Sewer, Umbrella maker, …
8	Plant and Machine Operators and Assemblers			
	81	Stationary Plant and Machine Operators		
		815	Textile, Fur and Leather Products Machine Operators	
			8152	Weaving and Knitting Machine Operators Carpet weaving machine operator, Knitting machine operator, …
			8153	Sewing Machine Operators Embroidery machine operator, Sewing machine operator,

Tabelle 3
Vergleich der Klassifikation der Berufe 2010 und der International Standard Classification of Occupations, 2008

	KldB 2010	**ISCO-08**
Status	Herausgeber der KldB 2010 ist die *Bundesagentur für Arbeit*. Die Klassifikation wird ebenfalls durch die *statistischen Ämter des Bundes und der Länder* verwendet.	Die ISCO-08 ist die offizielle Berufsklassifikation der *International Labour Organization.* Sie wird zur statistischen Berichterstattung von der *Europäischen Kommission* empfohlen.
Umfang	1.286 Berufsgattungen auf der untersten Ebene (5-Steller) 27.730 Berufsbenennungen im alphabetischen Verzeichnis	436 unit groups auf der untersten Ebene (4-Steller) 7.018 engl. Berufsbenennungen im Index of occupational titles
Klassifikationsobjekte	"Beruf" ist über drei zentrale Eigenschaften definiert: - „Der Berufsbegriff ist tätigkeits- und nicht personenbezogen." - „'Beruf' zeichnet sich durch ein Bündel von Tätigkeiten aus." - „'Beruf' wird durch zwei zentrale Dimensionen konstituiert: Berufsfachlichkeit und Anforderungsniveau."	"Job" (definiert als „a set of tasks and duties performed, or meant to be performed, by one person, including for an employer or in self-employment")
Zuordnung	„Die Klassifikationsmerkmale [sollten] homogen sein […], um eine eindeutige Zuordnung der Berufe zu gewährleisten. […] Berufe [werden] der Klassifikationseinheit auf der untersten Gliederungsebene zugeordnet, in der sich die Berufe befinden, die die größte Ähnlichkeit zu dem zuzuordnenden Beruf vorweisen. Bei der Zuordnung ist allein die tatsächlich ausgeübte Tätigkeit bzw. der Tätigkeitsschwerpunkt ausschlaggebend."	"[A]ll jobs […] can be assigned to one (and only one) of these unit groups. […] Decisions should be made on the basis of the tasks actually performed, rather than on the level of qualifications required in a particular country [or some other qualification-related aspect]."
Abdeckung/ Vollständigkeit	Obwohl dies nicht explizit formuliert wurde, sollen offenbar möglichst alle relevanten Berufe auf dem deutschen Arbeitsmarkt enthalten sein.	"all jobs in the world" Ausbildungen:

	KldB 2010	**ISCO-08**
	Auch Berufsausbildungen und Studiengänge werden in die KldB 2010 eingeordnet (vgl. das alphabetische Verzeichnis bzw. die Berufsinformationssysteme der Bundesagentur für Arbeit).	"Apprentices and trainees are classified to the occupation they are training for, if they are employed in the capacity of apprentice or trainee."
Gliederungsprinzip	Ähnlichkeit in den beiden Dimensionen Berufsfachlichkeit und Anforderungsniveau: (a) Berufsfachlichkeit ist definiert als „ein auf berufliche Inhalte bezogenes Bündel von Fachkompetenzen. Eine Fachkompetenz umfasst spezifische Kenntnisse und Fertigkeiten […], die auf einzelne Arbeitstätigkeiten zugeschnitten [und für die Ausübung eines Berufs erforderlich] sind." (b) Anforderungsniveau „bezieht sich auf die Komplexität der auszuübenden Tätigkeiten". Das Anforderungsniveau ist damit „eng an den formalen beruflichen Bildungsabschlüssen ausgerichtet. […] Berufserfahrung und/oder die informelle berufliche Ausbildung" können die formale Ausbildung aber ersetzen.	Ähnlichkeit von "Skill" (definiert als "ability to carry out the tasks and duties of a given job") Skill hat zwei Dimensionen: (a) "Skill level" (a "function of the complexity and range of tasks and duties") (b) "Skill specialization": (a) The field of knowledge required (b) The tools and machinery used (c) The materials worked on or with (d) The kinds of goods and services produced
Kategoriedefinitionen	Beschreiben den Inhalt jeder Kategorie mittels: (a) „eine[r] kurze Inhaltsbeschreibung" (b) „eine[r] Liste der Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, die üblicherweise den Kern der Klassifikationseinheit auszeichnen" (c) „Zugeordnete Berufe (Beispiele)" (d) „Negativabgrenzungen, indem ähnliche […] Berufe, die an anderer Stelle der Klassifikation verortet sind, benannt werden" Bei der Zuordnung von Berufen sollte auch auf Beschreibungen hierarchisch übergeordneter Kategorien zurückgegriffen werden.	Definieren den Inhalt jeder Kategorie mittels: (a) „lead statement [which] summarizes the scope and basic nature" (b) "statement of tasks [which] indicates main tasks typically, or usually, performed" (c) "examples of occupations classified here" (d) "related occupations classified elsewhere" (e) "Notes [to] clarify the boundaries between related groups

	KldB 2010	**ISCO-08**
Führungs-kräfte	„Führungskrafte“ in der KldB sind äquivalent zu Managern in der ISCO-08. „Leitung bzw. Führung ist als spezielle Berufsfachlichkeit zu interpretieren“ und hat daher eine „9“ als vierte Ziffer (mit Ausnahme der Gruppe der Top-Manager). „Aufgrund der hohen Komplexität wird einheitlich allen Führungskräften […] das Anforderungsniveau 4 zugewiesen.“	“Manager” sind in der “Major Group 1” zusammengefasst. „The critical difference [between managers and supervisors] is that supervisors are responsible only for supervision of the activities of other workers, whereas [managers] have overall responsibility for the operations of a business or an organizational unit. […] Managers usually have responsibility and make decisions about [at least one of the following]: (a) the overall strategic and operational direction […]; (b) budgets […]; (c) selection, appointment, and dismissal of staff.”
Aufsichts-kräfte	Aufsichtskräfte sind im gleichen berufsfachlichen Cluster verortet wie Führungskräfte (4. Ziffer = „9“). Dies basiert auf der „Annahme, dass der Tätigkeitsschwerpunkt [von Aufsichtskräften] in der Erledigung von Führungsaufgaben liegt. […] Da die Tätigkeit einer Aufsichtskraft im Vergleich zu einer Führungskraft weniger komplex […] ist, wurde allen Aufsichtskräften […] das Anforderungsniveau 3 zugewiesen.“	“Both managers and supervisors plan, organize, coordinate, control and direct the work done by others.” Im Gegensatz zu Managern gilt, dass “supervisors […] do not have authority to make decisions.” Für bestimmte Sektoren liegen 6 “supervisory unit groups” vor. “All other supervisory occupations are classified in the same unit group as the most skilled workers supervised.”
Angehörige der regulären Streitkräfte	„In Anlehnung an ISCO-08 […] erhalten die Militärberufe auch in der KldB 2010 auf oberster Ebene eine eigene Systematikposition.“ Berufsbezeichnungen, die außerdem an anderer Stelle in der Klassifikation eine passende Kategorie hätten (z.B. Stabsarzt, Militärmusiker), sind ihrem Dienstgrad entsprechend beim Militär eingeordnet.	Angehörige der regulären Streitkräfte sind ungeachtet ihrer beim Militär ausgeübten Tätigkeit der “Major Group 0: Armed Forces” zugeordnet (z.B. Stabsarzt).
Allgemeine Kategorien	„Berufs- und Tätigkeitsbezeichnungen, die innerhalb der Berufsfachlichkeit einer Berufsgruppe (3-Steller) keinen spezifischen	“Problems may arise […] when, in the case of some jobs, the range of tasks and duties performed does not

	KldB 2010	ISCO-08
	Tätigkeitsschwerpunkt erkennen lassen, werden den Berufsuntergruppen zugeordnet, die ‚keine Spezialisierung' vorsehen." (4. Stelle = 0)	correspond exactly to those specified in the classification. In such cases application of the following rules is suggested, in the order of precedence given below […]. (a) In cases where the tasks and duties performed require skills usually obtained though different levels of training and experience, jobs should be classified in accordance with those tasks and duties which require the highest level of skills. […] (b) In cases where the tasks and duties are connected with different stages of the production and distribution of goods process, tasks and duties related to the production stage should take priority over associated ones, such as those related to the sales and marketing of the same goods, their transportation or the management of the production process. […] (c) Where the tasks and duties performed are both at the same skill level and at the same stage of production, jobs should be classified according to the predominant tasks performed. […]"
Residualkategorie	"Liegt […] eine Spezialisierung vor und ist diese keiner anderen Berufsuntergruppe innerhalb der gewählten Berufsgruppe berufsfachlich zuzuordnen, so ist die Berufsbezeichnung eine ‚sonstige spezifische Tätigkeitsangabe'" (4. Stelle = 8).	Residual unit groups ("not elsewhere classified") haben als letzte Ziffer eine "9"

Quelle: Eigene Darstellung auf Basis von der KldB 2010 und ISCO-08

Für unsere weitere Arbeit verwenden wir die systematischen Einheiten der untersten Ebene aus der KldB 2010 und aus der ISCO-08. Dabei handelt es sich in der KldB 2010 um 1.286 sogenannte Berufsgattungen und in der ISCO-08 um 436 sogenannte unit groups, die wir klassifikationsübergreifend in diesem Text als *Berufskategorien* bezeichnen.

Die Definitionen der zu klassifizierenden Objekte, „Beruf“ in der KldB 2010 und „job“ in der ISCO-08, sind jeweils tätigkeitsbezogen und beides zeichnet sich durch ganze „Bündel von Tätigkeiten“ bzw. „sets of tasks and duties“ aus. Jeder Beruf/Job soll genau einer einzigen systematischen Einheit auf der untersten Gliederungsebene in der jeweiligen Klassifikation zugeordnet werden können. Für die Zuordnung soll in beiden Klassifikationen die tatsächlich ausgeübte Tätigkeit entscheidend sein.

Die ISCO-08 hat explizit den Anspruch, dass „all jobs in the world“ der Klassifikation zugeordnet werden können. Diese Aussage gilt in eingeschränkter Form auch für die KldB 2010. Frühere deutsche Berufsklassifikationen von 1961, 1970 und 1988 definieren Beruf als „die auf Erwerb gerichteten […] Arbeitsverrichtungen […], durch die der einzelne an der Leistung der Gesamtheit im Rahmen der Volkswirtschaft mitschafft“. Diese Definition erklärt, warum ehrenamtliche Tätigkeiten (z.B. „Schöffe“), nicht monetär vergütete Tätigkeiten (z.B. „Hausfrau“), theoretische Ausbildungen (z.B. „Schüler“, „Student“) und sonstige Spezialfälle (z.B. „Professor emerit.“, „Schrebergärtner“, „Briefmarkensammler“) in der aktuellen Berufsklassifikation KldB 2010 nicht enthalten sind. Da unsere Hilfsklassifikation anhand der KldB 2010 entwickelt wurde, sind dort ebenfalls keine Kategorien für derartige Bezeichnungen enthalten. Plausible Antworten auf die Frage nach dem Beruf finden also kein passendes Gegenstück in der Hilfsklassifikation und müssten bei Bedarf ggf. ergänzt werden.

Klassifikationen haben die Aufgabe, Berufe in möglichst homogenen systematischen Einheiten zusammenzufassen und diese systematischen Einheiten wiederum in übergeordneten Einheiten zu aggregieren. Dies erfordert, Vorstellungen über die Ähnlichkeit von Berufen und der systematischen Einheiten zu entwickeln. Die gesamte Struktur von Berufsklassifikationen basiert auf der Ähnlichkeit der systematischen Einheiten zueinander (vgl. Embury 1997). Die KldB 2010 und die ISCO-08 verwenden jeweils zwei Dimensionen von Ähnlichkeit: Zum einen beschreibt „Berufsfachlichkeit“ bzw. „skill specialisation“ die üblicherweise benötigten fachlichen Kompetenzen; zum anderen beschreibt „Anforderungsniveau“ bzw. „skill level“ die Komplexität der Tätigkeit, was eng an die beruflichen Bildungsabschlüsse angelehnt ist. Beides ist nicht personenbezogen, sondern soll alleine in Bezug auf die tatsächlich ausgeübte Tätigkeit die Ähnlichkeit zwischen verschiedenen Berufen bzw. systematischen Einheiten bemessen.

Aus unserer Sicht finden solche Ähnlichkeitsdefinitionen in erster Linie Verwendung bei der Erstellung von Klassifikationen und begründen für jede Ebene der Klassifikation, welche Ähnlichkeitsdimension zu gelten hat. In ISCO-08 ist beispielsweise die oberste Ebene (1-Steller) nach „skill level“ gegliedert und auf den unteren drei Ebenen

(2-4-Steller) werden Berufe nach ihrer „skill specialisation“ unterschieden. Embury (1997: 13) zufolge sind die explizit formulierten Ähnlichkeitsdefinitionen nicht nur bei der Erstellung der Klassifikationen erforderlich, sondern auch zur Einordnung neuer Berufe in die Klassifikation, denn „if the two occupations have few or no tasks in common […], the criteria will provide guidance on where to classify new occupations“. Zugleich räumt er aber auch ein, dass die Ähnlichkeit zwischen zwei Berufen bestimmt werden kann „by a direct comparison of the tasks involved“, sofern beide Berufe „almost the same sets of tasks“ haben. In den meisten Fällen lässt sich aufgrund der hohen Ähnlichkeit die passendste Berufskategorie vermutlich durch direkten Vergleich der „tasks involved“ auswählen. Beispielsweise braucht ein Kodierer, der „Krawattennäher“ klassifizieren möchte, sich nicht zuerst die Ähnlichkeitsdefinition von „skill“ (i.e., „field of knowledge required“, “tools and machinery used“, „materials worked on or with“) ins Gedächtnis zu rufen, sondern er wird lediglich die Aufgaben und Tätigkeiten („tasks“) des Krawattennähers mit den Kategoriebeschreibungen aus der Klassifikation vergleichen müssen. Die Berücksichtigung der wenig präzisen Ähnlichkeitsdefinitionen aus der ISCO-08 gibt also kaum Orientierungshilfe und wir messen ihnen zum Zwecke der Kodierung von Berufsangaben keine besondere Bedeutung bei.

Das Gesagte gilt auch für die Definition von Berufsfachlichkeit, der ersten Ähnlichkeitsdimension der KldB 2010. Das Anforderungsniveau, die zweite Ähnlichkeitsdimension in der KldB 2010, bringt aber gewisse Schwierigkeiten mit sich. Für die Zuordnung von Berufsbenennungen zu Berufskategorien verwendet die KldB 2010 ein zweistufiges Verfahren: Zuerst werden die Berufsbenennungen anhand ihrer berufsfachlichen Ähnlichkeit einem 4-Steller zugeordnet und anschließend werden die Anforderungsniveaus (5-Steller) der jeweiligen Benennungen bestimmt. Aus zwei Gründen wird dieses Verfahren problematisch, wenn nicht Berufsbenennungen sondern berufliche Tätigkeiten von Befragten in bestehende Kategorien der Klassifikation eingeordnet werden sollen.

a) Das zweistufige Verfahren ist nicht vollständig durchführbar, denn nicht für jeden 4-Steller sind alle vier möglichen Anforderungsniveaus in der Klassifikation enthalten.

b) Weiterhin suggeriert das zweistufige Vorgehen, dass Berufsfachlichkeit und Anforderungsniveau zwei zusammenhanglose Dimensionen seien, die am besten getrennt voneinander in zwei Fragen abgefragt würden. Entsprechende Vorschläge für unterschiedliche Fragen zur Erfassung des Anforderungsniveaus machen Paulus und Matthes (2013) und Müller (2014). Aus der Frage nach der beruflichen Tätigkeit und der zweiten Frage können sich dann konkurrierende Werte für das Anforderungsniveau ergeben, sodass die Autoren jeweils willkürlich erscheinende Priorisierungen zugunsten einer der beiden Fragen vornehmen. Als notdürftiger Ersatz für ansonsten oft fehlende Informationen mögen solche Verfahren zur Reduktion von Messfehlern geeignet sein; gleichzeitig stellt dieses Vorgehen aus prinzipieller Sicht ein Negativbeispiel für uns dar, denn warum sollten

die Antworten aus zwei zusammenhanglosen Fragen in einer einzigen Variablen zusammengefasst werden?

An diesen Schwierigkeiten zeigt sich, dass die KldB 2010 zur Systematisierung von Berufsbenennungen konzipiert wurde. Die Erfassung der beruflichen Tätigkeit von Erwerbstätigen ist hingegen nicht das originäre Ziel der KldB 2010. Zur Lösung dieses Problems und möglicherweise abweichend von der Intention der KldB 2010 interpretieren wir Berufsfachlichkeit und Anforderungsniveau als zusammenhängende Teildimensionen von „Beruf“. Sie begründen die Struktur der KldB 2010 und spiegeln sich daher in den Beschreibungen der Berufskategorien wider. Analog zur ISCO-08, wo den in Beschreibungen dargestellten „tasks involved“ Priorität gegenüber dem „skill level“ einer Kategorie eingeräumt wird, sind auch die Beschreibungen aus der KldB 2010 maßgeblich für unser Verfahren zur Berufskodierung.

Da die Beschreibungen der Kategorien für uns Priorität haben, können wir das Anforderungsniveau als Gliederungsprinzip weitestgehend ignorieren. Eine getrennte Erfassung des Anforderungsniveaus sollte üblicherweise nicht erforderlich sein. Bei Kenntnis eines vom Befragten im Beruf ausgeübten Tätigkeitsbündels sollten sich daraus – und nur daraus – beide Dimensionen ergeben und eine Zuordnung zu den Berufskategorien ermöglichen. Dieses Prinzip setzt voraus, dass Berufe, die sich in ihrem Anforderungsniveau unterscheiden (z.B. Altenpflegehelfer/in vs. Altenpfleger/in), auch hinreichend unterschiedliche Tätigkeitsbündel ausüben und diese Unterschiede aus den Beschreibungen der 5-stelligen Berufskategorien aus der KldB 2010 ersichtlich werden. Bei der Bearbeitung mussten wir jedoch feststellen, dass aus den teils sehr ähnlichen Beschreibungen von Kategorien nicht immer erkennbar ist, inwieweit sich die Tätigkeitsbündel der zugeordneten Berufe in wesentlichen Merkmalen unterscheiden. In zahlreichen Fällen konnten trotz Unterschieden beim Anforderungsniveau keine wesentlichen Unterschiede in den ausgeübten Tätigkeitsbündeln festgestellt werden. In solchen Ausnahmefällen wurden Folgefragen entwickelt, die das Anforderungsniveau explizit als nachgeordnete Dimension berücksichtigen. Das Prinzip, nur die Beschreibungen der Kategorien zu verwenden, ließ sich daher nicht durchgängig einhalten.

In der KldB 2010 und in der ISCO-08 liegen umfangreiche Beschreibungen bzw. Definitionen der Berufskategorien vor, die sich jeweils über mehrere hundert Seiten erstrecken. Das erklärte Ziel beider Klassifikationen ist, den Nutzer bei der Zuordnung von Berufsbezeichnungen in die Berufskategorien zu unterstützen, weshalb auf trennscharfe Beschreibungen der Kategorien untereinander geachtet worden sei. Für die Entwicklung der Hilfsklassifikation ist die Gültigkeit dieser Aussage eine wichtige Voraussetzung. Entsprechende Definitionen wurden in der ISCO seit der ersten Version von 1958 (ISCO-58, International Labour Office 1958) als unerlässlich angesehen. Als Begründung führt die ISCO-58 an, dass im internationalen Kontext Arbeitskräfte unterschiedliche Tätigkeiten ausüben können, obwohl ihre Berufe identische Bezeichnungen tragen. Nicht die Überschriften in Form von Berufsbezeichnungen, sondern erst die Definitionen der einzelnen Kategorien würden daher deutlich machen, welche

Tätigkeiten den jeweiligen Kategorien zugeordnet sind. Für die deutsche Klassifikation liegen entsprechende Beschreibungen erstmals in der Ausgabe von 1961 und erneut in der KldB 2010 vor. Sperling (1961) begründete für die damalige Ausgabe deren Notwendigkeit damit, dass die systematischen Einheiten künstlich geschaffene Zusammenfassungen mehrerer Berufe sind und daher Kriterien definiert werden müssen, anhand derer die Zugehörigkeit von Berufsbenennungen zur jeweiligen Einheit beurteilt werden kann.

Die übrigen Zeilen in Tabelle 3 betreffen spezielle Berufsgruppen (z.B. Führungskräfte, Aufsichtskräfte, …), die in beiden Klassifikationen jeweils gesondert behandelt werden. Unseren Umgang damit werden wir in Abschnitt 3.3 „Spezialfälle“ darlegen.

3 Leitgedanken zur Hilfsklassifikation

Berufsklassifikationen gliedern die berufliche Vielfalt in Form von Berufskategorien. Die Inhalte der einzelnen Berufskategorien sind (hauptsächlich) durch ihre jeweiligen Beschreibungen definiert. Zusätzlich stellen sie eine monohierarchische Systematik bereit, die die Berufskategorien nach Prinzipien der Ähnlichkeit zusammenfasst.

Eine wichtige Anwendung finden Berufsklassifikationen bei der statistischen Erfassung des Berufs von Erwerbstätigen. Genaue Handlungsanweisungen, wie dies zu erfolgen hat, sucht man in Berufsklassifikationen vergeblich. Die ISCO-08 beschränkt sich auf einige Empfehlungen, die im nationalen Kontext weiter ausgearbeitet werden müssen, und die KldB 2010 formuliert einige Regeln zur Klassifizierung von Berufsbenennungen. Darauf aufbauend ist die folgende Arbeitshypothese entstanden, was wir für den Idealfall halten, wie Erwerbstätige klassifiziert werden sollten.

Erwerbstätige sollten

a) anhand ihrer tatsächlich ausgeübten beruflichen Tätigkeit (und nicht etwa anhand der vorliegenden Antworten vom Befragten)
b) in die ihrer Definition nach am besten passende Berufskategorie auf der untersten hierarchischen Ebene zugeordnet werden („am besten passend“ lässt sich über die Ähnlichkeit von Berufsfachlichkeit, Anforderungsniveau bzw. skill näher definieren),
c) wobei im Zweifelsfall die vorherrschende Tätigkeit (Kerntätigkeit), die am meisten Arbeitszeit in Anspruch nimmt, für die Zuordnung ausschlaggebend ist (dies steht im expliziten Widerspruch zur ISCO-08, vgl. den nachfolgenden Abschnitt 3.3.3 „Berufe ohne Spezialisierung“).

Diese Zielsetzung erfordert einen Abgleich der beruflichen Tätigkeit des Befragten mit den Definitionen sämtlicher Berufskategorien, was entweder durch den Erwerbstätigen selbst oder durch einen Kodierer geschehen kann. Für ersteres ist es notwendig, dass Erwerbstätige die Definitionen der Berufskategorien kennen. Letzteres erfordert, dass sämtliche Befragte ihre berufliche Tätigkeit umfassend und mit allen Details beschreiben. Beides ist in Umfragen kaum realisierbar, weshalb in statistischen Erhebungen diesem Ideal üblicherweise nicht entsprochen werden kann.

Stattdessen ist die Berufskodierung von den verfügbaren Antworten abhängig und auch die Organisation des Kodierprozesses spielt eine wesentliche Rolle. Insbesondere die Nutzung von computergestützten Vorschlagssystemen, die Berufsbenennungen zur Kodierung vorschlagen, hat sich dabei durchgesetzt. Mit der Hilfsklassifikation haben wir eine Alternative entwickelt, die in Abschnitt 3.1 dargestellte Schwächen der bisherigen Vorschlagssysteme beheben soll.

In einer Annäherung an den Idealfall, wonach man komplette Definitionen von Berufskategorien vorschlagen sollte, haben wir mit der Hilfsklassifikation zum Vorschlagen besser geeignete Tätigkeitsbeschreibungen entwickelt, die die wichtigen Aspekte aus den Definitionen zusammenfassen. Die der Entwicklung zugrundeliegenden Überlegungen werden in den Abschnitten 3.2 und 3.3 beschrieben. Da wir eine Verwendung während des Interviews antizipieren, spielt es bei der Entwicklung keine Rolle, inwieweit die benötigten Informationen über die berufliche Tätigkeit tatsächlich vorliegen.

3.1 Defizite bisheriger Verfahren

Aus drei nachfolgend näher beschriebenen Gründen wurde die Entwicklung der Hilfsklassifikation erforderlich:

a) Eine simultane statistische Erfassung des Berufs anhand der KldB 2010 und anhand der ISCO-08 ist wünschenswert.

b) Berufsbenennungen sind unpräzise, zu allgemein oder mehrdeutig. Eine eindeutige Kommunikation über die ausgeübte Tätigkeit ist damit nicht immer möglich.

c) Wenn das computergestützte Vorschlagssystem zu viele ähnliche Berufsbenennungen enthält, wird es unübersichtlich und kann das Auffinden des passenden Berufs erschweren.

3.1.1 Simultane Kodierung nach KldB 2010 und ISCO-08

Üblicherweise ist es wünschenswert, berufliche Tätigkeiten von Befragten sowohl in die KldB 2010 als auch in die ISCO-08 zu kodieren, wodurch sich der Arbeitsaufwand verdoppelt. Zur Vereinfachung der Arbeit stellt die KldB 2010 einen Umsteigeschlüssel bereit, der Berufskategorien der KldB 2010 nach ISCO-08 verschlüsselt. Für 88% der Berufskategorien aus der KldB 2010 existieren eindeutige Umstiege in ISCO-08. Wenn wir aber die Umfragedaten von Schierholz et al. (2018), die in der KldB 2010 vorliegen, nach ISCO-08 umschlüsseln wollen, so erhalten wir lediglich für 78 Prozent der Befragten einen eindeutigen ISCO-Code. Für die restlichen Kategorien ordnet der Umsteigeschlüssel mehr als eine Kategorie zu und eine erneute Kodierung der zugrundeliegenden Freitextangaben wird erforderlich.

Unser Anliegen ist es, dass die berufliche Tätigkeit nur einmal kodiert werden muss. Zu diesem Zweck empfiehlt die ISCO-08 in Anlehnung an Hoffmann (1994) die Verwendung eines Kodier-Index, der für jede Berufsbenennung zwei Codes enthält und Berufsbenennungen damit zugleich der nationalen als auch der internationalen Berufsklassifikation zuordnet. Unsere Hilfsklassifikation übernimmt diese Funktion des

Kodier-Index und weist jeder Hilfskategorie zwei Codes zu. Der ISCO-08 zufolge können dafür zusätzliche Einträge im Kodier-Index notwendig sein, die die Unterschiede zwischen einzelnen Kategorien in beiden Klassifikationen deutlich machen. Dabei bezweifeln wir aber, dass einzelne Wörter (Berufsbenennungen) aus einem Kodier-Index die feinen Unterschiede zwischen einzelnen Kategorien entsprechend abbilden können. Stattdessen halten wir Hilfskategorien für erforderlich, die die bestehenden Berufskategorien feiner aufgliedern und auf diese Weise den charakteristischen Inhalt jeder möglichen Kombination von Berufskategorien aus der KldB 2010 und der ISCO-08 beschreiben.

3.1.2 Unpräzise Berufsbenennungen

Da Berufsbenennungen standardmäßig zur Kommunikation über Berufe und zur Berufskodierung verwendet werden, wollen wir zunächst argumentieren, warum die Benennungen uns zur Verwendung im Interview nicht geeignet erscheinen. Dabei ist hervorzuheben, dass die folgenden Argumente einseitig sind und nicht für alle Berufe gelten. Unpräzise Berufsbenennungen führen aber zu erheblichen Schwierigkeiten bei der möglichst exakten Messung des Berufs und begründen, warum die Hilfsklassifikation entwickelt wurde.

Es wurde bereits gezeigt, dass heutige Berufsbenennungen oft nicht mehr anschaulich die Tätigkeit beschreiben, sondern ihre Aussagekraft nachgelassen hat. Folgende Berufsbenennungen, die der aktuellen Berufsdatenbank der Bundesagentur für Arbeit (vgl. Paulus/Matthes 2013) entnommen wurden[1], bestätigen dies:

Maschineneinrichter/in (spanlose Metallbearbeitung); Maschineneinrichter/in (Zerspanungstechnik); Helfer/in – Metalloberflächenbearbeitung; Helfer/in – Metallbearbeitung; Verfahrensmech. – Hütten-/Halbzeugindustrie Stahl-Umformung; Techniker/in – Maschinentechnik (Anlagentechnik); Techni-ker/in – Maschinentechnik (Automatisierungstechnik); Laminierer/in (Kunststoffverarbeitung); Helfer/in – Kunststoff, Kautschuk; Automatenfachmann/-frau (ohne Fachrichtungen); Automatenfachmann/-frau – Automatenmechatronik

Diese Berufsbenennungen wurden ausgewählt, da sie alle aus mehreren Wörtern bestehen. Offenbar gelingt es für die aufgeführten Berufe nicht, eine aus einem einzigen Wort bestehende, aussagekräftige Benennung zu finden, sondern es werden Schlagworte aneinandergereiht. Dies deutet darauf hin, dass kurze, zusammenhängende Texte eine bessere Darstellungsform sein könnten. Ein Weglassen der hinteren Wörter ist keine Option, denn sie sind zur Präzisierung erforderlich und die KldB 2010 ordnet alle aufgelisteten Benennungen unterschiedlichen Berufskategorien zu.

[1] Berufsbenennungen werden in diesem Text mit Schrägstrich getrennt (z.B. Maschineneinrichter/in), sofern dies der Originalschreibweise aus den jeweils referenzierten Datenbanken oder Klassifikation entspricht.

Als Antwortoptionen im Interview sind derartige Berufsbenennungen wenig geeignet. Mit den Erläuterungen in Klammern bzw. hinter dem Bindestrich stehen Interviewer vor der Herausforderung, wie sie derartige Texte nach den Vorgaben des standardisierten Interviews vorlesen sollen. Ohne ausführliches Training der Interviewer ist zu befürchten, dass diese die entsprechenden Zusätze beim Vorlesen einfach ignorieren. Selbst wenn die vollständigen Berufsbenennungen vorgelesen werden, sind die Unterschiede zwischen einzelnen Berufsbenennungen für den Laien nicht unmittelbar verständlich. Für ein erfolgreiches Interview ist es aber wichtig, dass Interviewer und Befragte die Essenz der Antwortoptionen möglichst einfach verstehen können, was bei den angegebenen Beispielen zu Problemen führen kann.

Wenn man im Interview mit einer offenen Frage nach der „ausgeübten Tätigkeit“ fragt, antworten viele Befragte mit einer Berufsbenennung. Drei Argumente deuten darauf hin, dass solche Berufsbenennungen die ausgeübte Tätigkeit nicht bestmöglich beschreiben:

a) Einige häufige Antworten auf die Frage nach der beruflichen Tätigkeit sind:

   Bürokaufmann; Kaufmännischer Angestellter; Maschinenbediener; Projektleiter; Sachbearbeiter; Verkäufer; Verwaltungsangestellter; Wissenschaftlicher Mitarbeiter

   Viele Befragte aus diesen Berufen präzisieren diese Angabe und nennen beispielsweise die Branche, ohne dass sie explizit dazu aufgefordert wurden. Da die Tätigkeiten der Befragten trotz gleicher Berufsbenennung sehr unterschiedlich sein können und auch häufig unterschiedlich kodiert werden, sind zusätzliche Angaben zur Kodierung dringend erforderlich.

b) Stooß und Saterdag (1979) und Dostal, Schade und Parmentier (1999) weisen darauf hin, dass die Befragten nicht notwendigerweise die tatsächlich ausgeübte Tätigkeit beschreiben, sondern ihr berufliches Selbstverständnis auch anhand des Ausbildungsabschlusses („Volkswirt“), statusbezogener Bezeichnungen („Abteilungsleiter“, „Regierungsrat“) und anderer Begrifflichkeiten ausdrücken können, je nachdem was dem Befragten im Moment der Befragung gerade wichtig erschien.

c) Auch muss dem Befragten eine passende Berufsbenennung spontan in den Sinn kommen, weshalb möglicherweise eine Bezeichnung der Berufsausbildung oder aus dem Arbeitsvertrag genannt wird, obwohl sich die aktuell ausgeübte Tätigkeit in der Zwischenzeit geändert hat und eine andere Berufsbenennung die ausgeübte Tätigkeit passender beschreibt. Ein Befragter antwortet beispielsweise „Pilot“ (KldB-Berufskategorie 52313) obwohl er inzwischen zusätzlich an der Organisation des Flugbetriebs mitwirkt („Chefpilot“, KldB-Berufskategorie 52314). Oder die Antwort lautet „Rettungsassistent“ (KldB-Berufskategorie 81324) obwohl „Rettungsassistent – Fahrdienst“ oder „Lehrrettungsassistent“ (KldB-Berufskategorien 52182 bzw. 84213) passender wäre. In ähnlicher Weise existieren in der KldB 2010 noch viele andere Berufsbenennungen, die sich weiter spezifizieren lassen und dann zu unterschiedlichen Kodierungen führen.

Aus diesen Gründen ist eine vom Befragten genannte Berufsbenennung indikativ für die ausgeübte Tätigkeit, aber eine genauere Präzisierung, wie sie bislang häufig mit einer zweiten offenen Frage versucht wird, ist meist wünschenswert.

Diese zusätzlichen Informationen finden praktische Verwendung bei der Berufskodierung. Schierholz et al. (2018) berichten von 418 Befragten (39 Prozent der Berufstätigen aus der Studie), deren Berufsbenennungen aus dem Interview identisch zu Berufsbenennungen aus dem alphabetischen Verzeichnis der KldB 2010 sind und sich auf diese Weise vollautomatisch kodieren lassen. Zusätzlich wurden zwei professionelle Kodierer mit der Kodierung beauftragt. Für 27 Prozent der automatisch kodierbaren Antworten vergibt mindestens einer der beiden Kodierer einen Code, der vom alphabetischen Verzeichnis abweicht. Offensichtlich befolgen die Kodierer den paradigmatischen Kodierablauf nicht durchgängig und verwenden stattdessen in einigen Fällen weitere Informationen. Dies bestätigt unsere These, dass Berufsbenennungen oft Interpretationsspielraum lassen und eine eindeutige Kodierung der zugrundeliegenden beruflichen Tätigkeit anhand von Benennungen nicht immer möglich ist.

Computergestützte Vorschlagssysteme erlauben die Eingabe eines Begriffs und lassen den Kodierer bzw. den Befragten eine Berufsbenennung aus einer kurzen Liste auswählen. Falls dabei die ursprünglich eingegebene Berufsbenennung zur Auswahl steht, ist es naheliegend diese auch auszuwählen – obwohl ggf. zusätzliche Informationen zur Verfügung stehen, die eine andere Berufsbenennung plausibler erscheinen lassen. Zur Vermeidung eines derartigen Confirmation Bias (Gilbert/Krull/Malone 1990) wollen wir anstelle der Berufsbenennungen Beschreibungen von Tätigkeiten einblenden und auf diese Weise eine bloße Wiederholung der zuvor genannten Berufsbenennung vermeiden.

Erfahrungen von Schierholz et al. (2018) zeigen, dass Berufsbenennungen manchmal zutreffend erscheinen, aber unpassenden Berufskategorien zugeordnet sind. Beispielsweise antwortete eine Person im Interview, dass sie „Zimmermädchen“ ist und ihre Aufgaben „Betten machen, Zimmer herrichten“ sind. Nachfolgend wählte diese Person, vermutlich aus Mangel an besseren Alternativen, die Antwortoption „Helferin/Helfer – Reinigung“ aus. Diese Berufsbenennung ist in der KldB 2010 der Berufskategorie „54101 Berufe in der Reinigung (ohne Spezialisierung) – Helfer-/Anlerntätigkeiten“ zugeordnet. Passender für diese Person wäre jedoch die Berufskategorie „63221 Berufe im Hotelservice – Helfer-/Anlerntätigkeiten“ gewesen. Wenn also die Antwortoption „Helferin/Helfer – Reinigung“ ausgewählt wird, ist nicht garantiert, dass die zugeordnete Berufskategorie 54101 für die ausgeübte Tätigkeit tatsächlich zutrifft. Stattdessen halten wir es für nötig, die Benennung „Helferin/Helfer – Reinigung“ mithilfe einer Beschreibung zu präzisieren, sodass nur Personen diese Antwortoption auswählen, wenn die Kategorie 54101 auch tatsächlich zutrifft.

Zusammengefasst können wir festhalten: Berufsbenennungen sind unpräzise. Wir wissen nicht, ob Befragte im Interview die passendste Bezeichnung für ihre berufliche Tätigkeit nennen, oder bloß eine grob zutreffende Beschreibung geben. In der Praxis

verwenden Kodierer daher zusätzliche Informationen des Befragten. Dabei können für die berufliche Tätigkeit eines Befragten mehrere Berufsbenennungen aus dem Kodier-Index möglich sein, die aber regelmäßig unterschiedlichen Berufskategorien zugeordnet sind. Nach welchen Prinzipien die Berufsbenennungen in Kategorien zusammengefasst wurden und wo die Grenzen zwischen einzelnen Berufskategorien liegen, wird anhand von Berufsbenennungen aber nur ungenau zum Ausdruck gebracht. Es ist daher leicht möglich, dass einzelne Berufsbenennungen einen passenden Eindruck erwecken, aber die zugeordnete Berufskategorie für die ausgeübte Tätigkeit nicht korrekt ist (siehe Zimmermädchen-Beispiel). Dies ist nicht verwunderlich, denn in erster Linie stellen Berufsbenennungen Hilfsmittel zur alltäglichen Kommunikation dar und wurden nicht für wissenschaftliche Zwecke entwickelt. Diese Erkenntnisse zu Berufsbenennungen lassen uns zweifeln, ob der paradigmatische Kodierablauf zu einer genauen Erfassung des Berufs geeignet ist. Alternativ erscheint es sinnvoll, dass eine passende Berufskategorie möglichst direkt ohne den Umweg über Berufsbenennungen aus der Klassifikation ausgewählt wird. Die Hilfsklassifikation soll dies unterstützen.

3.1.3 Übersichtliche Darstellung

Computergestützte Vorschlagssysteme wurden entwickelt, um mögliche Berufsbenennungen übersichtlich darzustellen. Dazu ist es sinnvoll, nicht alle denkbaren Berufsbenennungen anzuzeigen, sondern bloß die relevanten Benennungen im System zu verwenden. Es folgen einige Beispiele von insgesamt 53 Berufsbenennungen, die die KldB 2010 der Berufskategorie „84214 Lehrkräfte in der Sekundarstufe – hoch komplexe Tätigkeiten“ zuordnet.

Lehrer/in (Uni) – Gesamtschulen, Fachlehrer/in – Waldorfschulen, Klassenlehrer/in – Waldorfschulen, Lehrer/in – Gymnasien (Sekundarstufe I und II), Biologielehrer/in, Deutschlehrer/in, Englischlehrer/in, Gymnasiallehrer/in, Hauptschullehrer/in, Lehrer/in – Regelschulen (Hauptschule)

Anstatt zahlreiche nahezu identische Berufsbenennungen einzublenden, ist es sinnvoller, alles in nur einer Hilfskategorie zusammenzufassen, die die Bezeichnung „Lehrer/in in der Sekundarstufe“ tragen könnte. Zwar ließe sich argumentieren, dass beispielsweise Hauptschullehrer/innen und Gymnasiallehrer/innen unterschiedliche Berufe ausüben und daher getrennt voneinander erfasst werden müssten, doch die vorgenommene Zusammenlegung dieser Berufe folgt dem Vorgehen der KldB 2010, in der ebenfalls keine feingliedrigere Unterteilung vorgenommen wird. Die Zusammenfassung ermöglicht daher, Berufe in der gleichen Genauigkeit zu erfassen wie dies in der KldB 2010 geschieht. Auf diese Weise bleibt Platz, bei Eingabe des Suchwortes „Lehrer“ noch weitere Hilfskategorien mit Bezeichnungen wie „Förderschullehrer/in“, „Grundschullehrer/in“ und „Schulleiter/in“ übersichtlich zur Auswahl darzustellen. Diese sind in der KldB 2010 anderen Berufskategorien zugeordnet und könnten bei zu vielen Auswahlmöglichkeiten leicht übersehen werden.Offensichtlich kann auch mehr als eine Berufsbenennung für die Tätigkeit eines Befragten zutreffen, was bei

der Auswahl zu Verwirrung führen kann. Dieser Schwierigkeit lässt sich entgegenwirken, indem man die Berufsbenennungen in einer Berufskategorie zusammenfasst und ihr gemeinsames, charakteristisches Merkmal beschreibt.

3.2 Vorgehen bei der Entwicklung

Besonderes Augenmerk wird in der neuen Hilfsklassifikation auf prägnant beschriebene Kategorien gelegt, die möglichst disjunkt zueinander sind und sich eindeutig voneinander abgrenzen lassen. Zunächst verdeutlichen wir dies anhand von zwei beispielhaften Hilfskategorien aus der neuen Hilfsklassifikation und beschreiben nachfolgend die zugrundeliegenden Überlegungen. Zusätzliche Details zum Entwicklungsprozess stellen wir im elektronischen Anhang bereit.

Tabelle 4
Zwei beispielhafte Hilfskategorien

Berufsbezeichnung: Callcenteragent/in (Inbound) **Tätigkeit**: Bearbeitung von Anfragen, Aufträgen oder Reklamationen (Inbound), z.B. per Telefon oder E-Mail **Tätigkeitsbeschreibung**: z.B. Kundenanfragen beantworten und Kunden über verschiedene Produkte beraten; Aufträge in Computersystemen erfassen; Informationen an Kunden versenden **KldB 2010**: 92122 Berufe im Dialogmarketing – fachlich ausgerichtete Tätigkeiten **ISCO-08**: 4222 Contact centre information clerks **Abgrenzungen**: Verkäufer/in – Telemarketing, Telefonist/in, Teamleiter/in – Callcenter, Schalterauskunft
Berufsbezeichnung: Verkäufer/in – Telemarketing **Tätigkeit**: Kontaktaufnahme mit Kunden, meist per Telefon, um Waren und Dienstleistungen zu verkaufen **Tätigkeitsbeschreibung**: z.B. Produkte per Telefon oder E-Mail bewerben; Info-Broschüren und Waren versenden; Termine mit Handelsvertretern vereinbaren; Marketing-Datenbanken aktualisieren **KldB 2010**: 92122 Berufe im Dialogmarketing – fachlich ausgerichtete Tätigkeiten **ISCO-08**: 5244 Contact centre salespersons **Abgrenzungen**: Callcenteragent/in (Inbound), Teamleiter/in – Callcenter

Quelle: Hilfsklassifikation

3.2.1 Disjunkte Hilfskategorien und eindeutige Zuordnung

Um den Befragten eine eindeutige Auswahl zu ermöglichen, sollen die Antwortoptionen in Umfragen üblicherweise paarweise disjunkt sein. In gleicher Weise haben die KldB 2010 und die ISCO-08 den Anspruch, dass jeder Beruf genau einer einzigen Kategorie zugeordnet werden kann. Auch bei der Erstellung der Hilfsklassifikation verfolgen wir dieses Ziel.

Aus prinzipieller Sicht wird dies über die Konstruktion der Hilfsklassifikation erreicht: Sei A_i die i-te Berufskategorie aus der KldB 2010 und sei B_j die j-te Berufskategorie aus der ISCO-08. Die Schnittmenge dieser beiden Mengen definiert den Inhalt unserer neuen Hilfskategorie, $C_{ij} := A_i \cap B_j$, die also alle Berufe/Tätigkeitsbündel enthält, die sowohl Element von A_i als auch Element von B_j sind. Unter der Voraussetzung jeweils paarweise disjunkter Berufskategorien in beiden offiziellen Klassifikationen

sind auch die derart konstruierten Hilfskategorien wieder paarweise disjunkt und folgen den Grenzen der zugrundeliegenden Berufskategorien. Weiterhin kann man aufgrund dieser Definition folgern, dass für einen Befragten, der in Hilfskategorie C_{ij} kodiert wurde, die Berufskategorien A_i und B_j zutreffen.

Im dargestellten Beispiel entspricht die Hilfskategorie des „Callcenteragent/in (Inbound)“ der Schnittmenge aus der KldB 2010-Kategorie „92122 Dialogmarketing – Fachkraft“ und der ISCO-Kategorie „4222 Kundeninformationsfachkraft im Callcenter“. Das Beispiel „Verkäufer/in – Telemarketing“ entstand aus der Schnittmenge von „92122 Dialogmarketing – Fachkraft“ und „5244 Telefonverkäufer“. Die Berufskategorie 92122 muss für unsere Hilfsklassifikation aufgeteilt werden, denn die ISCO-08 unterscheidet Mitarbeiter von Callcentern danach, ob sie für die Kundeninformation oder für den Verkauf zuständig sind.

Obwohl sich der Inhalt jeder Hilfskategorie mathematisch einfach definieren lässt, ist die Identifikation und Beschreibung des Inhalts in praktischer Hinsicht die eigentliche Herausforderung bei der Erstellung der Hilfsklassifikation. Im Gegensatz zur präzisen Sprache der Mathematik, wo sich Mengen eindeutig und paarweise disjunkt definieren lassen, sind Berufskategorien und auch unsere Hilfskategorien in Worten umschrieben, die Spielraum für Interpretationen lassen. Dostal (2002) beschreibt zutreffend, dass Berufe „ausgefranst“ sind, sie also einen „Kernbereich von konstituierenden Elementen [haben], der durch einen Randbereich von optionalen Zusatzelementen eingehüllt wird.“ Zur Identifikation des Kernbereichs bzw. des charakteristischen Aufgabenbündels der einzelnen Berufskategorien verwenden wir im Regelfall die Überschrift und die einleitenden Sätze der jeweiligen Definition. Die Definitionen von ähnlichen Berufskategorien berücksichtigen wir, um die Grenzen zu anderen Berufskategorien zu identifizieren, sofern diese Grenzen nicht explizit in den „Notes“ von ISCO-08-Kategorien genannt sind. Soweit möglich stellen wir auf diese Weise sicher, dass wir die Unterschiede zwischen den beschriebenen Tätigkeiten verstehen und sie entsprechend als disjunkte Tätigkeiten interpretieren. Bei Bedarf ziehen wir ggf. weitere Informationen wie die Berufsbenennungen, das BERUFENET der Bundesagentur für Arbeit oder anderes verwandtes Material zurate, um eine bessere Vorstellung der zugeordneten Berufe und der dort ausgeübten Tätigkeiten zu erhalten.

In der KldB 2010 kommt es aber auch vor, dass sich die in den genannten Dokumenten beschriebenen Tätigkeiten gar nicht oder nur sehr geringfügig zwischen mehreren Berufskategorien unterscheiden. Insbesondere wenn Berufskategorien der KldB 2010 sich nur im Anforderungsniveau unterscheiden, weisen die beschriebenen Tätigkeiten oft eine sehr hohe Ähnlichkeit auf. Wenn wir in solchen Fällen keine disjunkten Tätigkeiten für die Hilfsklassifikation ableiten können, betrachten wir die zugrundeliegenden Berufskategorien gemeinsam als ob nur eine einzige Tätigkeit beschrieben worden wäre und erzeugen nur eine einzige Hilfskategorie. Die Hilfskategorie ist dann mehreren Berufskategorien derselben Klassifikation zugeordnet. Um dennoch eine eindeutige Zuordnung zu erreichen, formulieren wir zusätzliche Folgefragen, die

spezifisch auf die jeweilige Hilfskategorie und das trennende Kriterium der zugrundeliegenden Kategorien zugeschnitten sind. Sie orientieren sich regelmäßig an der Definition des Anforderungsniveaus, oft operationalisiert über die üblicherweise erforderliche Ausbildung wie sie im BERUFENET erfasst ist, und ermöglichen eine eindeutige Zuordnung auf Basis dieser zweiten Dimension.

3.2.2 Verständliche Darstellung der Hilfskategorien

Befragte, Interviewer und Kodierer sollen mit der Hilfsklassifikation ein Instrument erhalten, mit dem die Auswahl der passenden Tätigkeit möglichst einfach wird. Dazu müssen die Inhalte der Hilfskategorien kurz und übersichtlich dargestellt sowie einfach verständlich sein. Gleichzeitig ist eine präzise Sprache notwendig, denn wenn Antwortoptionen nur in vagen, ungenauen Begriffen beschrieben werden, sind Antworten nicht disjunkt und es fällt in Grenzfällen besonders schwer, die am besten passende Kategorie zu bestimmen. Aus diesem Grund empfehlen Tourangeau, Rips und Rasinski (2000), soweit wie möglich die wichtigen Begriffe im Fragetext in Form von Hinweisen näher zu erläutern, um auf diese Weise sicherzustellen, dass alle Befragten die Frage möglichst gleich verstehen.

Inhaltlich sind unsere Hilfskategorien, wie beschrieben, Schnittmengen von Berufskategorien aus der KldB 2010 und aus der ISCO-08 und daher Zusammenfassungen von mehreren, einander ähnlichen Berufen. Beruf selbst wurde definiert als „Bündel von Tätigkeiten“ und entsprechend enthält auch jede Hilfskategorie wieder zahlreiche miteinander zusammenhängende Tätigkeiten, die aber in Berufsklassifikationen nie vollständig aufgelistet werden. Für die einzelnen Elemente einer Hilfskategorie ist es auch keineswegs nötig, dass in jedem Einzelfall alle beispielhaft aufgeführten Tätigkeiten tatsächlich ausgeübt werden. Die Beschreibungen der Kategorien sind daher zwangsweise vage gehalten. Wir können bloß versuchen, die Kernbereiche der einzelnen Hilfskategorien möglichst präzise zu benennen und dabei die wichtigsten Aspekte aus den Definitionen der Berufskategorien zu übernehmen. Wie auch in den Berufskategorien müssen die Randbereiche von Hilfskategorien aber vage bleiben.

Für die Erstellung der Hilfsklassifikation sind die Definitionen der zugrundeliegenden Berufskategorien eine entscheidende Hilfe und viele Beschreibungen können wir wortwörtlich übernehmen. Aus einer gewissen Sicht wäre es ideal, wenn die kompletten Definitionen der Berufskategorien im Interview vollständig vorgelesen würden bzw. bei der Kodierung Verwendung fänden, denn diese Definitionen beschreiben am ausführlichsten, nach welchen Kriterien die Berufskategorien einzelne Berufe zusammenfassen und was Teil der jeweiligen Berufskategorie ist. Die Definitionen sind allerdings lang und ausführlich. Für Kodierer wäre es zu mühsam, wenn sie immer in den Definitionen nachschlagen müssten. Ein vollständiges Vorlesen der Definitionen im Interview verbietet sich von selbst. Unsere Hilfskategorien sollen daher eine Kurzfassung für Nutzungsformen darstellen, in denen die vollständigen Definitionen zu lang sind.

Zur Darstellung der Hilfskategorien verwenden wir folgende Felder, die eine Verwendung der Hilfsklassifikation zur Kodierung während des Interviews ermöglichen sollen:

a) *Berufsbezeichnung*: Diese Bezeichnung soll auf einen Blick zeigen, ob die Hilfskategorie möglicherweise zutreffend ist und ggf. vom Interviewer vorgelesen werden. Als alleinige Entscheidungsgrundlage halten wir die Berufsbezeichnung nicht für ausreichend, denn Inhalte von Hilfskategorien lassen sich nicht immer in einer Berufsbezeichnung abkürzen. Der Interviewer kann anhand dieser Bezeichnung bewerten, ob er die zugehörige Tätigkeit vorlesen möchte.

b) *Tätigkeit* (oft im Sinne einer zweckgerichteten Aufgabe oder Funktion): Die Tätigkeit soll dem Befragten vorgelesen werden. Im Regelfall kann anhand dieses Textes entschieden werden, ob die Hilfskategorie für die berufliche Tätigkeit des Befragten zutrifft. Sie beschreibt den charakteristischen Inhalt und die Wesensart (häufig unter Verwendung des Berufszwecks) der zugeordneten Berufe in der Hilfskategorie.

c) *Tätigkeitsbeschreibung*: Die Tätigkeitsbeschreibung wird im Regelfall nicht benötigt, aber unterstützt bei Rückfragen und dient der weiteren Präzisierung der Inhalte der Berufskategorie. Sie enthält beispielhaft einige Aktivitäten, die Personen in den zugeordneten Berufen üblicherweise ausüben.

d) *Abgrenzung*: Nach Möglichkeit sollten alle eventuell passenden Hilfskategorien zur Auswahl vorgeschlagen werden. Erst wer die Alternativen kennt, kann entscheiden, was am passendsten ist. Zu diesem Zweck geben Abgrenzungen an, welche anderen Hilfskategorien ebenfalls vorgeschlagen werden sollen, wenn eine Hilfskategorie mit hoher Wahrscheinlichkeit ausgewählt werden wird. Bei Abgrenzungen vom Typ „hoch“ ist die Ähnlichkeit besonders groß, sodass diese Kategorien zwingend vorgeschlagen werden müssen. Abgrenzungen vom Typ „mittel“ sind zwar ebenfalls ähnlich, aber zur Reduktion der Anzahl vorgeschlagener Kategorien und damit einhergehender erhöhter Übersichtlichkeit sind sie nicht zwingend vorzuschlagen.

Die Inhalte der Hilfskategorien C_{ij} sollten im Optimalfall identisch mit der Schnittmenge der zugrundeliegenden Berufskategorien, $C_{ij} = A_i \cap B_j$, sein. Wenn wir aber in wenigen Worten beschreiben, was in den Definitionen von A_i und B_j ausführlich dargestellt wird, muss es zwangsläufig zu Abweichungen kommen. Dies birgt die Gefahr, dass eine Hilfskategorie mehr Elemente als die Schnittmenge enthalten kann, $C_{ij} \supset A_i \cap B_j$. In diesem Fall wäre es falsch, Elemente aus C_{ij} automatisch auch den Berufskategorien A_i und B_j zuzuordnen (vgl. oben aufgeführtes Beispiel zum „Helfer/in – Reinigung“). Die Vermeidung derartiger Fehlkodierungen hat für uns höchste Priorität, damit die Hilfsklassifikation eine geeignete Hilfestellung bei der Kodierung bietet. Dazu muss der Text bei Tätigkeit derart eingeschränkt, kleinteilig und präzise formuliert werden, sodass diese Hilfskategorie in Grenzfällen nicht ausgewählt wird. Auf diese Weise verkleinert sich C_{ij} zu einer Teilmenge von $A_i \cap B_j$. Dies impliziert zugleich, dass die Gesamtheit aller Berufe in der Hilfsklassifikation weniger vollständig enthalten sein wird als in den offiziellen Klassifikationen. Bei der Zuordnung von

Grenzfällen sollten entsprechend stets die ausführlichen Definitionen der zugrundeliegenden Berufskategorien zurate gezogen werden.

Einige Berufskategorien sind einander sehr ähnlich und unterscheiden sich nur hinsichtlich eines einzigen Merkmals. Mit den bisher vorgestellten Möglichkeiten bräuchten wir dafür zwei Hilfskategorien:

Tätigkeit bei Hilfskategorie 1: Führungstätigkeit mit Personalverantwortung und strategischer Entscheidungsbefugnis im Warenlager
Tätigkeit bei Hilfskategorie 2: Führungstätigkeit mit Personalverantwortung aber ohne strategische Entscheidungsbefugnis im Warenlager

Um derart umständliche Formulierungen zu vermeiden, erlaubt die Hilfsklassifikation Folgefragen. Die Tätigkeiten beider Berufskategorien werden dabei zusammengefasst und erst mithilfe der Folgefrage wird das zusätzliche Merkmal abgefragt. Die vollständige Hilfskategorie ist dann wie folgt:

Berufsbezeichnung: Lagerleiter/in
Tätigkeit: Führungstätigkeit mit Personalverantwortung im Warenlager
Tätigkeitsbeschreibung: z.B. Wareneingang und Warenausgang überwachen; Qualitätskontrollen durchführen; Personaleinsätze planen und Fortbildungen organisieren
Folgefrage: Sind Sie befugt strategische Entscheidungen zu treffen, z.B. zur Einführung neuer Verfahren, zu finanziellen Investitionen oder zur Einstellung und Entlassung von Personal?
Antwort 1: Ja -> Zuordnung zur Kldb-Kategorie 51394 und zur ISCO-Kategorie 1324
Antwort 2: Nein -> Zuordnung zur Kldb-Kategorie 51393 und zur ISCO-Kategorie 4321
Default: Hier wird festgelegt, welche Zuordnung erfolgen soll, wenn keine Antwort gegeben wurde.
Abgrenzungen: Verladeaufseher/in, Logistikleiter/in

Folgefragen sollen die Verständlichkeit erhöhen. Anstatt in einer Antwortoption abzufragen, ob die Eigenschaften A (Führungstätigkeit mit Personalverantwortung) und B (im Warenlager) und C (strategische Entscheidungsbefugnis) zutreffen, fragen wir in einem ersten Schritt nur nach A und B und erst in einer Folgefrage nach C. Folgefragen sind insbesondere dann hilfreich, wenn sie eine zusätzliche Dimension (z.B. Art der Führungstätigkeit, Anforderungsniveau) betreffen. Sie erlauben es umfassender zu beschreiben, was gemeint ist, ohne Befragte mit übermäßig langen und verschachtelten Antworttexten zu verwirren.

3.3 Spezialfälle

Nicht alle Berufskategorien konnten anhand der dargestellten, allgemeinen Prinzipien bearbeitet werden. Insbesondere unser Umgang mit speziellen Berufsgruppen, die in

beiden Berufsklassifikationen hervorgehoben werden (vgl. Tabelle 2), bedarf einer näheren Erläuterung.

3.3.1 Militärberufe

Laut der ISCO-08 sind alle Angehörigen der regulären Streitkräfte in speziell für das Militär vorgesehenen Berufskategorien zu erfassen. Dies umfasst alle Personen, die aufgrund ihres Berufs zu militärischem Gehorsam verpflichtet sind. Auch Beschäftigte, die nicht-militärische Tätigkeiten beim Militär ausüben (z.B. Arzt, Koch, Sekretär, LKW-Fahrer), sollen entsprechend in die militärischen Berufskategorien kodiert werden. Offensichtlich überschneiden sich die Tätigkeiten von zivilen Berufskategorien mit den militärischen Kategorien. Daher ist es leicht möglich, dass beispielsweise ein Stabsarzt sich als Arzt ausgibt und auf diese Weise fälschlicherweise einer zivilen Kategorie zugeordnet wird. Zur korrekten Kodierung ist stets Wissen darüber erforderlich, ob die Tätigkeit beim Militär ausgeübt wird. In Anlehnung an die ISCO-08 verfolgt die KldB 2010 das gleiche Prinzip.

Bei der Entwicklung der Hilfsklassifikation haben wir versucht, die Hilfskategorien so präzise wie möglich zu beschreiben, um Fehlzuordnungen auszuschließen. Diesem Prinzip zufolge müssten wir bei zahlreichen zivilen Berufskategorien explizit dazuschreiben, dass Angehörige des Militärs dort ausgeschlossen sind. Beispielsweise könnte die Tätigkeit in der Hilfskategorie eines Allgemeinarztes beschrieben werden als: „Untersuchung von Patienten und Diagnose von Krankheiten im nicht-militärischen Bereich“. Dem Stabsarzt würde auf diese Weise explizit mitgeteilt, dass diese Hilfskategorie für ihn nicht zutreffend ist. Da dies aber die Verständlichkeit der Hilfskategorien stark beeinträchtigen würde, haben wir dies nicht umgesetzt.

Unsere Hilfskategorie für „Angehörige der regulären Streitkräfte“ ist daher nicht in der eigentlich gewünschten Weise disjunkt zu den anderen Hilfskategorien. Wenn eine korrekte Kodierung von Militärangehörigen erforderlich ist, empfehlen wir, wie in vielen Umfragen üblich, die Mitgliedschaft beim Militär mithilfe einer Frage nach der beruflichen Stellung abzufragen. In einer Folgefrage kann dann der Dienstgrad erhoben werden, wie er für die Kodierung nach der ISCO-08 bzw. nach der KldB 2010 benötigt wird. Eine offene Abfrage des Berufs und die Nutzung der Hilfsklassifikation zur Kodierung von Militärangehörigen sollte nach Möglichkeit vermieden werden.

3.3.2 Aufsichts- und Führungskräfte

Für Personen, deren Aufsichts- oder Führungstätigkeiten die berufsfachlichen Aufgaben dominieren, enthalten Berufsklassifikationen spezielle Berufskategorien. In der ISCO-08 und in der KldB 2010 werden drei spezifische Leitungsfunktionen unterschieden:

a) Geschäftsführer und Vorstände koordinieren die Gesamtaktivitäten eines Unternehmens. Sie sind in der ISCO-08 der Berufskategorie „1120 Managing Directors and Chief Executives“ und in der KldB 2010 der Berufskategorie „71104“ zugeordnet.

b) ISCO-08 zufolge tragen Führungskräfte die Verantwortung für die strategische und operative Ausrichtung ihrer Organisationseinheit, für das Budget und/oder für die Einstellung und Entlassung von Personal. Führungskräfte sind in ISCO-08 der Major Group 1 „Managers“ zugeordnet und tragen in der KldB 2010 die Endziffern „94“.

c) Im Gegensatz zu Führungskräften haben Aufsichtskräfte einer Definition aus ISCO-08 zufolge keine Entscheidungsbefugnis beim Budget oder bei der Einstellung und Entlassung von Personal. Stattdessen sind sie verantwortlich, die Aktivitäten anderer Beschäftigter zu beaufsichtigen. In der ISCO-08 gibt es sechs Berufskategorien für Aufsichtskräfte: „3121 Mining Supervisors“, “3122 Manufacturing Supervisors“, “3123 Construction Supervisors”, “3341 Office Supervisors”, “5151 Cleaning and Housekeeping Supervisors in Offices, Hotels and Other Establishments” und “5222 Shop Supervisors”. Aufsichtskräfte in anderen als den genannten Bereichen werden ISCO-08 zufolge den gleichen Berufskategorien zugeordnet, in denen auch die jeweils beaufsichtigten Beschäftigten klassifiziert sind. In der KldB 2010 sind zahlreiche Berufskategorien für Aufsichtskräfte enthalten, die die Endziffern „93“ tragen.

Für Geschäftsführer und Vorstände sowie für Führungskräfte verweist die KldB 2010 explizit auf die Definitionen aus der ISCO-08. Entsprechend sind die jeweiligen Definitionen in ISCO-08 und in der KldB 2010 äquivalent. Für Aufsichtskräfte nehmen wir ebenfalls eine Äquivalenz an, obwohl die KldB 2010 dies nicht explizit schreibt.

In vielen Umfragen ist es üblich, mit geschlossenen Fragen zu erfassen, ob eine Person Aufsichtstätigkeiten ausübt. Als Faustregel gilt manchmal, dass Aufsichts- oder Führungskraft ist, wer Personalverantwortung für mindestens zehn Personen hat. Dies lässt sich in Umfragen mittels folgender Fragen erfassen: „Gehört es zu Ihren beruflichen Aufgaben, die Arbeit anderer Arbeitskräfte zu beaufsichtigen oder ihnen zu sagen, was sie tun müssen?“ und falls „Ja“: „Wie viele andere Arbeitskräfte beaufsichtigen Sie direkt?“ In einer von Schierholz et al. (2018) durchgeführten Umfrage antworteten 132 von 1031 Befragten (13%), dass sie mindestens zehn Arbeitskräfte beaufsichtigen. Allerdings stimmt dies oft nicht mit der Kodierung von professionellen Berufskodierern überein: Bloß 49 (37%) derjenigen Personen, die nach eigenen Angaben mindestens zehn Arbeitskräfte beaufsichtigen, werden auch von professionellen Berufskodierern als Aufsichts- oder Führungskraft kodiert. Beide Konzepte sind offenbar nicht deckungsgleich und die oben genannte Faustregel erscheint uns vor diesem Hintergrund fragwürdig. Umgekehrt werden die beruflichen Tätigkeiten derjenigen, die nach eigener Angabe keine Arbeitskräfte beaufsichtigen, nur in Ausnahmefällen als Aufsichts- oder Führungskraft kodiert.

Daneben gibt es zahlreiche weitere Möglichkeiten, wie sich Aufsichtstätigkeiten erfassen lassen. Pollak et al. (2010) vergleichen einige Fragen und stellen in Abhängigkeit von der verwendeten Formulierung deutliche Unterschiede fest, wie viele Personen sich selbst als Aufsichtskraft bezeichnen. Eine allgemein akzeptierte Formulierung zur Erfassung von Aufsichtskräften konnte sich in der Wissenschaft bisher nicht durchsetzen.

Unsere Strategie zum Umgang mit Aufsichts- und Führungskräften ist wie folgt: Wenn eine Hilfskategorie eine Tätigkeit als Aufsichts- oder Führungskraft impliziert, beschreiben wir diesen Sachverhalt mit berufsspezifischen Formulierungen und heben dies als Kernbereich der Tätigkeit besonders hervor. Gegebenenfalls werden Folgefragen benötigt, um zwischen Aufsichtskräften und Führungskräften zu unterscheiden (vgl. das Beispiel „Lagerleiter/in“). Die Definitionen der ISCO-08 zu Aufsichtskräften, Führungskräften und Geschäftsführern werden dabei berücksichtigt und dienen als Ergänzung zu den Definitionen der Berufskategorien. Im Gegensatz zum offiziellen Umsteigeschlüssel der KldB 2010 und im Einklang mit der dort vorgelegten Definition betrachten wir „Führungskräfte“ als äquivalent zu „Managern“ in ISCO-08, d.h. wenn eine Hilfskategorie einer Führungskräfte-Kategorie in der KldB 2010 zugeordnet ist, muss auch immer eine Manager-Kategorie in ISCO-08 zugeordnet sein.

3.3.3 Berufe ohne Spezialisierung

Berufskategorien sind über Tätigkeitsbündel definiert, die für die jeweilige Berufskategorie charakteristisch sind. Einige Berufe weisen aber eine hohe Bandbreite verschiedener Aufgaben auf, die sich unterschiedlichen Tätigkeitsbündeln zuordnen lassen, weshalb eine eindeutige Zuordnung zu einer einzigen Berufskategorie nicht ohne weiteres möglich ist. Die KldB 2010 und die ISCO-08 schlagen unterschiedliche Strategien zum Umgang mit diesen Fällen vor.

Berufe ohne Spezialisierung in der KldB 2010. Ähnlich wie die Vorgängerklassifikationen seit 1970 sieht die KldB 2010 eigenständige Berufskategorien „ohne Spezialisierung“ vor. Berufe, die zwar auf übergeordneter Ebene in die Klassifikation eingeordnet werden können, aber auf der untersten berufsfachlichen Ebene keine Spezialisierung erkennen lassen, werden dieser Berufskategorie zugeordnet. Als Beispiel nennt die KldB 2010 den Beruf „Pferdewirt/in“. Da unbekannt ist, ob der Tätigkeitsschwerpunkt des Pferdewirts in der Aufzucht und Versorgung der Tiere („1131 Berufe in der Pferdewirtschaft – Pferdezucht“) oder in der Ausbildung von Pferden und Reitern („1132 Berufe in der Pferdewirtschaft – Reiten“) liegt, ist die Berufsbenennung „Pferdewirt/in“ der systematischen Einheit „1130 Berufe in der Pferdewirtschaft (ohne Spezialisierung)“ zugeordnet. Berufskategorien ohne Spezialisierung sind durch eine „0“ an vierter Stelle gekennzeichnet.

Eine nähere Betrachtung zeigt, dass Berufskategorien ohne Spezialisierung für unterschiedliche Zwecke verwendet werden:

a) In einem Beruf können unterschiedliche Aufgaben übernommen werden, die sich mehreren Berufskategorien zuordnen lassen. Im Beispiel wäre dies der Fall, wenn der Pferdewirt tatsächlich für die Aufzucht und Versorgung der Tiere und zugleich auch für die Ausbildung von Pferden und Reitern zuständig wäre.

b) Auch Berufskategorien, die dem Namen nach ohne Spezialisierung sind, können spezialisierte Berufe beschreiben. Beispielsweise nimmt die Berufskategorie „81404 Ärztinnen/Ärzte (ohne Spezialisierung)“ besonderen Bezug auf (spezialisierte) Fachärzte für Allgemeinmedizin. Im Gegensatz zum Pferdewirt wird hier

also nicht angenommen, dass Ärzte ohne Spezialisierung die Aufgaben verschiedener spezialisierter Berufskategorien (z.B. von Fachärzten für innere Medizin oder von Fachärzten für Chirurgie) wahrnehmen.

c) In vielen Fällen fehlen aber auch Informationen über die ausgeübte Tätigkeit. Es ist gut möglich, dass ein Befragter, der im Interview „Pferdewirt“ antwortet, tatsächlich nur für die Aufzucht und Versorgung der Tiere zuständig ist und daher aufgrund seiner beruflichen Tätigkeit der Berufskategorie 11312 zugeordnet werden sollte. Eine Zuordnung zur 11302, wie sie der Kodier-Index vornimmt, ist dann falsch. Fehlende Informationen vom Befragten sind aber eine Schwierigkeit bei der Erfassung des Berufs, die für die Entwicklung der Hilfsklassifikation nur insoweit eine Rolle spielt, als hier erneut die Problematik von vagen Berufsbenennungen zum Vorschein kommt. Da wir eine derartige Vagheit aber gerade vermeiden wollen, müssen unsere Hilfskategorien präziser beschrieben werden als das, was in allgemeinen Berufsbenennungen wie „Pferdewirt/in“ und „Arzt/Ärztin“ offenbart wird.

Für die Entwicklung der Hilfsklassifikation besteht die Herausforderung, dass Hilfskategorien disjunkt sein sollen. Im Prinzip gehen wir für Berufskategorien ohne Spezialisierung so vor wie sonst auch und übernehmen alle wichtigen Aspekte aus den Definitionen der zugrundeliegenden Berufskategorien. Dabei kann es von Nöten sein besonders hervorzuheben, dass Berufskategorien ohne Spezialisierungen ein vielfältiges Tätigkeitsspektrum enthalten. Im Beispiel zur 11302 ist in der zugehörigen Hilfskategorie als Tätigkeit „Pflege und Training von Pferden“ angegeben. Dies soll deutlich machen, dass eine Tätigkeit, die alleine der Pferdeaufzucht oder alleine der Ausbildung gewidmet wäre, hier nicht zugeordnet werden soll.

Inwieweit es auf diese Weise gelungen ist, dass Befragte die Unterschiede zwischen stärker spezialisierten Hilfskategorien und weniger spezialisierten Hilfskategorien verstehen, bleibt der weiteren Prüfung vorbehalten. Ggf. sollten die entsprechenden Hilfskategorien auch aus der Hilfsklassifikation entfernt werden, wenn es nicht gelungen ist, die Allgemeinheit der jeweiligen Hilfskategorie darzustellen.

Berufskategorien ohne Spezialisierung können, wie wir am Beispiel des Pferdewirts gesehen haben, die Tätigkeitsbündel von mehreren Berufskategorien umfassen. Auf diese Weise können Berufe, die auf unterster berufsfachlicher Ebene keine Spezialisierung erkennen lassen, der Berufsklassifikation zugeordnet werden. Wenn ein Beruf aber noch breiter aufgestellt ist, wenn z.B. der Pferdewirt auch noch Esel züchtet, stellt die soeben beschriebene Methode aus der KldB 2010 keine Kriterien bereit, nach der ein Pferde- und Eselzüchter der Berufsklassifikation zugeordnet werden kann.

Berufe ohne Spezialisierung in der ISCO-08. Das Prinzip von Berufskategorien ohne Spezialisierung („general occupations“) wurde in der internationalen Standardklassifikation von 1968 eingeführt und erst kurz darauf von der KldB übernommen. In den ISCO-Überarbeitungen von 1988 und 2008 wurde aber darauf verzichtet. Stattdessen gibt die ISCO-08 drei Regeln an, anhand derer Berufe klassifiziert werden können,

die eine hohe Bandbreite von verschiedenen Aufgaben wahrnehmen und daher mehreren Berufskategorien zugeordnet werden könnten.

a) Die beruflichen Tätigkeiten (tasks and duties), die das höchste skill level erfordern, haben Priorität gegenüber Tätigkeiten, die ein geringeres skill level erfordern.

b) Die beruflichen Tätigkeiten, die mit der Produktion von Gütern in Verbindung stehen, haben Priorität gegenüber Tätigkeiten, die mit der Verteilung und dem Verkauf derselben Güter in Verbindung stehen.

c) Die beruflichen Tätigkeiten, die im Beruf vorherrschen und besonders zeitaufwändig sind, haben Priorität gegenüber anderen Tätigkeiten.

Diese Regeln sollen ISCO-08 zufolge in der angegebenen Reihenfolge Verwendung finden. Es erscheint uns aber zweifelhaft, ob im Kontext der Kodierung von Antworten aus einer Umfrage die benötigten Informationen in Erfahrung gebracht werden können. Die Hilfsklassifikation unterstützt diese komplexen Entscheidungsregeln daher nicht und ggf. müssten Kodierer entsprechend geschult werden.

Für den Einsatz der Hilfsklassifikation im Interview wird es nötig sein eine Entscheidungsregel anzugeben, nach der Befragte eine Hilfskategorie auswählen können, wenn mehrere zutreffen. Da diese Entscheidungsregel einfach verständlich und allgemein anwendbar sein sollte, kommen die ersten beiden Entscheidungsregeln aus der ISCO-08 nicht in Betracht. Die dritte Entscheidungsregel, die eine Zuordnung nach vorherrschender Tätigkeit und zeitlichem Aufwand vornimmt, erscheint aber geeignet und dürfte sich weitestgehend mit einem intuitiven Verständnis decken. Auch entspricht diese dritte Regel weitestgehend dem Vorgehen aus der KldB 2010, wonach der „Tätigkeitsschwerpunkt“ für die Zuordnung von Berufsbenennungen ausschlaggebend war.

3.3.4 Residualkategorien für sonstige spezifische Berufe

In der ISCO-08 und in der KldB 2010 sind jeweils Residualkategorien für berufliche Tätigkeiten vorgesehen, die anderweitig nicht zugeordnet werden können. Derartige Berufe weisen eine klare Spezialisierung auf und eine grobe Klassifizierung auf übergeordneter Ebene ist daher möglich. Wenn auf unterster Ebene aber keine andere Berufskategorie zutrifft, werden derart spezialisierte Berufe in einer Residualkategorie zusammengefasst (4. Ziffer = „8“ in der KldB 2010 bzw. 4. Ziffer = „9“ in der ISCO-08). Zum Beispiel lassen sich die Berufe „Amtsarzt/-ärztin“, „Neurochirurg/in“, „Umweltmediziner/in“ auf übergeordneter Ebene der „814 Human- und Zahnmedizin“ zuordnen. Die Berufskategorien gliedern dies weiter in verschiedene fachärztliche Spezialisierungen auf. Da von diesen Spezialisierungen aber keine zutrifft, werden die genannten Berufe in einer Berufskategorie „81484 Ärztinnen/Ärzte (sonstige spezifische Tätigkeitsangabe)“ zusammengefasst.

Aufgrund dieser Konstruktionsweise können die Residualkategorien vielfältige Berufe aufweisen, die untereinander kaum Überschneidungsbereiche haben. In vielen Fällen würde es die Verständlichkeit der Hilfskategorie erschweren, wenn wir versuchen

würden, die unterschiedlichen Berufe in einer einzigen Hilfskategorie zusammenzufassen. Stattdessen legen wir in solchen Fällen mehrere Hilfskategorien für die verschiedenen Berufe an. Das BERUFENET der Bundesagentur für Arbeit bildet dabei oft die Grundlage, anhand der wir die einzelnen Hilfskategorien erstellen.

Im Einzelfall stellte sich die Bearbeitung von Residualkategorien aber immer wieder als schwierig und zeitaufwändig heraus. 16 Residualkategorien, die sich nicht klar von anderen Kategorien aus der KldB 2010 unterscheiden oder die nur sehr spezifische, seltene Berufsbenennungen enthalten, wurden daher bei der Bearbeitung ausgelassen (vgl. Anhang).

4 Diskussion

Seit nahezu 100 Jahren verfolgen Statistiker das Ziel, den Beruf im Sinne der ausgeübten Tätigkeit zu erfassen. Die Kategorien aus Berufsklassifikationen sind daher über ihre jeweils typischen Tätigkeiten definiert und im Interview werden Personen nach ihrer "berufliche[n] Tätigkeit" befragt. Zur Kodierung werden aber meist Kodier-Indizes mit teils unpräzisen Berufsbenennungen verwendet und nicht etwa die tätigkeitsbezogenen Definitionen der Berufskategorien. Berufsbenennungen sind aber bloß unpräzise Informationsbündel, die für eine detaillierte Beschreibung der ausgeübten Tätigkeit oft nicht ausreichen.

Als Mittelweg zwischen unpräzisen Berufsbenennungen in Kodier-Indizes und den umfassenden, aber zugleich unübersichtlichen Definitionen der Berufskategorien in offiziellen Klassifikationen entwickelten wir eine Hilfsklassifikation mit 1226 Hilfskategorien. Da wir Tätigkeiten aus den offiziellen Berufskategorien nur im üblicherweise benötigten Detailgrad beschreiben, ist das neue Instrument zum Einsatz in computergestützten Vorschlagsystemen zur Berufskodierung geeignet. Insbesondere eine Online-Nutzung während der Datenerhebung wird angestrebt, sodass Befragte die am besten passende Hilfskategorie selbst auswählen können. Indem wir die Ungenauigkeit von Berufsbenennungen überwinden, hoffen wir eine direkte Kodierung der beruflichen Tätigkeit ohne den Umweg über Berufsbenennungen zu ermöglichen und so die Qualität der Berufskodierung zu erhöhen.

Zur Erstellung der Hilfsklassifikation wurden die Definitionen der Berufskategorien aus der nationalen deutschen Berufsklassifikation KldB 2010 sowie aus der internationalen Berufsklassifikation ISCO-08 verwendet. Aus diesem Grund reicht es aus, Befragte nur ein einziges Mal anhand der Hilfsklassifikation zu kodieren. Wenn die passendste Kategorie der Hilfsklassifikation bekannt ist, lassen sich daraus direkt die zutreffenden Kategorien aus den offiziellen Berufsklassifikationen ableiten, was den Aufwand der Kodierung reduziert.

Oberste Priorität bei der Entwicklung war stets, dass bei Zutreffen einer Hilfskategorie auch die zugrundeliegenden Berufskategorien zutreffen müssen. In Grenzfällen ist es vorzuziehen, dass Befragte nicht anhand der Hilfsklassifikation, sondern anhand der Definitionen aus der KldB 2010 und der ISCO-08 klassifiziert werden. Aus diesem

Grund wurde keine Vollständigkeit der Hilfsklassifikation angestrebt in dem Sinne, dass sie für alle Beschäftigen eine passende Hilfskategorie enthält. Möglichst präzise Beschreibungen der einzelnen Hilfskategorien sollen sicherstellen, dass bei Nicht-Auffindbarkeit einer passenden Hilfskategorie auf keinen Fall eine Falschkodierung erfolgt. Wenn keine passende Hilfskategorie gefunden werden kann, ist eine Mehrfachkodierung anhand der offiziellen Klassifikationen mit konventioneller Methodik weiterhin erforderlich.

Zur Entwicklung haben wir hauptsächlich Definitionen der Berufskategorien aus der KldB 2010 und aus der ISCO-08 verwendet. Dies erforderte an vielen Stellen eigene Interpretationen, was uns an den jeweiligen Definitionen wichtig erschien und die sich von den Intentionen der Verfasser unterscheiden können. Experten, die sich mit einzelnen Berufen als auch mit der jeweiligen Klassifikation besonders gut auskennen, könnten unsere Hilfskategorien sicher noch besser auf die offiziellen Berufskategorien abstimmen.

Bisher haben wir die Hilfsklassifikation stets als unterstützendes Produkt zur Kodierung in die KldB 2010 und in die ISCO-08 beschrieben. Ausgangspunkt war hierbei die Annahme, dass die Berufskategorien der jeweiligen Klassifikation berufliche Tätigkeiten beschreiben und untereinander paarweise disjunkt sind. Unser Konstruktionsprinzip zur Erstellung der Hilfsklassifikation ist nur unter dieser Voraussetzung sinnvoll, denn nur dann beschreiben unsere Hilfskategorien paarweise disjunkte berufliche Tätigkeiten. Jedoch ist diese Annahme wohl nicht vollständig gültig, denn bei der Entwicklung der KldB 2010 ging es weniger um eine disjunkte Gliederung beruflicher Tätigkeiten, sondern vielmehr um die Systematisierung von Berufsbenennungen anhand der drei Gliederungskriterien Berufsfachlichkeit, Führungstätigkeit und Anforderungsniveau. Sofern diese Disjunktheits-Annahme nicht erfüllt war – und auch aufgrund der allgemeinen Interpretationsbedürfigkeit von Berufskategorien –, mussten bei der Entwicklung der Hilfsklassifikation an vielen Stellen Wertungen und Erwägungen vorgenommen werden, die sich nicht aus den zugrundeliegenden Klassifikationen ableiten lassen. Daher ist es auch fragwürdig, ob die Hilfsklassifikation bloß als ein unterstützendes Produkt zur Berufskodierung betrachtet werden sollte, wie dies ursprünglich geplant und in diesem Artikel dargelegt wurde.

Alternativ könnte man die Hilfsklassifikation auch als ein eigenständiges Schema betrachten, welches ohne engen Bezug zu den offiziellen Berufsklassifikationen alle wichtigen Tätigkeitsbündel auf dem deutschen Arbeitsmarkt enthält. Ein solch eigenständiges Schema wäre wertvoll, denn die KldB 2010 ist zur Systematisierung von Ausbildungs- und Berufsbenennungen konzipiert, aber eine Klassifizierung von Befragten anhand ihrer beruflichen Tätigkeit ist nicht originäres Ziel der Klassifikation. Nach dieser Sichtweise wäre es empfehlenswert, unsere Hilfsklassifikation einem Realitätscheck zu unterziehen. Es ist gut möglich, dass die Hilfskategorien nicht zutreffend beschreiben, was Beschäftigte selber als Kernbereiche ihrer beruflichen Tätigkeit sehen, oder dass Beschäftigte ihre Tätigkeiten in ganz anderen Kombinationen ausführen als von uns antizipiert. Unser Tätigkeitsschema sollte in diesem Fall der

Realität angeglichen und erweitert werden. Eine derartige Umstellung der Hilfsklassifikation zu einem Tätigkeitsschema steht aber im Widerspruch zu unserem Ziel, eine möglichst eindeutige Zuordnung zu den Berufskategorien der offiziellen Klassifikationen zu ermöglichen. Während die Nutzung der offiziellen Berufsklassifikationen bereits etabliert ist, müsste man entsprechende Verfahren für ein davon losgelöstes Tätigkeitsschema erst entwickeln.

Mit der Hilfsklassifikation steht ein neues Instrument zur Unterstützung der Berufskodierung bereit. Im nächsten Schritt soll die Hilfsklassifikation empirisch in Interviewsituationen getestet werden.

Literatur

Bundesagentur für Arbeit (2017): Tätigkeitsschlüssel-Online (http://bns-ts.arbeitsagentur.de/) (abgerufen am 18.09.2017).

Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010 (https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/KldB2010-Nav.html) (abgerufen am 18.09.2017).

Bushnell, Diane (1998): An Evaluation of Computer-assisted Occupation Coding in New Methods for Survey Research. In: Proceedings of the International Conference. Southampton: Association for Survey Computing, S. 23–36.

Cantor, David; Esposito, John (1992): Evaluating interviewer style for collecting industry and occupation information. In: Proceedings of the Survey Research Section. American Statistical Association, S. 661–666.

Conrad, Frederick; Couper, Mick; Sakshaug, Joseph (2016): Classifying open-ended reports: Factors affecting the reliability of occupation codes. In: Journal of Official Statistics, 32. Jg., Nr. 1, S. 75–92.

Damelang, Andreas; Schulz, Florian; Vicari, Basha (2015): Institutionelle Eigenschaften von Berufen und ihr Einfluss auf berufliche Mobilität in Deutschland. In: Schmollers Jahrbuch, 135. Jg., Nr. 3, S. 307–333.

Demszky von der Hagen, Alma;Sperling Voß, G. Günter (2010): Beruf und Profession. In: Böhle, Fritz; Voß, G. Günter; Wachtler, Günther (Hg.) (2010): Handbuch Arbeitssoziologie. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 751–803.

Dostal, Werner (2002): Der Berufsbegriff in der Berufsforschung des IAB. In: Kleinhenz, Gerhard (Hg): IAB-Kompendium Arbeitsmarkt- und Berufsforschung. Beiträge zur Arbeitsmarkt- und Berufsforschung, 250. Jg., S. 463–474.

Dostal, Werner; Schade, Hans-Joachim; Parmentier, Klaus (1999): Möglichkeiten und Grenzen der quantitativen Berufsforschung am IAB. Eine Bestandsaufnahme. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 1. Jg., Nr. 99, S. 41–60.

Dostal, Werner; Stooß, Friedemann; Troll, Lothar (1998): Beruf – Auflösungstendenzen und erneute Konsolidierung. In: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 31. Jg., Nr. 3, S. 438–460.

Elias, Peter; Birch, Margaret; Ellison, Ritva (2014): CASCOT International version 5 (http://www2.warwick.ac.uk/fac/soc/ier/software/cascot/internat/) (abgerufen am 18.09.2017).

Elias, Peter (1997): Occupational classification (ISCO-88). Concepts, methods, reliability, validity and cross-national comparability. (http://dx.doi.org/10.1787/304441717388) (abgerufen am 18.09.2017).

Embury, Brian (1997): Constructing a map of the world of work. How to develop the structure and contents of a national standard classification of occupations. Working Paper No. 95/2, Geneva: Bureau of Statistics, International Labour Office.

Fürst, Gerhard (1929): Zur Methode der deutschen Berufsstatistik. In: Allgemeines Statistisches Archiv, 19. Jg., S. 1–29.

Geis, Alfons (2011): Handbuch für die Berufsvercodung. Coding Documentation (https://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/handbuch_der_berufscodierung_110304.pdf) (abgerufen am 18.09.2017).

Geis, Alfons; Hoffmeyer-Zlotnik, Jürgen HP. (2000): Stand der Berufsvercodung. In: ZUMA-Nachrichten., 47. Jg., S. 103–128.

Gilbert, Daniel T.; Krull, Douglas S.; Malone, Patrick S. (1990): Unbelieving the Unbelivable. Some Problems in the Rejection of False Information. In: Journal of Personality and Social Psychology, 59. Jg., Nr. 4, S. 601–613.

Hoffmann, Eivind (1994): Mapping a national classification of occupations into ISCO-88. outline of a strategy. In: Chernyshev, I. (Hg.) (1994): Labour Statistics for a Market Economy. Challenges and Solutions in the Transition Countries of Central and Eastern Europe and the Former Soviet Union, Budapest: Central European University Press, S. 203–209.

International Labour Office (2012): International Standard Classification of Occupations: ISCO-08. Geneva: International Labour Organization.

International Labour Office (1958): International Standard Classification of Occupations. Geneva.

International Labour Office (1923): Systems of Classification of Industries and Occupations. Studies and Reports, Series N, No. 1. Geneva.

Loos, Christiane; Eisenmenger, Matthias; Bretschi, David (2013): Das Verfahren der Berufskodierung im Zensus 2011. In: Wirtschaft und Statistik, 3. Jg., S. 173–184.

Meerwarth, Rudolf (1925): Nationalökonomie und Statistik. Eine Einführung in die empirische Nationalökonomie. In: Handbuch der Wirtschafts- und Sozialwissenschaften in Einzelbänden, Bd. 7., Berlin: Walter de Gruyter.

Meier, Urs (2003): Handbuch zur Berufsdatenbank. Bundesamt für Statistik (https://www.bfs.admin.ch/bfsstatic/dam/assets/337799/master) (abgerufen am 29.05.2017).

Müller, Anne (2014): The implementation of the German Classification of Occupations 2010 in the IAB Job Vacancy Survey. In: IAB-Forschungsbericht 10/2014. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.

Paulus, Wiebke; Matthes, Britta (2013): Klassifikation der Berufe. Struktur, Codierung und Umsteigeschlüssel. In: FDZ-Methodenreport 08/2013. Nürnberg: Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung.

Pollak, Reinhard; Bauer, Gerrit; Müller, Walter; Weiss, Felix; Wirth, Heike (2010): The comparative measurement of supervisory status. In: Rose, David; Harrison, Eric (2014): Social Class in Europe. An introduction to the European Socio-economic Classification. Routledge.

Prigge, Michaela; Köhr, Martha; Pfeiffer, Norbert; Blettner, Maria; Beutel, Manfred; Wild, Philipp; Münzel, Thomas; Blankenberg, Stefan; Seidler, Andreas; Letzel, Stephan; Latza, Ute; Liebers, Falk (2014): Codierung der Tätigkeitsangaben im Basiskollektiv der Gutenberg-Gesundheitsstudie unter Anwendung der Klassifikation der Berufe KldB 2010. Darstellung des Vorgehens und der Datenqualität. Zeitschrift für Arbeitswissenschaft, 68. Jg., Nr. 3, S. 153–162.

Rauchberg, Heinrich (1888): Die deutsche Berufs- und Betriebszählung vom 5. Juni 1882. In: Statistische Monatsschrift 14, S. 569–603.

Schierholz, Malte; Gensicke, Miriam; Tschersich, Nikolai; Kreuter, Frauke (2018): Occupation coding during the interview. In: Journal of the Royal Statistical Society, Series A, 181. Jg., Nr. 2, S. 379–407.

Schönbach, Klaus (1979): Probleme der Verschlüsselung von Berufstätigkeiten. In: Pappi, Franz Urban (Hg.) (1979): Sozialstrukturanalysen mit Umfragedaten. Königstein im Taunus: Athenäum-Verlag, S. 41–57.

Smith, Adam (1998, orig. 1776) Wealth of Nations – A Selected Edition. Edited by Kathryn Sutherland, New York: Oxford University Press.

Sperling, Hans (1961): Zur Theorie und Methode der Berufsklassifizierung. In: Schmollers Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft. Berlin: Duncker & Humblot, 81. Jg., S. 705–720

Statistisches Bundesamt (2016): Demographische Standards. Ausgabe 2016. Band 17 der Reihe Statistik und Wissenschaft. Wiesbaden.

Statistisches Bundesamt (1961): Klassifizierung der Berufe. Stuttgart: Kohlhammer.

Stooß, Friedemann; Saterdag, Hermann (1979): Systematik der Beruf und der beruflichen Tätigkeiten. In: Pappi, Franz Urban (Hg.) (1979): Sozialstrukturanalysen mit Umfragedaten. Königstein im Taunus: Athenäum-Verlag, S. 41–57.

Stooß, Friedemann (1977): Die Systematik der Berufe und der beruflichen Tätigkeiten. In: Seifert, Karl Heinz (Hg.) (1977): Handbuch der Berufspsychologie. Göttingen: Verlag für Psychologie Hogrefe, S. 69–98.

Tijdens, Kea (2014): Dropout rates and response times of an occupation search tree in a web survey. Journal of Official Statistics, 30. Jg., Nr.1, S. 23–43.

Tijdens, Kea (2010) Measuring occupations in web-surveys. The WISCO database of occupations. In: Amsterdam Institute for Advanced labour Studies. Working Paper No. 10/86. Amsterdam: University of Amsterdam.

Tourangeau, Roger; Rips, Lance R.; Rasinski, Kenneth (2000): The Psychology of Survey Response. Cambridge: Cambridge University Press.

Watson, Tony James (2012): Sociology, work and organization. 6th edition, London: Routledge.

Weeden, Kim A.; Grusky, David B. (2005): The Case for a New Class Map. In: American Journal of Sociology, 111. Jg., Nr. 1, S. 141–212.

Weber, Max (1980, orig. 1921) Wirtschaft und Gesellschaft. Grundriß der verstehenden Soziologie. Studienausgabe, 5., rev. Aufl. besorgt von Johannes Winckelmann, Tübingen: Mohr.

Willms, Angelika (1983): Historische Berufsforschung mit amtlicher Statistik. Rekonstruktion der Entwicklung der Berufsstatistik in Deutschland und Entwurf einer Klassifikation vergleichbarer Berufsfelder, 1925–1980. Mannheim: VASMA-Projekt, Arbeitspapier Nr. 30.

Impressum

IAB-Discussion Paper 13/2018
14. Mai 2018

Herausgeber
Institut für Arbeitsmarkt- und Berufsforschung der Bundesagentur für Arbeit
Regensburger Straße 104
90478 Nürnberg

Redaktion
Ricardo Martinez Moya, Jutta Palm-Nowak

Technische Herstellung
Renate Martin

Rechte
Nachdruck - auch auszugsweise - nur mit Genehmigung des IAB gestattet

Website
http://www.iab.de

Bezugsmöglichkeit
http://doku.iab.de/discussionpapers/2018/dp1318.pdf

ISSN 2195-2663

Rückfragen zum Inhalt an:

Malte Schierholz
Telefon 0911.179 6022
E-Mail Malte.Schierholz@iab.de

Anhang zu: *Vorstellung einer Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung*

Malte Schierholz[1,2], Lorraine Brenner[1], Lea Cohausz[1], Lisa Damminger[1], Lisa Fast[1], Ann-Kathrin Hörig[1], Anna-Lena Huber[1], Annabell Petry[1], Laura Tschischka[1], Theresa Ludwig[1]

[1] Mannheimer Zentrum für Europäische Sozialforschung, Universität Mannheim, Deutschland

[2] Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg, Deutschland

Inhalt

Formatierungsvorgaben 2

Organisatorisches Vorgehen 9

Ausgewählte Schwierigkeiten 10

Beispiel: Verwendetes Material zur Bearbeitung der KldB-Berufskategorie 92122... 14

Formatierungsvorgaben

Die neu geschaffene Hilfsklassifikation wird in der Datei „hilfsklassifikation.xml“ bereitgestellt. Die darin enthaltenen Angaben und das zugrundeliegende Schema werden im Folgenden anhand von Abbildung 1 für eine einzelne Kategorie erläutert.

```
- <kategorie>
    <id>4-stellig</id>
    <taetigkeit>string</taetigkeit>
    <taetigkeitsbeschreibung>string</taetigkeitsbeschreibung>
    <bezeichnung>string</bezeichnung>
  - <untergliederung>
    - <default>
        <kldb schluessel="5-stellig">string</kldb>
        <isco schluessel="4-stellig">string</isco>
        <dkz codenr="char(9)" refid="integer">string</dkz>
        ...
        <dkz codenr="char(9)" refid="integer">string</dkz>
      </default>
      <fragetext>string</fragetext>
    - <antwort position="numeric">
        <text>string</text>
        <kldb schluessel="5-stellig">string</kldb>
        <isco schluessel="4-stellig">string</isco>
        <dkz codenr="char(9)" refid="integer">string</dkz>
        ...
        <dkz codenr="char(9)" refid="integer">string</dkz>
      </antwort>
      ...
    - <antwort position="numeric">
        <text>string</text>
        <kldb schluessel="5-stellig">string</kldb>
        <isco schluessel="4-stellig">string</isco>
        <dkz codenr="char(9)" refid="integer">string</dkz>
        ...
        <dkz codenr="char(9)" refid="integer">string</dkz>
      </antwort>
    </untergliederung>
    <abgrenzung refid="4-stellig" typ="hoch">string</abgrenzung>
    ...
    <abgrenzung refid="4-stellig" typ="hoch">string</abgrenzung>
    <abgrenzung refid="4-stellig" typ="mittel">string</abgrenzung>
    ...
    <abgrenzung refid="4-stellig" typ="mittel">string</abgrenzung>
  </kategorie>
```

Abbildung 1: Darstellung einer Kategorie im xml-Format

Jede Kategorie beginnt mit dem Start-tag <kategorie> und endet mit dem End-tag </kategorie>. Dazwischen wird die Kategorie mithilfe der folgenden Elemente definiert, die zwingend vorkommen müssen:

id

Beispiel:

<id>1002</id>

Funktion: Eindeutiger Schlüssel für die Kategorie

Anforderungen: Einmalig verwendete 4-stellige Ziffer

taetigkeit

Beispiel:

<taetigkeit>Bearbeitung von Anfragen, Aufträgen oder Reklamationen (Inbound), z.B. per Telefon oder E-Mail</taetigkeit>

Funktion: Beschreibung der charakteristischen Kerntätigkeit dieser Kategorie.

Anforderungen: Die Tätigkeit muss eindeutig abgrenzbar zu anderen Tätigkeiten sein und zu diesem Zweck *hinreichend präzise*. Unterschiede zu verwandten Kategorien sollten klar erkennbar sein. Zugleich sollte sie möglichst *kurz* und *leicht verständlich* sein, da dies vom Interviewer vorgelesen werden muss. Die Befragten sollten diese <taetigkeit> nur dann auswählen, wenn die zugeordneten Kategorien aus der KldB und aus der ISCO auch tatsächlich zutreffen.

Stil: Im Regelfall verwenden wir für die <taetigkeit> substantivierte Verben anstelle der Infinitivformen (z.B. „Fremdsprachenunterricht in außerschulischen Bildungseinrichtungen“ anstelle von „in außerschulischen Bildungseinrichtungen eine Fremdsprache unterrichten“). Ausnahmen sind e rlaubt, wenn dies die Verständlichkeit erhöht. Falls Berufsbenennungen enthalten sind, wird zum besseren Lesefluss nur die männliche Form verwendet.

Häufige Fehler:

- Verwendung von überflüssigen Wörtern, obwohl sie keine relevanten Informationen enthalten. Nur das Wichtigste sollte in der der <taetigkeit> stehen, weniger Wichtiges ggf. in der <taetigkeitsbeschreibung>.
- Wenn eine Berufsbezeichnung vorhanden ist, die die entsprechende Kategorie in der KldB präzise beschreibt (z.B. „Kutscher/in“, „Ergotherapeut/in“), sollte der gleiche Wortstamm auch bei der <taetigkeit> verwendet werden. (z.B. „Führen von Pferdekutschen zu Transportzwecken“, „ergotherapeutische Behandlung von Patienten“) anstelle von Umschreibungen (z.B. „Führen von von Pferden gezogenen Fahrzeugen zu Transportzwecken“)

taetigkeitsbeschreibung

Beispiel:

<taetigkeitsbeschreibung>z.B. Kundenanfragen beantworten und Kunden über verschiedene Produkte beraten; Aufträge in Computersystemen erfassen; Informationen an Kunden versenden</taetigkeitsbeschreibung>

Funktion: Beispielhafte Aufgaben dieser Hilfskategorie oder nähere Erläuterungen. Im Gegensatz zur <taetigkeit> wird die <taetigkeitsbeschreibung> im Interview nicht vorgelesen, sondern soll bei Nachfragen an den Interviewer zur Verfügung stehen.

Stil: Jedes Beispiel sollte mit einem Infinitiv enden. Einzelne Beispiele sollen nicht mit Komma, sondern mithilfe eines Semikolons deutlich voneinander getrennt werden (z.B. „z.B. Roherze und Hilfsstoffe aufbereiten; Hoch- oder Schmelzöfen einrichten, bedienen und kontrollieren; Wartungs- und

Reparaturarbeiten durchführen“). In vielen Fällen ist es vorteilhaft die Beispiele thematisch oder dem Arbeitsablauf entsprechend zu ordnen. In anderen Fällen sollte das Wichtigste vorne stehen. Falls Berufsbenennungen enthalten sind, wird zum besseren Lesefluss nur die männliche Form verwendet.

Herkunft: Üblicherweise basiert die <taetigkeitsbeschreibung> auf den üblichen Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, die in den Definitionen der Berufskategorien aus der KldB 2010 genannt sind.

bezeichnung

Beispiel:

<bezeichnung>Callcenteragent/in (Inbound)</bezeichnung>

Funktion: Eine möglichst kurze und prägnante Berufsbezeichnung. Mit einem Blick soll ein Interviewer anhand dieser Bezeichnung erkennen, ob diese Kategorie möglicherweise für den Befragten zutreffen könnte und er die die <taetigkeit> daher vorlesen sollte.

Stil: Es wird jeweils die männliche und die weibliche Form der Berufsbenennung benötigt, sodass der Bezug zu beiden Geschlechtern deutlich wird. Die Beschränkung auf ein Geschlecht ist nicht nötig, da die <bezeichnung> nicht im Fließtext verwendet werden soll und auf einen guten Lesefluss daher weniger Wert gelegt wird.

Herkunft: Berufsbenennungen aus der KldB 2010, aus der DKZ oder Freitextangaben aus dem Interview können als Vorbild dienen.

Hintergrundinfos: Üblicherweise wurde bei der Entwicklung der Hilfsklassifikation zuerst eine <taetigkeit> beschrieben und nachfolgend eine dazu passende <bezeichnung> ausgewählt.

Verschiedene Alternativen für die passendste <bezeichnung> müssen im Einzelfall gegeneinander abgewogen werden. Folgende einander entgegengesetzte Überlegungen sind dabei von Bedeutung:

- Vor dem Hintergrund des gesellschaftlichen und beruflichen Wandels existieren zu einigen Berufen ältere und neuere Benennungen. Auch die reglementierten Ausbildungsberufe erfahren immer wieder Namensänderungen. Da wir die aktuelle Berufslandschaft abbilden wollen, ist die Verwendung neuerer Bezeichnungen erstrebenswert. Andererseits sollten die Bezeichnungen möglichst plakativ, allgemein verständlich und im besten Fall ein Teil der Umgangssprache sein, was für ältere Bezeichnungen sprechen kann.
- Berufsbenennungen sollten präzise und identifikationsstiftend sein. Wenn Befragte sich mit ihnen identifizieren können, erleichtert dies ihre Auswahl. Andererseits könnten Personen, die sich nicht einem präzise benannten Beruf zugehörig fühlen, sondern bloß ähnliche Tätigkeiten ausüben, davon abgeschreckt werden eine präzise Benennung auszuwählen, obwohl die zugehörige Kategorie für sie die Richtige wäre. Um das Risiko einer derartigen falschen Nicht-Auswahl zu verringern, können auch weniger präzise Berufsbenennungen sinnvoll sein.

untergliederung und default

Die Untergliederung ermöglicht es, mithilfe einer Folgefrage die zugrundeliegende Kategorie näher einzugrenzen. Sofern keine Folgefrage erforderlich ist – der Regelfall –, muss nur der Kindknoten <default> enthalten sein, der angibt mit welchen Kategorien aus der DKZ, aus der KldB 2010 und aus der ISCO-08 die neue Kategorie standardmäßig in Verbindung steht.

kldb

Beispiel:

<kldb schluessel = "92122">Berufe im Dialogmarketing - fachlich ausgerichtete Tätigkeiten</kldb>

Funktion: Wenn diese Hilfskategorie ausgewählt wird, erfolgt zugleich eine Kodierung in die genannte Berufskategorie.

Anforderungen: Jeweils im <default>-Teil und für jede einzelne Antwortoption muss genau eine Berufskategorie aus der KldB 2010 angegeben werden. Ausnahmen sind bloß möglich, wenn zwei Folgefragen gestellt werden (vgl. Abschnitt „Ausgewählte Schwierigkeiten“).

Stil: Das Schlüsselattribut gibt die 5-stellige Nummer dieser Kategorie an, der Inhalt des Tags enthält den Namen der Kategorie.

isco

Beispiel:

<isco schluessel = "4222">Contact centre information clerks</isco>

Funktion: Wenn diese Hilfskategorie ausgewählt wird, erfolgt zugleich eine Kodierung in die genannte Berufskategorie.

Anforderungen: Jeweils im <default>-Teil und für jede einzelne Antwortoption muss genau eine Berufskategorie aus der ISCO-08 angegeben werden. Ausnahmen sind bloß möglich, wenn zwei Folgefragen gestellt werden (vgl. Abschnitt „Ausgewählte Schwierigkeiten“).

Stil: Das Schlüsselattribut gibt die 4-stellige Nummer dieser Kategorie an, der Inhalt des Tags enthält den Namen der Kategorie.

dkz

Beispiel:

<dkz codenr="92122-104" refid="35308">Servicefachkraft - Dialogmarketing</dkz>
<dkz codenr="92122-101" refid="7001">Callcenteragent/in</dkz>
<dkz codenr="92122-102" refid="14107">Fachkaufmann/-frau - Teleservice</dkz>
<dkz codenr="92122-103" refid="14994">E-Mail-Agent/in</dkz>

Funktion: Die angegebenen DKZs waren uns bei der Erstellung der Hilfsklassifikation bekannt. Sie sind hier zum Zwecke der Dokumentation des Entwicklungsprozesses angegeben.

Herkunft: Die *Dokumentationskennziffer* (DKZ) ist eine interne Berufsdatenbank der Bundesagentur für Arbeit.[1] Sie findet Verwendung im Rahmen der Beratung und Vermittlung von Arbeitskräften, im BERUFENET sowie zu statistischen Zwecken. Um die aktuelle berufliche Landschaft in Deutschland abzubilden, wird die DKZ laufend aktualisiert.

Jeder in der DKZ erfasste Beruf hat eine Bezeichnung (z.B. „Servicefachkraft – Dialogmarketing“). Der Bezeichnung sind jeweils eine ID (z.B. 35308) und eine Berufskennziffer/Codenummer (z.B. 92122-104) zugeordnet.

Die ersten 5 Ziffern der Codenummer entsprechen dem 5-stelligen Schlüssel aus der KldB 2010. Bei der Erstellung der Hilfsklassifikation wurden die DKZs jeweils den am besten passendsten Hilfskategorien zugeordnet (üblicherweise einer einzigen). Im Regelfall stimmen die KldB-Schlüssel der Hilfskategorien und die ersten fünf Ziffern der DKZ daher überein.

abgrenzung

Beispiel:

[1] Die DKZ-Datenbank ist im Downloadportal der Bundesagentur für Arbeit unter https://download-portal.arbeitsagentur.de/files/ nach Registrierung verfügbar.

<abgrenzung typ = "hoch" refid = "1003">Verkäufer/in - Telemarketing</abgrenzung>
<abgrenzung typ = "mittel" refid = "3504">Teamleiter/in - Callcenter</abgrenzung>
<abgrenzung typ = "mittel" refid = "3218">Schalterauskunft</abgrenzung>
<abgrenzung typ = "hoch" refid = "3220">Telefonist/in</abgrenzung>

Funktion: Mit dem <abgrenzung>-Tag wird angegeben, welche anderen Kategorien ebenfalls vorgelesen werden sollten, um dem Befragten eine bessere Auswahl zu ermöglichen. Wenn diese nicht vorgelesen werden, kann es passieren, dass der Befragte der vorgelesenen Kategorie zustimmt, obwohl eine andere passender wäre. Um derartige Falschkodierungen zu vermeiden, sollten auch ähnliche Kategorien vorgelesen werden, bei denen Verwechslungsgefahr besteht.

Anforderungen: Es können beliebig viele Abgrenzungen angegeben werden. Über das *typ*-Attribut sind zwei Arten von Abgrenzungen möglich:

- typ = „hoch“: Die angegebenen Kategorien sind zwingend vorzulesen, da ansonsten sehr viele Fehlklassifikationen zu befürchten sind.
- typ = „mittel“: die angegebenen Kategorien sind nach Möglichkeit vorzulesen, wenn die Anzahl der vorgeschlagenen Kategorien im Rahmen bleibt. Die Fehlklassifikationswahrscheinlichkeit ist gering.

Weitere Hinweise: Abgrenzungen spiegeln die subjektive Einschätzung der Ähnlichkeit von zwei Hilfskategorien wider. Eine Abgrenzung ist fast immer erforderlich zwischen Hilfskategorien, die aus einer einzigen Kategorie der KldB 2010 entstanden ist. Üblicherweise verwenden wir keine Abgrenzungen, wenn sich das Anforderungsniveau von zwei Berufskategorien aus der KldB 2010 um mindestens 2 unterscheidet (z.B. keine Abgrenzung zwischen 24511 und 24513)

fragetext und antwort

Beispiel (für die „Kaufmännische/r Betriebsleiter/in“ bezeichnete Kategorie (ID: 3211)):

<fragetext>Sind Sie befugt strategische Entscheidungen zu treffen, z.B. zur Einführung neuer Verfahren, zu finanziellen Investitionen, oder zur Einstellung und Entlassung von Personal?</fragetext>
<antwort position = "1">
  <text>Ja</text>
  <kldb schluessel = "71394">Führungskräfte - Unternehmensorganisation und -strategie</kldb>
  <isco schluessel = "1213">Policy and planning managers</isco>
</antwort>
<antwort position = "2">
  <text>Nein</text>
  <kldb schluessel = "71393">Aufsichtskräfte - Unternehmensorganisation und -strategie</kldb>
  <isco schluessel = "3341">Office supervisors</isco>
</antwort>

Funktion: Einige Berufskategorien unterscheiden sich bloß in Details. Zur besseren Verständlichkeit ist es daher manchmal hilfreich, zuerst nach einer gemeinsamen <taetigkeit> zu fragen und die Details erst danach in einer Folgefrage zu erfragen. Je nachdem welche Antwort bei der Folgefrage ausgewählt wird, erfolgt die Kodierung in die genannten Berufskategorien. Wenn keine Antwort vorliegt, erfolgt die Kodierung anhand der Berufskategorien im <default>-tag.

Anforderungen: Im Regelfall wird keine Folgefrage gestellt (siehe die Erläuterungen bei <untergliederung>). Nur wenn eine Folgefrage gestellt wird, sind die Elemente <fragetext> (Anzahl: üblicherweise exakt eine, allerdings sind Ausnahmen möglich, vgl. Abschnitt Ausgewählte

Schwierigkeiten“) und <antwort> (Mindestanzahl: 2) notwendig. Über das *position*-Attribut ist festgelegt, in welcher Reihenfolge die Antwortoptionen angezeigt werden.

Ein oder mehrere <dkz>-Tags sind als Kindelemente des <antwort>-tags ebenfalls erlaubt: Dies deutet darauf hin, dass DKZ-Berufe der jeweiligen Antwortoption besonders nahe stehen.

Details: Folgefragen können in ihrer Formulierung flexibel gestaltet werden und so die jeweils wesentlichen Unterscheidungsmerkmale von Berufskategorien hervorheben. Häufig betreffen die Folgefragen aber die fachliche Spezialisierung, das Anforderungsniveau oder Aufsichts- und Führungstätigkeiten. In diesen Fällen wurden die Formulierungen der Folgefragen wie folgt vereinheitlicht. Bei genauer Kenntnis einzelner Berufe könnten Experten sicherlich noch bessere Folgefragen für die entsprechenden Berufe formulieren.

Die fachliche Spezialisierung kann auf unterschiedliche Weise beschrieben werden. Daher wurden drei Standardformulierungen gewählt, die gut zu den Antwortoptionen passen. Im Einzelfall schien es häufig auch sinnvoll, anstelle der generischen Fragen berufsspezifische Formulierungen zu verwenden.

Im Abschnitt 2.2 wurde beschrieben, dass sich die Hilfsklassifikation vorrangig an den Beschreibungen der einzelnen Berufskategorien orientiert. Nur wenn die Beschreibungen sich inhaltlich kaum unterscheiden, wurde das Anforderungsniveau als nachrangiges Kriterium verwendet um die Unterschiede herauszuarbeiten. Dabei orientieren sich unsere Standardformulierungen stark an der Definition des Anforderungsniveaus aus der KldB 2010. Bei vielen Berufskategorien ist das Anforderungsniveau durch die Laufbahngruppe von Beamten bzw. durch die üblicherweise erforderliche Ausbildung bestimmt. Dabei ist zu beachten, dass nicht die tatsächlich abgeschlossene Ausbildung eines Berufstätigen zählt, sondern der erforderliche Bildungsabschluss bloß ein Proxy für die Komplexität der Tätigkeit ist und daher als rein arbeitsplatzbezogenes Merkmal zu verstehen ist. Dies ist in unseren Standardformulierungen berücksichtigt, wenn nach der Art von Ausbildung gefragt wird, die für die jeweilige Tätigkeit *in der Regel* erforderlich ist. Sofern in den Antwortoptionen Fachrichtungen und Namen einzelner Ausbildungen genannt sind, beruht dies auf unserer Recherche im BERUFENET und dient der Veranschaulichung. Im Einzelfall kam es auch vor, dass unterschiedliche Anforderungsniveaus von Berufskategorien nicht auf unterschiedliche Ausbildungen zurückgeführt werden konnten. In diesem Fall erlaubt die Frage nach der Notwendigkeit neues Wissen zu lernen eine Einordnung in Anforderungsniveau 1 bzw. 2.

Die KldB 2010 übernimmt in ihrer Definition von Aufsichts- und Führungstätigkeiten die Definitionen aus ISCO-08. Aufsichtskräfte in der KldB 2010 sind also äquivalent zu supervisors aus ISCO-08 und Führungskräfte sind äquivalent zu managers. Entsprechend waren auch die Definitionen aus der ISCO-08 für unsere Standardformulierungen maßgeblich. ISCO-08 zufolge beaufsichtigen Aufsichtskräfte die Aktivitäten anderer Arbeiter. Demgegenüber haben nur Führungskräfte die Befugnis, Entscheidungen zur Einführung neuer Verfahren, zu finanziellen Investitionen oder zur Einstellung und Entlassung von Personal zu treffen. Beides spiegelt sich in unseren Formulierungen wider.

Bei der Entwicklung unserer Standardfragen wurden Formulierungen aus dem Panel Arbeitsmarkt und soziale Sicherung (PASS) berücksichtigt und teilweise wortgleich übernommen.[2]

[2] Trappmann, M., Beste, J., Bethmann, A. and Müller, G. (2013). The pass panel survey after six waves, *Journal for Labour Market Research* **46**(4): 275–281.

Standardfragen zur fachlichen Spezialisierung

Mit welchem spezifischen Fachgebiet beschäftigen Sie sich?

In welchem Bereich sind Sie vorwiegend tätig?

Welche Aufgaben führen Sie dabei in der Regel aus?

Standardfragen zum Anforderungsniveau

Sind Sie Beamte/r im einfachen, mittleren, gehobenen oder höheren Dienst? oder *Sind Sie Beamte/r im mittleren, gehobenen oder höheren Dienst?*

1 im einfachen Dienst oder vergleichbar (nur bei der ersten Frage)

2 im mittleren Dienst oder vergleichbar

3 im gehobenen Dienst oder vergleichbar

4 im höheren Dienst oder vergleichbar

Welche Art von Ausbildung ist für Ihre Tätigkeit in der Regel erforderlich?

Beispielantworten:

1 eine abgeschlossene Berufsausbildung

2 eine vertiefende berufliche Weiterbildung mit Mindestdauer von 18 Monaten

3 ein abgeschlossenes Bachelorstudium (der Betriebswirtschaftslehre, Wirtschaftswissenschaften oder eines vergleichbaren Fachs)

4 ein abgeschlossenes Masterstudium

Ist für Ihre Tätigkeit in der Regel XY erforderlich?

Beispiel für XY:

- eine abgeschlossene Berufsausbildung (als Z)

- eine berufliche Weiterbildung (als Z)

- eine Prüfung zum (z.B.) Fachwirt

- ein abgeschlossenes Bachelorstudium (im Bereich Z)

- ein abgeschlossenes Masterstudium (im Bereich Z)

Mussten Sie für diese Tätigkeit neues Wissen lernen?

Zur Unterscheidung zwischen 1er und höheren Niveaus

Zur Unterscheidung von 1 und 2 wurde jedoch auch oft diese Frage verwendet: Ist für ihre Tätigkeit in der Regel eine abgeschlossene Berufsausbildung (als Z) erforderlich?

Standardfragen zu Aufsichts- und Führungstätigkeiten

Gehört es zu Ihren beruflichen Aufgaben, die Arbeit anderer Arbeitnehmer zu beaufsichtigen oder ihnen zu sagen, was sie tun müssen?

zur Unterscheidung zwischen 93 und untergeordneten KldBs (2er, 3er oder 4er)

Sind Sie befugt strategische Entscheidungen zu treffen, z.B. zur Einführung neuer Verfahren, zu finanziellen Investitionen, oder zur Einstellung und Entlassung von Personal?

zur Unterscheidung zwischen 93er und 94er

Ist ein wesentlicher Bestandteil ihrer Arbeit, andere Arbeitnehmer zu beaufsichtigen oder als Führungskraft strategische Entscheidungen für das Unternehmen zu treffen?

zur Unterscheidung von 93er, 94er und 4er bzw. 3er oder 2er

Organisatorisches Vorgehen

Folgende Dokumente wurden zu Rate gezogen, um ein umfassendes Bild der einzelnen Berufskategorien zu erhalten:

- Die offiziellen Kategoriebeschreibungen der KldB 2010 (Band 2) elektronische Fassung („Gliederung mit Erläuterungen" von https://www.klassifikationsserver.de, abgerufen am 28.03.2016)
- Das alphabetische Verzeichnis der KldB 2010 (Band 1), elektronische Fassung, („Stichwörter" von https://www.klassifikationsserver.de, abgerufen am 28.03.2016)
- Falls diese Kategoriebeschreibungen nicht präzise genug sind, greifen wir auf umfassende Berufsbeschreibungen zu einzelnen Berufen zurück, die im BERUFENET der Bundesagentur für Arbeit laufend aktualisiert werden. Insbesondere finden sich dort Tätigkeitsinhalte (z.B. unter https://berufenet.arbeitsagentur.de/berufenet/faces/index?path=null/kurzbeschreibung/taetigkeitsinhalte&such=Internationale%2Fr+Luftverkehrsassistent%2Fin&dkz=76530).
- Tabellarischer Umsteigeschlüssel von der KldB 2010 zur ISCO-08 (https://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/Arbeitshilfen/Umsteigeschluessel/Umsteigeschluessel-Nav.html, Erstellungsdatum: 15.09.2011)
- DKZ mit ISCO-Umsteigeschlüsseln (Internes Dokument des Instituts für Arbeitsmarkt- und Berufsforschung, Datenstand: 28.10.2015). Nur zum damaligen Zeitpunkt gültige Berufe (ohne Berufsausbildungen) wurden verwendet.
- ISCO-08 Struktur & Erläuterungen[3] (deutsche Fassung: http://www.statistik.at/kdb/downloads/csv/ISCO08_DE_COT_20151120_150453.txt). Wenn es auf einzelne Formulierungen aus der ISCO-08 ankommt, ist die englische Originalfassung maßgeblich, die in der Datenbank von Statistik Austria am besten zugänglich ist (http://www.statistik.at/kdb/downloads/csv/ISCO08_EN_COT_20151120_150801.txt, abgerufen am 11.05.2016).
- Berufsangaben von Befragten aus der IAB-Befragung *Arbeiten und Leben im Wandel* (ALWA)[4]

[3] Statistik Austria (2011): ISCO 08 – gemeinsame deutschsprachige Titel und Erläuterungen auf Basis der englischsprachigen Version 1.5a von April 2011. Wien. URL: http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=049974 (Publikation), http://www.statistik.at/KDBWeb/kdb.do?FAM=BERUF&&NAV=DE&&KDBtoken=null (Datenbankzugriff)

[4] Drasch, Katrin; Matthes, Britta; Munz, Manuel; Paulus, Wiebke; Valentin, Margot-Anna (2012): Arbeiten und Lernen im Wandel * Teil V: Die Codierung der offenen Angaben zur beruflichen Tätigkeit,

Basierend auf diesen Daten wurde automatisch für jede einzelne Kategorie (5-Steller) der KldB 2010 ein Überblick über diese Berufsgattung und ähnliche Berufe erstellt. Im Beispiel am Ende dieses Anhangs ist dieser Überblick exemplarisch für die Kategorie 92122 dargestellt. Dieser Überblick enthält:

- Eine umfassende Darstellung der jeweiligen zugrundeliegenden Kategorie. Dies enthält die Kategoriebeschreibung aus der KldB 2010, die zugeordneten Berufsbezeichnungen aus der DKZ, aus dem alphabetischen Verzeichnis der KldB 2010 und Freitext-Antworten aus der in diese Kategorie kodierte Antworten aus der ALWA-Befragung.
- Eine ähnliche Darstellung zum Überblick über alle verwandten Berufskategorien. Verwandte Berufskategorien sind solche die entweder 1.) in der gleichen Berufsuntergruppe sind (4-Steller identisch) oder 2.) eine Sonderfunktion in der gleichen Berufsgruppe einnehmen (3-Steller identisch und 4 Ziffer ist „0“ oder „9“) oder 3.) Berufe beschreiben, welche bei der zugrundeliegenden Kategorie explizit als „nicht einzubeziehende Berufe“ genannt sind.
- ISCO-08-Codes aus beiden Umsteigeschlüsseln für die zugrundeliegende Kategorie.
- ISCO-08-Codes aus beiden Umsteigeschlüsseln für alle verwandten Berufskategorien.
- ISCO-Kategoriebeschreibungen für alle ISCO-08-Kategorien, die 1.) mit der zugrundeliegenden Kategorie assoziiert sind oder 2.) mit einer der verwandten Kategorien assoziiert sind oder 3.) zu den ISCO-Kategorien, die mit der zugrundeliegenden Kategorie assoziiert sind, verwandt sind. In diesem Fall sind verwandte Berufskategorien solche die unter „Some related occupations classified elsewhere“ genannt sind.

Für jede beliebige Kategorie der KldB 2010 liegt somit ein Überblick über diese Kategorie und assoziierte Kategorien in der ISCO-08 sowie über verwandte Kategorien in beiden Klassifikationen vor. Im Regelfall sollte dieser Überblick alle Kategorien enthalten, die bei der Erstellung der zugeordneten Kategorien für die Hilfsklassifikation beachtet werden müssen.

Alle Kategorien der KldB 2010 wurden nacheinander abgearbeitet. Auf Basis von Überblicksdokumenten, wie unten exemplarisch für die 92122 dargestellt, wurden Hilfskategorien erstellt, die auf die 92122 verlinken. Falls sich die Tätigkeiten in mehreren Berufskategorien der KldB 2010 stark überschneiden, wurden die entsprechenden Berufskategorien zusammengelegt und die zugehörigen Überblicksdokumente gemeinsam bearbeitet.

Bei der Erstellung ist äußerste Sorgfalt nötig. Zur Qualitätskontrolle wurden fertige Hilfskategorien noch von einer weiteren Person reviewt und nötigenfalls überarbeitet. Die Entwicklung erfolgte hauptsächlich durch studentische Hilfskräfte. Insgesamt wurden einige Tausend Arbeitsstunden zur Entwicklung benötigt.

Abgrenzungen der Kategorien untereinander konnten im ersten Bearbeitungsschritt noch nicht erstellt werden, da die abzugrenzenden Kategorien zu dem Zeitpunkt noch nicht fertig waren. Daher wurden zunächst bloß Abgrenzungen zu KldB-Kategorien aufgenommen, die in einem zweiten, abschließenden Bearbeitungsschritt noch einmal durch Abgrenzungen zu Hilfskategorien (IDs) ersetzt wurden.

Ausgewählte Schwierigkeiten

Bei der Bearbeitung sind zahlreiche Schwierigkeiten aufgetreten, die im Einzelfall entschieden werden mussten. Im Folgenden legen wir für einige ausgewählte Fälle dar, welche Lösungen wir dabei gefunden haben und auf welchen Überlegungen sie beruhen.

Ausbildung und Branche. *FDZ-Methodenreport 04/2012*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg.

Hohe Ähnlichkeit von Berufskategorien

Viele Kategorien aus der KldB 2010 und aus der ISCO-08 weisen untereinander starke Überschneidungen auf. Ein Beispiel derartiger Überschneidung sind die bereits in Abschnitt 3.2. verwendeten ISCO-08-Kategorien „4222 Contact centre information clerks" und „5244 Contact centre salespersons". Beide Kategorien beschreiben Mitarbeiter von Callcentern. Theoretisch lassen sich beide Kategorien gut unterscheiden, denn die einen sind für die Entgegennahme von Anrufen und Information der Kunden zuständig wohingegen die anderen im Telefonverkauf proaktiver tätig sind. In der Praxis ist es aber auch sehr gut möglich, dass eine einzige Person in beiden Bereichen arbeitet und diese theoretische Trennung nicht zielführend ist. Wir schlagen vor, dass in diesem Fall die vorherrschende Tätigkeit (Kerntätigkeit), die am meisten Arbeitszeit in Anspruch nimmt, für die Zuordnung ausschlaggebend sein sollte (vgl. Abschnitt 3). Üblicherweise wurden in solchen Fällen mehrere Hilfskategorien erstellt (vgl. Abschnitt 3.2 für die entsprechende Umsetzung beim Callcenter-Mitarbeiter). In einigen Fällen sind Kategorien aus der ISCO-08 bzw. aus der KldB 2010 aber zueinander so ähnlich, dass die kleinen Unterschiede in getrennten Hilfskategorien nicht entsprechend zur Geltung kommen würden. Anstelle der Erstellung von getrennten Hilfskategorien wurden in diesen Fällen die ähnlichen Kategorien in einer einzigen Hilfskategorie zusammengefasst. Erst eine Folgefrage ermöglicht dann, weitere Details zu erheben auf deren Basis ein Befragter korrekt den ISCO-08 und KldB-2010- Kategorien zugeordnet werden kann.

Umsteigeschlüssel zur Einordnung von Hilfskategorien in ISCO-08

Umsteigeschlüssel von der KldB 2010 und von der DKZ waren für uns maßgeblich, um den Hilfskategorien Kategorien aus der ISCO-08 zuzuordnen. Dabei treten zwei Schwierigkeiten auf:

- Offenbar existieren in der KldB 2010 einige Berufe zu denen die ISCO-08 keine passgenauen Kategorien bereithält. Ist zum Beispiel der „Lebensmitteltechniker/in - Fischerzeugnisse" (KldB-Kategorie 29243), der die industrielle Herstellung von Fischerzeugnissen überwacht, wirklich als ein Chemiebetriebstechniker (ISCO-Kategorie 3116) anzusehen? In Ermangelung einer besser passenden ISCO-08-Kategorie nehmen wir diese Zuordnung vor, wobei wir uns bei entsprechenden Entscheidungen meist am Umsteigeschlüssel der KldB 2010 orientiert haben.
- In einigen Fällen (und häufiger die DKZ-Umstiege betreffend) halten wir die Zuordnungen aus den Umsteigeschlüsseln für unpassend. Wenn eine besser passende Kategorie ins Auge fällt, wurde der Fehler nicht übernommen und diese besser passende Kategorie aus der ISCO-08 wurde der Hilfskategorie stattdessen zugeordnet.

Behandlung von Informatikern in der KldB 2010

Die KldB 2010 verwendet zwei unterschiedliche berufsfachliche Dimensionen für Informatiker. Einerseits erfolgt eine Gliederung nach Anwendungsfeld (z.B. 43114 Wirtschaftsinformatik, 43124 Technische Informatik, 43134 Bio- und Medizininformatik), zugleich erfolgt aber auch eine Gliederung nach Aufgabenbereich (z.B. 43214 IT-Systemanalyse, 43343 IT-Systemadministration, 43353 Datenbankentwicklung und –administration, 43414 Softwareentwicklung). In diesem Fall ist es gut möglich, dass die berufliche Tätigkeit des Befragten sich mehr als einer einzigen Berufskategorie zuordnen lässt. Da die Kategorien in der KldB 2010 nicht disjunkt sind, lässt sich eine eindeutige Zuordnung auch nicht für die Hilfsklassifikation erreichen.

Diese Schwierigkeit berücksichtigen wir nicht weiter, sondern übernehmen die wichtigsten Aspekte aus den Definitionen der KldB-Berufskategorien für unsere Beschreibungen der Hilfskategorien. Der Befragte kann auf diese Weise selber entscheiden, ob sein Beruf sich besser über sein Anwendungsfeld oder sein Aufgabenbereich beschreiben lässt.

Entscheidet sich der Befragte für eine Klassifizierung nach Anwendungsfeld (z.B. Wirtschaftsinformatik, dessen berufliche Tätigkeit wir in der Hilfsklassifikation beschreiben als „Planung und Einführung neuer IT-Lösungen für das eigene Unternehmen“), gibt es eine weitere Herausforderung: Einerseits erfordert die KldB 2010 eine Bestimmung des Anforderungsniveaus, weshalb wir eine Folgefrage nach der üblicherweise erforderlichen Ausbildung stellen. Andererseits sieht die ISCO-08 keine Gliederung nach dem Anwendungsfeld vor, sondern erfordert die Bestimmung des Aufgabenbereiches, weshalb eine weitere Folgefrage erforderlich ist. Dieses Projekt ist zwar mit dem Ziel gestartet, immer bloß maximal eine Folgefrage zuzulassen, aber im Endprodukt wurden einige Ausnahmen von dieser Regel nötig.

Behandlung von Büro- und Sekretariatskräften

Bisher wird bei der Kodierung besonders häufig die Berufskategorie 71402 „Büro- und Sekretariatskräfte (ohne Spezialisierung)“ ausgewählt. Berufsbezeichnungen wie „Bürokaufmann“ und „Sekretär“ sind dieser Berufskategorie zugeordnet. Auch die Definition dieser Berufskategorie ist sehr umfassend ausgerichtet. Der Umsteigeschlüssel der KldB 2010 sieht für diese Berufskategorie die ISCO-Kategorie 4120 „Secretaries (general)“ vor. Allerdings ist diese ISCO-Kategorie deutlich enger gefasst als die KldB-Kategorie, denn nur wenn die Tätigkeit hauptsächlich mit „transcription, formatting and processing of correspondence and other documentation“ befasst ist, soll sie der ISCO-Kategorie 4120 zugeordnet werden. Wer hingegen allgemein Büroarbeit und administrative Tätigkeiten ausübt, soll in der ISCO-08 der deutlich allgemeiner formulierten Kategorie 4110 „General office clerks“ zugeordnet werden.

Diese Zuordnungen aus der KldB 2010 haben bedeutende Konsequenzen: Wenn jemand nur antwortet er sei „Bürokaufmann“, könnte man ihn nach traditionellen Verfahren den Berufskategorien 71402 (KldB) bzw. 4120 (ISCO) zuordnen. Beides wird allerdings keineswegs für alle Bürokaufleute richtig sein, denn beispielsweise Bürokaufleute, die in der Buchhaltung tätig sind, müssten ganz anderen Berufskategorien zugeordnet werden. Sowohl die Zuordnung der Berufsbezeichnung als auch die Zuordnung der KldB-Berufskategorie zur ISCO können daher für uns nicht ausschlaggebend sein, sondern wir orientieren uns an den Definitionen der Berufskategorien. Jedoch sind unsere Interpretation der Definitionen und unsere Formulierungen der Hilfskategorien dabei ausschlaggebend und werden die Ergebnisse einer Kodierung anhand der Hilfsklassifikation stark beeinflussen. Da die Definition der 71402 (und anderer Berufskategorien) aber nicht besonders präzise ist, entstehen Freiheitsgrade, die wir mit unseren Formulierungen füllen.

Da sich die Tätigkeiten von 71402 und 71403 nicht zu sehr unterscheiden, legen wir beides in einer einzigen Hilfskategorie zusammen. Die dort beschriebene Tätigkeit lautet: „Sekretariatstätigkeit mit Aufgaben wie Korrespondenz, Terminplanung, Finanzen und Büroorganisation“. Auch wenn diese Beschreibung für die Tätigkeit eines Befragten zutrifft, sind noch weitere Berufskategorien zusätzlich zur 71402 und 71403 möglich. Um die passendste zu finden, wird eine Folgefrage gestellt. Damit die Tätigkeit so genau wie möglich verschlüsselt werden kann, werden konkrete Tätigkeiten dabei zuerst abgefragt (z.B. „persönliche Unterstützung von Führungskräften bei organisatorischen Aufgaben“) und allgemeine Tätigkeiten erst spät (z.B. „Erledigung verschiedener Büro- und Verwaltungstätigkeiten nach vorgegebenen Verfahren“).

Wer eine Berufsausbildung als „Bürokaufmann/-frau“ absolviert hat, antwortet gegebenenfalls mit dieser Berufsbenennung, obwohl die Berufskategorie 71402 für seine berufliche Tätigkeit möglicherweise nicht zutreffend ist. Zahlreiche Berufskategorien sind ähnlich zur 71402/3. Wünschenswert wäre es, wenn diese ähnlichen Berufskategorien zur Auswahl im computergestützten Vorschlagssystem eingeblendet werden, sodass der Befragte nicht fälschlicherweise diese Hilfskategorie auswählt. Es ist aber zu befürchten, dass der Platz zur Anzeige aller abgegrenzten

Hilfskategorien nicht ausreicht. Daher ist speziell für die Antwort „Bürokaufmann/-frau“ auch denkbar, zunächst eine Folgefrage zu stellen, die etwa wie folgt formuliert sein könnte: „Ist Ihre Tätigkeit in einem der folgenden Bereiche spezialisiert?“ mit Antwortoptionen „Fachliche Planung der Produktion“/“Einkauf und Beschaffung“/“Kundenmanagement, Marketing, Absatz“/“Buchhaltung, Kostenrechnung, Controlling“/“Recht“/“Personal und betriebliche Organisation“/“keine Spezialisierung“. Abhängig von der Antwort könnte man dann zur Anzeige die auf keinen Fall zutreffenden Hilfskategorien herausfiltern.

Ausgelassene Residualkategorien

Bei einigen Kategorien aus der KldB 2010, insbesondere bei Residualkategorien für sonstige spezifische Berufe (4. Ziffer = „8“, vgl. Abschnitt 2.3.4), weisen die Kategoriebeschreibungen sehr starke Überschneidungen zu anderen Kategorien auf. Sofern für uns keine unterschiedlichen Tätigkeiten in den verschiedenen Kategorien erkennbar sind oder andersartige Probleme bestehen, wurden derartige Kategorien nicht in die Hilfsklassifikation aufgenommen. Folgende Kategorien wurden daher bei der Bearbeitung ausgelassen: 11183, 11402, 22182, 22183, 22184, 26382, 26383, 41383, 71382, 71383, 72214, 73282, 73283, 73284, 73293, 81382, 81784, 82283, 91484. Nur sehr unvollständig wurde die 81783 bearbeitet.

Beispiel: Verwendetes Material zur Bearbeitung der KldB-Berufskategorie 92122 Dialogmarketing - Fachkraft

Wie bereits erwähnt wurde, haben wir vollautomatisch (und daher möglicherweise fehlerhaft) für jede Berufskategorie alle ggf. relevante Informationen zusammengestellt. Dies soll nun an einem Beispiel illustriert werden. Auf Basis des folgenden Materials wurden die im Artikel beispielhaft genannten Hilfskategorien zum „Callcenteragent/in (Inbound)“ und zum „Verkäufer/in – Telemarketing“ entwickelt.

Berufe im Dialogmarketing - fachlich ausgerichtete Tätigkeiten

921 Werbung und Marketing

92 Werbung, Marketing, kaufmännische und redaktionelle Medienberufe

9 Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Kategoriebeschreibung

Inhalt:

Diese Systematikposition umfasst alle Berufe im Dialogmarketing, deren Tätigkeiten fundierte fachliche Kenntnisse und Fertigkeiten erfordern. Angehörige dieser Berufe kontaktieren bestehende oder potenzielle Kunden/Kundinnen per Telefon oder über sonstige elektronische Kommunikationsmedien, um Waren und Dienstleistungen zu bewerben, Abschlüsse zu tätigen oder Verkaufsbesuche zu vereinbaren. Zudem sind sie in der Kundeninformation tätig und beantworten Kundenanfragen.

Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, üblicherweise:

- telefonische Gespräche, Nachrichten oder Bestellungen entgegennehmen und bearbeiten, Anforderungen feststellen, Auskünfte erteilen oder Termine vereinbaren
- Kunden und Kundinnen über zusätzliche Produkte und Dienstleistungen beraten
- Waren und Dienstleistungen per Telefon oder über E-Mail unter Einhaltung formaler Abläufe (Scripts) und nach Kontaktlisten bewerben
- Interesse an Waren und Dienstleistungen wecken und nach einem Verkaufsabschluss oder einer Terminvereinbarung mit Handelsvertreter/innen streben
- die Bearbeitung und Versendung von Waren bzw. die Erbringung von Dienstleistungen organisieren, Informationspaket und Broschüren an die Kunden und Kundinnen übermitteln
- computergestützte Aufzeichnungen zu den getätigten Telefongesprächen, E-mail-Kontakten und den erzielten Erfolgen führen und den Vorgesetzten vorlegen

Zugeordnete Berufe (Beispiele):

Call-Center-Agent/in Fachkaufmann/-frau – Teleservice Kaufmann/-frau – Dialogmarketing Servicefachkraft – Dialogmarketing

Aus der DKZ (codenr [id] Bezeichnung):

- 92122-104 [35308] Servicefachkraft - Dialogmarketing (Tätigkeitsinhalte, Fähigkeiten)

- 92122-105 [35310] Kaufmann/-frau - Dialogmarketing (Tätigkeitsinhalte, Fähigkeiten)
- 92122-101 [7001] Callcenteragent/in (Tätigkeitsinhalte, Fähigkeiten)
- 92122-102 [14107] Fachkaufmann/-frau - Teleservice (Tätigkeitsinhalte, Fähigkeiten)
- 92122-103 [14994] E-Mail-Agent/in (Tätigkeitsinhalte, Fähigkeiten)

Aus dem alphabetischen Verzeichnis:

Callcenteragent/in, E-Mail-Agent/in, Fachkaufmann/-frau - Teleservice, Fachkraft - Merchandising, Fachkraft - Telefonmarketing für Blinde und Sehbehinderte, Hotline-Mitarbeiter/in, Kaufmann/-frau - Dialogmarketing, Kaufmann/-frau - Telekommunikation, Servicefachkraft - Dialogmarketing, Telefonagent/in, Telefonverkäufer/in, Telekommunikationskaufmann/-frau

Nicht einzubeziehende Berufe:

- Kaufmännische/r Assistent/in, Wirtschaftsassistent/in Werbung (92112 Werbung und Marketing - Fachkraft)
- Kaufmann/-frau Marketingkommunikation (92112 Werbung und Marketing - Fachkraft)
- Anzeigenverkäufer/in (92382 Verlags-, Medienkaufleute(ssT)-Fachkraft)

Aus einer Telefonumfrage die 20 häufigsten Freitextantworten, die in diese Kategorie kodiert wurden (nur zur ersten Frage nach dem Beruf; absolute Häufigkeit der Nennung in Klammern):

Callcenter-Agent (12), Callcenteragent (9), Call Center Agent (8), Call-Center-Agent (5), Telemarketing (5), Telefonmarketing (4), Callcenter Agent (4), Call-Center Agent (4), Call-Center-Agentin (3), Call Agent (3), Call Agentin (3), Trafficer+Produktioner (3), Callcenter Agentin (2), Telefonagentin (2), Callcenter agent (2), Callcenter (2), Callcentermitarbeiter (2), Coulzentermitarbeiter (1), Collagent (1), Im callcenter (1)

Verwandte Berufskategorien KldB 2010

Bei Bedarf müssen weitere Definitionen aus der KldB, Band 2, und Beschreibungen aus dem BERUFENET zum Verständnis der Kategorien und Begriffe herangezogen werden.

Aus der Berufsgruppe

921 Werbung und Marketing

Enthält zugeordnete Aufsichts-und Führungskräfte (4. Stelle = 9), die Kategorie ohne Spezialisierung (4. Stelle = 0) sowie alle Kategorien aus der gleichen Berufsuntergruppe (gleiche 4-Steller).

92194 Führungskräfte - Werbung und Marketing

921 Werbung und Marketing

92 Werbung, Marketing, kaufmännische und redaktionelle Medienberufe

9 Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Inhalt:

Angehörige dieser Berufe übernehmen Führungsaufgaben in Werbung und Marketing, welche einen hohen Komplexitätsgrad aufweisen und ein entsprechend hohes Kenntnis- und Fertigkeitsniveau erfordern. Sie orga-nisieren, planen, koordinieren und überwachen die Werbeaktivitäten eines Unternehmens.

Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, üblicherweise:

- die Aktivitäten eines Unternehmens oder einer Organisation in den Bereichen Werbung und Marketing planen, leiten und koordinieren
- Kommunikationsstrategie ausarbeiten, zentrale Botschaften formulieren, Kreativbriefings für die Zusammenarbeit mit Agenturen erstellen und Gestaltungskonzepte beurteilen
- Mediaplan erstellen und führen, die gesamte Einkaufsverantwortung übernehmen
- Verträgen mit Kunden/Kundinnen oder Zeitungen, Radio- oder Fernsehsendern und Werbeagenturen aushandeln
- die Mitarbeiter/innen im Bereich Werbung und Marketing leiten und führen
- betriebliche und administrative Verfahren festlegen und leiten
- Budget festlegen und verwalten, Ausgaben kontrollieren und einen effizienten Ressourceneinsatz sicherstellen
- die Auswahl, Schulung und Leistung von Mitarbeiter/innen überwachen

Zugeordnete Berufe (Beispiele):

Call-Center-Manager/in Marketingleiter/in Werbeleiter/in

Aus der DKZ:

Werbeleiter/in, Leiter/in - Marketing

Nicht einzubeziehende Berufe:

- Vertriebsleiter/in (61194 Führung - Einkauf und Vertrieb)
- Marketingsbetriebswirt/in (Hochschule) (92114 Werbung und Marketing - Experte)
- Leiter/in Presse- und Öffentlichkeitsarbeit (92294 Führung - Öffentlichkeitsarbeit)
- Verleger/in (Medien, Musik) (92394 Führung - Verlags- und Medienwirtschaft)

92123 Berufe im Dialogmarketing - komplexe Spezialistentätigkeiten

921 Werbung und Marketing

92 Werbung, Marketing, kaufmännische und redaktionelle Medienberufe

9 Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Inhalt:

Diese Systematikposition umfasst alle Berufe im Dialogmarketing, deren Tätigkeiten Spezialkenntnisse und -fertigkeiten erfordern. Angehörige dieser Berufe übernehmen organisatorische Aufgaben in Call-Centern und beaufsichtigen die Arbeit von Bürokräften im Dialogmarketing.

Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, üblicherweise:

- die Arbeit der Bürokräfte im Dialogmarketing beaufsichtigen und koordinieren
- Arbeitseinsatz planen, Einsatzzeiten, Pausen sowie Arbeitsaufgaben festlegen und zuteilen
- arbeitsbezogene Probleme klären, Fortschritts- und andere Berichte erstellen und der Geschäfts-leitung vorlegen
- Mitarbeiter/innen im Zusammenhang mit den Arbeitsaufgaben, Sicherheitsverfahren und Unterneh-mensrichtlinien schulen und anleiten oder die Durchführung von Schulungen veranlassen
- die Arbeitsleistung von Mitarbeiter/innen evaluieren, entsprechende Personalmaßnahmen empfehlen
- bei Rekrutierung, Befragung und Auswahl von Mitarbeiter/innen unterstützen

Zugeordnete Berufe (Beispiele):

Call-Center-Fachwirt/in Call-Center-Trainer/in Teamleiter/in Call-Center

Aus der DKZ:

Teamleiter/in - Callcenter, Trainer/in, Supervisor/in - Callcenter, Fachwirt/in - Callcenter, Betriebswirt/in (Fachschule) - Callcentermanagement

Nicht einzubeziehende Berufe:

- Fachberater/in Vertrieb (61123 Vertrieb (außer IKT) - Spezialist)
- Vertriebsleiter/in (61194 Führung - Einkauf und Vertrieb)
- Fachkaufmann/-frau Werbung und Kommunikation (92113 Werbung und Marketing - Spezialist)
- Werbeleiter/in (92194 Führung - Werbung und Marketing)

Nicht einzubeziehende Kategorien (von oben)

92112 Berufe in Werbung und Marketing - fachlich ausgerichtete Tätigkeiten

921 Werbung und Marketing

92 Werbung, Marketing, kaufmännische und redaktionelle Medienberufe

9 Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Inhalt:

Diese Systematikposition umfasst alle Berufe in Werbung und Marketing, deren Tätigkeiten fundierte fachliche Kenntnisse und Fertigkeiten erfordern. Angehörige dieser Berufe wirken dabei mit, Produkte und Dienst-leistungen bekannt zu machen und die Nachfrage zu

steigern. Sie führen die ihnen übertragenen Aufgaben unter der Anleitung von akademischen Fachkräften aus.

Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, üblicherweise:

- Aufträge von Kunden und Kundinnen für Werbe- und Marketingaktionen entgegennehmen sowie die Korrespondenz mit den Kunden und Kundinnen vorbereiten bzw. nach Absprache führen
- bei der Entwicklung von Kommunikationskonzepten für medienübergreifende Kampagnen oder für Einzelmaßnahmen assistieren
- interne und externe Herstellungsprozesse überwachen, z.B. kontrollieren, ob Satz und Layout des Prospektes und die Tonalität des Textes mit den Vorgaben aus dem Kommunikationskonzept über-einstimmen, ob die Farben und Schriften dem Corporate Design des Kunden entsprechen und ob der zeitliche Aufwand gemäß der Budgetplanung eingehalten wird
- bei der Vertragsgestaltung mitwirken, z.B. Rechte und Lizenzen für Bilder einholen
- Rechnungen für die erbrachten Leistungen erstellen

Zugeordnete Berufe (Beispiele):

Kaufmann/-frau Marketingkommunikation Kaufmännische/r Assistent/in, Wirtschaftsassistent/in Werbung Marketingfachkraft, -assistent/in Werbekaufmann/-frau

Aus der DKZ:

Kfm. Ass./Wirtschaftsassistent/in - Werbung, Werbekaufmann/-frau, Kaufmann/-frau - Marketingkommunikation, Marketingfachkraft/-assistent/in

Nicht einzubeziehende Berufe:

- Mediengestalter/in Digital und Print (23212 Digital-,Printmediengestaltung-Fachkraft)
- Kaufmann/-frau Dialogmarketing (92122 Dialogmarketing - Fachkraft)
- Kaufmännische/r Assistent/in, Wirtschaftsassistent/in Medien (92302 Verlags-,Medienkaufleute(oS) - Fachkraft)
- Anzeigenverkäufer/in (92382 Verlags-, Medienkaufleute(ssT)-Fachkraft)

92382 Verlags- und Medienkaufleute (sonstige spezifische Tätigkeitsangabe) - fachlich ausgerichtete Tätigkeiten

923 Verlags- und Medienwirtschaft

92 Werbung, Marketing, kaufmännische und redaktionelle Medienberufe

9 Sprach-, Literatur-, Geistes-, Gesellschafts- und Wirtschaftswissenschaften, Medien, Kunst, Kultur und Gestaltung

Inhalt:

Diese Systematikposition umfasst alle Verlags- und Medienkaufleute, deren Tätigkeiten fundierte fachliche Kenntnisse und Fertigkeiten erfordern und die in der übergeordneten Systematikposition 923 Verlags- und Medienwirtschaft nicht anderweitig erfasst sind.

Angehörige dieser Berufe verkaufen z.B. Anzeigen in Medien oder arbeiten im Filmbetrieb mit.

Aufgaben, Tätigkeiten, Kenntnisse und Fertigkeiten, üblicherweise:

- für Print- und elektronische Medien Anzeigenflächen verkaufen, Anzeigenkunden/-kundinnen betreuen und das Neukundengeschäft auf- und ausbauen
- über spezifische Werbewirkungen, die Verwendung von Schriftarten, Fotos, Logos und anderen gestalterischen Elementen sowie über Größe und Preise der Anzeigen informieren
- mögliche Sponsoren oder Förderstellen für Filmproduktionen recherchieren und der Produktions-leitung vorschlagen
- eingehende Rechnungen prüfen und nach Absprache mit der Filmproduktionsleitung Zahlungen abwickeln

Zugeordnete Berufe (Beispiele):

Anzeigenverkäufer/in Assistent/in Filmgeschäftsführung

Aus der DKZ:

Anzeigenverkäufer/in, Assistent/in - Filmgeschäftsführung

Nicht einzubeziehende Berufe:

- Mediengestalter/in Digital und Print (23212 Digital-,Printmediengestaltung-Fachkraft)
- Buchhändler/in (62512 Buchhandel - Fachkraft)
- Kaufmann/-frau Marketingkommunikation (92112 Werbung und Marketing - Fachkraft)
- Kaufmann/-frau Dialogmarketing (92122 Dialogmarketing - Fachkraft)
- Medienkaufmann/-frau Digital und Print (92302 Verlags-,Medienkaufleute(oS) - Fachkraft)
- Redaktionsassistent/in (92412 Redakteure, Journalisten - Fachkraft)

Umsteigeschlüssel

Für 92122 Dialogmarketing - Fachkraft

Aus dem offiziellen Umsteigeschlüssel (erster Eintrag = Schwerpunkt):

- 4222 Kundeninformationsfachkräfte in Call Centers
- 5244 Telefonverkäufer

Aus dem DKZ-Umsteigeschlüssel (unsortiert):

- 3341 Office supervisors (für die Berufe: Kaufmann/-frau - Dialogmarketing)
- 4222 Contact centre information clerks (für die Berufe: Servicefachkraft - Dialogmarketing, Callcenteragent/in, Fachkaufmann/-frau - Teleservice, E-Mail-Agent/in)

Für verwandte Berufskategorien

Aus dem offiziellen Umsteigeschlüssel:

KldB	Verwandte Berufskategorien KldB 2010	ISCO	Bezeichnung
92194	Führungskräfte - Werbung und Marketing	1222	Führungskräfte in Werbung und Öffentlichkeitsarbeit
92123	Berufe im Dialogmarketing - komplexe Spezialistentätigkeiten	3341	Sekretariatsleiter
92112	Berufe in Werbung und Marketing - fachlich ausgerichtete Tätigkeiten	4419	Bürokräfte und verwandte Berufe, anderweitig nicht genannt
92382	Verlags- und Medienkaufleute (sonstige spezifische Tätigkeitsangabe) - fachlich ausgerichtete Tätigkeiten	4419	Bürokräfte und verwandte Berufe, anderweitig nicht genannt

Aus dem DKZ-Umsteigeschlüssel (unsortiert):

KldB	DKZ-Berufsbezeichnungen	ISCO	Bezeichnung
92194	Leiter/in - Marketing	1221	Sales and marketing managers
92194	Werbeleiter/in	2431	Advertising and marketing professionals
92123	Trainer/in, Supervisor/in - Callcenter	2320	Vocational education teachers
92123	Teamleiter/in - Callcenter, Fachwirt/in - Callcenter, Betriebswirt/in (Fachschule) - Callcentermanagement	3341	Office supervisors
92112	Kfm. Ass./Wirtschaftsassistent/in - Werbung, Werbekaufmann/-frau, Kaufmann/-frau - Marketingkommunikation,	4419	Clerical support workers not elsewhere classified

	Marketingfachkraft/-assistent/in		
92382	Anzeigenverkäufer/in	3339	Business services agents not elsewhere classified
92382	Assistent/in - Filmgeschäftsführung	4419	Clerical support workers not elsewhere classified

Kategoriebeschreibungen (ISCO-08, deutsch)

Wenn einzelne Formulierungen relevant sind, ist die englische Originalfassung zu konsultieren.

Zugeordnete Kategorien

4222 Kundeninformationsfachkräfte in Call Centers

422 Berufe im Bereich Kundeninformation

42 Bürokräfte mit Kundenkontakt

4 Bürokräfte und verwandte Berufe

Definition:

Kundeninformationsfachkräfte in Call Centers bieten Kunden Beratung und Information, antworten auf Kundenanfragen über die Waren, Dienstleistungen oder Geschäftsbedingungen einer Gesellschaft oder Organisation und bearbeiten Finanztransaktionen mithilfe von Telefon oder elektronischen Kommunikationsmedien wie E-Mail. Diese Dienstleistungen können in Geschäftsräumen erbracht werden, die weit entfernt von den Kunden oder von sonstigen Standorten der Organisationen oder Gesellschaften sind, über die Informationen erteilt werden.

Aufgaben umfassen -

(a) Bearbeitung eingehender Anrufe und Nachrichten von Kunden, sei es zur Beantwortung von Anfragen, zur Abwicklung geforderter Dienstleistungen oder zur Bearbeitung von Beschwerden;
(b) Feststellung der Anforderungen und Eingabe von Ereignissen in ein Computersystem;
(c) Erledigung von Aufgaben für andere Geschäftseinheiten, falls relevant;
(d) Fakturierung oder Bearbeitung von Zahlungen bei Bedarf;
(e) Versand von Briefen, Informationsblättern und sonstigen Dokumenten an Kunden;
(f) Beratung von Kunden über zusätzliche Produkte oder Dienstleistungen.

Beispiele für hier zugeordnete Berufe:

Kundeninformationsfachkraft in Call Center

Nicht in dieser Berufsgattung klassifizierte Berufe:

- Telefonist - s. 4223
- Interviewer in der Marktforschung - s. 4227
- Verkaufskraft im Telemarketing - s. 5244
- Verkaufskraft in Call Centers - s. 5244
- Verkaufskraft in Kundenkontaktzentrum - s. 5244

Anmerkungen

Nur Fachkräfte, die Informationsanfragen beantworten und/oder Transaktionen direkt abwickeln, fallen in die Berufsgattung 4222, Kundeninformationsfachkräfte in Call Centers. Jene, die spezielle Dienstleistungen anbieten wie Reiseberater, werden der entsprechenden Berufsgattung zugeordnet, unabhängig davon, ob sie ihre Tätigkeit in Call Centers verrichten oder nicht.

5244 Telefonverkäufer

524 Sonstige Verkaufskräfte

52 Verkaufskräfte

5 Dienstleistungsberufe und Verkäufer

Definition:

Telefonverkäufer kontaktieren bestehende und potenzielle Kunden per Telefon oder über sonstige elektronische Kommunikationsmedien, um Waren und Dienstleistungen zu bewerben, Abschlüsse zu tätigen und Verkaufsbesuche zu vereinbaren. Sie können von einem Kundenkontaktzentrum oder von einer nicht zentral organisierten Einrichtung aus tätig werden.

Aufgaben umfassen -

(a) Bewerbung von Waren und Dienstleistungen per Telefon oder über E-Mail unter Einhaltung formaler Abläufe (Scripts) und nach Kontaktlisten;
(b) Weckung von Interesse an Waren und Dienstleistungen sowie Streben nach einem Verkaufsabschluss oder einer Terminvereinbarung mit Handelsvertretern;
(c) Organisation der Bearbeitung und Versendung von Waren bzw. der Erbringung von Dienstleistungen, Übermittlung von Informationspaketen und Broschüren an die Kunden;
(d) Vereinbarung von Treffen für Handelsvertreter;
(e) Führung von Aufzeichnungen für nachfolgende Maßnahmen und zur Aktualisierung der Marketing-Datenkbanken anhand des Status der einzelnen Kunden;

(f) Berichtslegung über die Tätigkeit von Mitbewerbern und über Fragen, die sich im Zuge der Kundenkontakte stellen, an die Vorgesetzten;
(g) Führung von Statistiken über Kundenbesuche und dabei erzielte Erfolge;
(h) Vorlage regelmäßiger Berichte über Telemarketing-Tätigkeiten und Ergebnisse.

Beispiele für hier zugeordnete Berufe:

Verkaufskraft im Telemarketing

Verkaufskraft in Call Centers

Verkaufskraft in Kundenkontaktzentrum

Internet-Verkaufskraft

Telemarketer

Nicht in dieser Berufsgattung klassifizierte Berufe:

- Kundeninformationsfachkraft in Call Centers - s. 4222

Anmerkungen

3341 Sekretariatsleiter

334 Sekretariatsfachkräfte

33 Nicht akademische betriebswirtschaftliche und kaufmännische Fachkräfte und Verwaltungsfachkräfte

3 Techniker und gleichrangige nichttechnische Berufe

Definition:

Sekretariatsleiter beaufsichtigen und koordinieren die Aktivitäten von Arbeitnehmern in Hauptgruppe 4, Bürokräfte und verwandte Berufe.

Aufgaben umfassen -

(a) Koordinierung, Zuteilung und Prüfung der Arbeit von Bürokräften, die folgende Aufgaben erfüllen: Textverarbeitung, Führung und Ablage von Aufzeichnungen, Betätigung von Telefonen und Telefonanlagen; Dateneingabe, Desktop-Publishing und andere Aktivitäten wie allgemeine Büro- und Verwaltungstätigkeiten;
(b) Festlegung von Arbeitsplänen und Verfahren und Koordinierung der Aktivitäten mit anderen Arbeitseinheiten oder Abteilungen;
(c) Klärung von arbeitsbezogenen Problemen und Erstellung und Vorlage von Fortschritts- und anderen Berichten;
(d) Schulung und Anleitung von Mitarbeitern im Zusammenhang mit Arbeitsaufgaben, Sicherheitsverfahren und Unternehmensrichtlinien oder Veranlassung der Durchführung von Schulungen;

(e) Bewertung der Arbeitsleistung von Mitarbeitern und Einhaltung von Regelungen und Empfehlung entsprechender Personalmaßnahmen;
(f) Unterstützung bei Rekrutierung, Befragung und Auswahl von Mitarbeitern.

Beispiele für hier zugeordnete Berufe:

Büroleiter

Datenerfassungsleiter

Registraturleiter

Büropersonalleiter

Nicht in dieser Berufsgattung klassifizierte Berufe:

- Aufsichtskraft in der medizinischen Dokumentation - s. 3252

Anmerkungen

Verwandte Berufskategorien ISCO-08

1221 Führungskräfte in Vertrieb und Marketing

122 Führungskräfte in Vertrieb, Marketing und Entwicklung

12 Führungskräfte im kaufmännischen Bereich

1 Führungskräfte

Definition:

Führungskräfte in Vertrieb und Marketing planen, leiten und koordinieren die Verkaufs- und Marketingaktivitäten eines Unternehmens oder einer Organisation oder von Unternehmen, die anderen Unternehmen und Organisationen Verkaufs- und Marketingdienste anbieten.

Aufgaben umfassen -

(a) Planung und Organisation von speziellen Verkaufs- und Marketingprogrammen auf der Grundlage von Verkaufsaufzeichnungen und Marktbeurteilungen;
(b) Festlegung von Preislisten, Preisnachlaß- und Lieferbedingungen, Verkaufsförderungsbudgets, Verkaufsmethoden, speziellen Initiativen und Kampagnen;
(c) Festlegung und Leitung von operative und administrativen Verfahren im Zusammenhang mit Verkaufs- und Marketingaktivitäten;
(d) Leitung und Management der Tätigkeiten von Verkaufs- und Marketingpersonal;
(e) Planung und Leitung täglicher Abläufe;
(f) Festlegung und Verwaltung von Budgets und Kontrolle der Ausgaben zur Sicherstellung eines effizienten Ressourceneinsatzes;
(g) Überwachung von Auswahl, Aus- und Weiterbildung und Leistung der Mitarbeiter;

(h) Vertretung des Unternehmens oder der Organisation bei Verkaufs- und Marketingkongressen, Fachmessen und auf anderen Foren.

Beispiele für hier zugeordnete Berufe:

Führungskraft im Marketing

Führungskraft im Verkauf

Anmerkungen

1222 Führungskräfte in Werbung und Öffentlichkeitsarbeit

122 Führungskräfte in Vertrieb, Marketing und Entwicklung

12 Führungskräfte im kaufmännischen Bereich

1 Führungskräfte

Definition:

Führungskräfte in Werbung und Öffentlichkeitsarbeit planen, leiten und koordinieren die Aktivitäten von Unternehmen und Organisationen in den Bereichen Werbung, Public Relations und Öffentlichkeitsinformation oder die Aktivitäten von Unternehmen, die anderen Unternehmen und Organisationen ähnliche Dienstleistungen anbieten.

Aufgaben umfassen -

(a) Planung, Leitung und Koordinierung der Aktivitäten eines Unternehmens oder einer Organisation in den Bereichen Werbung und Öffentlichkeitsarbeit;
(b) Aushandlung von Verträgen mit Kunden oder Zeitungen, Radio- oder Fernsehsendern, Sport- und Kulturorganisationen und Werbeagenturen;
(c) Planung und Verwaltung von Informationsprogrammen zur Information von Gesetzgebern, Massenmedien und allgemeiner Öffentlichkeit über Pläne, Leistungen und Standpunkte von Unternehmen oder Organisationen;
(d) Leitung und Führung der Tätigkeiten von Werbe- und Public Relations-Personal;
(e) Festlegung und Verwaltung von Budgets, Kontrolle von Ausgaben und Sicherstellung eines effizienten Ressourceneinsatzes;
(f) Festlegung und Leitung operativer und administrativer Verfahren;
(g) Planung und Leitung der täglichen Aktivitäten;
(h) Überwachung von Auswahl, Aus- und Weiterbildung und Leistung von Mitarbeitern.

Beispiele für hier zugeordnete Berufe:

Führungskraft in der Werbung

Führungskraft in der Öffentlichkeitsarbeit

Anmerkungen

2320 Lehrkräfte im Bereich Berufsbildung

232 Lehrkräfte im Bereich Berufsbildung

23 Lehrkräfte

2 Akademische Berufe

Definition:

Lehrkräfte im Bereich Berufsbildung lehren oder vermitteln berufsbildende Fächer in Erwachsenen- und Weiterbildungsinstitutionen und an Schüler von Berufsbildenden Schulen. Sie bereiten Schüler auf die Arbeit in bestimmten Berufen oder Berufsfeldern vor, für die normalerweise keine Universitäts- oder Hochschulbildung erforderlich ist.

Aufgaben umfassen -

(a) Erstellung von Lehrplänen oder Planung von Kursinhalten und Unterrichtsmethoden;
(b) Ermittlung des Ausbildungsbedarfs von Schülern oder Arbeitnehmern und Herstellung und Aufrechterhaltung von Kontakten mit Einzelpersonen, Branchen und anderen Bildungssektoren zur Gewährleistung der Bereitstellung relevanter Aus- und Weiterbildungsprogramme;
(c) Abhaltung von Vorlesungen und Führung von Diskussionen zur Vergrößerung des Wissens und der Kompetenz von Schülern;
(d) Anweisung und Überwachung von Schülern in der Verwendung von Werkzeug, Ausrüstung und Materialien und Verhinderung von Verletzungen und Beschädigungen;
(e) Beobachtung und Evaluierung der Arbeit der Schüler zur Feststellung des Fortschritts, Vermittlung von Feedback und Unterbreitung von Verbesserungsvorschlägen;
(f) Durchführung von mündlichen oder schriftlichen Leistungstests zur Messung des Fortschritts, zur Bewertung der Unterrichtseffektivität und zur Beurteilung der Kompetenz;
(g) Erstellung von Berichten und Führung von Aufzeichnungen wie Noten, Anwesenheitslisten und Details der Schulungsaktivitäten;
(h) Überwachung von Einzel- oder Gruppenprojekten, Schulpraktika, Laborarbeit und anderen Schulungen;
(i) individuelle Unterweisung und instruierende oder fördernde Anweisungen;
(j) Abhaltung von praktischen Übungen zur Unterrichtung und Demonstration von Prinzipien, Techniken, Verfahren oder Methoden bestimmter Fächer.

Beispiele für hier zugeordnete Berufe:

Ausbilder - Automobiltechnik

Ausbilder - Kosmetologie

Lehrer im Bereich Berufsbildung

Nicht in dieser Berufsgattung klassifizierte Berufe:

- Lehrer im Sekundarbereich - s. 2330

Anmerkungen

2431 Akademische und vergleichbare Fachkräfte in Werbung und Marketing

243 Akademische und vergleichbare Fachkräfte in Vertrieb, Marketing und Öffentlichkeitsarbeit

24 Betriebswirte und vergleichbare akademische Berufe

2 Akademische Berufe

Definition:

Akademische und vergleichbare Fachkräfte in Werbung und Marketing entwickeln und koordinieren Werbestrategien und -kampagnen, bestimmen den Markt für neue Güter und Dienstleistungen und identifizieren und entwickeln Marktchancen für neue und bestehende Güter und Dienstleistungen.

Aufgaben umfassen -

(a) Planung, Entwicklung und Organisation von Werberichtlinien und -kampagnen zur Unterstützung der Absatzziele;
(b) Beratung von Unternehmensleitung und Kunden hinsichtlich Strategien und Kampagnen zur Erreichung von Zielmärkten und Schaffung von Kundenbewusstsein und für die effektive Hervorhebung der Eigenschaften von Gütern und Dienstleistungen zu Werbezwecken;
(c) Verfassung von Werbetexten und Mediaskripten sowie Organisation von TV- und Filmproduktionen und Medienplatzierung;
(d) Sammeln und Analysieren von Daten über Konsumentenmuster und -präferenzen;
(e) Interpretation und Prognose aktueller und zukünftiger Konsumtrends;
(f) Erforschung der potenziellen Nachfrage nach neuen Gütern und Dienstleistungen und der entsprechenden Marktmerkmale sowie Erhebung und Analysierung von Daten und anderen statistischen Informationen;
(g) Unterstützung von Wachstum und Entwicklung des Unternehmens durch Festlegung und Umsetzung von Marketingzielen, Richtlinien und Programmen;
(h) Auftragsvergabe für und Durchführung von Marktstudien zur Identifikation von Marktchancen für neue und bestehende Güter und Dienstleistungen;
(i) Beratung hinsichtlich aller Marketingelemente wie Produktmix, Preise, Werbung und Verkaufsförderung, Verkaufs- und Vertriebskanäle.

Beispiele für hier zugeordnete Berufe:

Werbespezialist

Marktforschungsanalytiker

Marketingspezialist

Anmerkungen

3252 Fachkräfte im Bereich medizinische Dokumentation und Information

325 Sonstige Assistenzberufe im Gesundheitswesen

32 Assistenzberufe im Gesundheitswesen

3 Techniker und gleichrangige nichttechnische Berufe

Definition:

Fachkräfte im Bereich medizinische Dokumentation und Information entwickeln, pflegen und setzen Verarbeitungs-, Speicherungs- und Abrufsysteme von Gesundheitsaufzeichnungen in medizinischen Einrichtungen und anderen Gesundheitspflegeeinrichtungen um, zwecks Erfüllung der gesetzlich vorgeschriebenen professionellen, ethischen und administrativen Anforderungen zur Führung von Aufzeichnungen in der Erbringung von Gesundheitsdienstleistungen.

Aufgaben umfassen -

(a) Planung, Entwicklung, Pflege und Betrieb verschiedener Indizes und Speicher- und Abrufsysteme für Gesundheitsaufzeichnungen zur Sammlung, Klassifizierung, Speicherung und Analyse von Informationen;
(b) Transkription, Zusammenstellung und Verarbeitung von medizinischen Patientenaufzeichnungen, Aufnahme- und Entlassungsdokumenten und anderen medizinischen Berichten in Aufzeichnungssysteme zwecks Bereitstellung von Daten für die Überwachung und Überweisung von Patienten und zur Verbesserung von epidemiologischer Überwachung, Forschung, Verrechnung, Kostenkontrolle und Pflegeverbesserung;
(c) Prüfung von Aufzeichnungen im Hinblick auf Vollständigkeit, Richtigkeit und Einhaltung von Bestimmungen;
(d) Übersetzung von narrativen Beschreibungen und numerischen Informationen anhand von medizinischen Aufzeichnungen und anderen Dokumenten über die Erbringung von Gesundheitsdienstleistungen in Codes im Zusammenhang mit Standardklassifikationssystemen;
(e) Schutz der Sicherheit von medizinischen Aufzeichnungen zwecks Sicherstellung der Aufrechterhaltung der Vertraulichkeit und der Freigabe von Informationen an befugte Personen und Behörden gemäß den Bestimmungen;
(f) Beaufsichtigung von Büro- und Verwaltungsfachkräften, die an der Verwaltung von Krankenakten mitwirken.

Beispiele für hier zugeordnete Berufe:

Klinischer Kodierer

Krankheitsregister-Dokumentar

Fachkraft für medizinische Informationsdienste

Analytiker von Krankenakten

Medizinischer Dokumentationsassistent

Medizinischer Dokumentar

Nicht in dieser Berufsgattung klassifizierte Berufe:

- Sekretariatsfachkraft im Gesundheitswesen - s. 3344
- Datenerfasser - s. 4132
- Ablagekraft in der Dokumentation und Registratur - s. 4415

Anmerkungen

Die in dieser Berufsgattung erfassten Berufe erfordern normalerweise die Kenntnis von medizinischer Terminologie, rechtlichen Aspekten von Gesundheitsinformationen, Gesundheitsdatenstandards und computer- oder papiergestützter Datenverwaltung, die durch formelle Bildung und/oder praktisches Training erworben wird.

3339 Fachkräfte für unternehmensbezogene Dienstleistungen, anderweitig nicht genannt

333 Fachkräfte für unternehmensbezogene Dienstleistungen

33 Nicht akademische betriebswirtschaftliche und kaufmännische Fachkräfte und Verwaltungsfachkräfte

3 Techniker und gleichrangige nichttechnische Berufe

Definition:

Diese Berufsgattung beinhaltet Fachkräfte für unternehmensbezogene Dienstleistungen, die in Untergruppe 333, Fachkräfte für unternehmensbezogene Dienstleistungen, anderweitig nicht genannt sind. Diese Berufsgattung beinhaltet zum Beispiel Personen, die Geschäftskontakte herstellen, unternehmensbezogene Dienstleistungen verkaufen (wie z.B. Werbeflächen in Medien), Verträge für Auftritte von Sportlern, Unterhaltern und Künstlern, für die Veröffentlichung von Büchern, die Produktion von Theaterstücken oder für Aufzeichnung, Darbietung und Verkauf von Musik arrangieren und Vermögenswerte und Waren in Auktionen verkaufen.

Aufgaben umfassen -

(a) Beschaffung von Informationen über zu verkaufende Dienstleistungen und über die Bedürfnisse der potenziellen Käufer;
(b) Aushandlung von Verträgen im Auftrag von Verkäufer oder Käufer und Erklärung der Kauf- und Zahlungsbedingungen für den Kunden;
(c) Unterzeichnung von Verträgen im Auftrag von Verkäufer oder Käufer und Sicherstellung, dass der Vertrag eingehalten wird;
(d) Sicherstellung, dass die gekaufte unternehmensbezogene Dienstleistung für den Käufer in der vereinbarten Art innerhalb der vereinbarten Frist bereitgestellt wird;
(e) Versteigerung von Vermögenswerten wie Autos, Waren, Viehbeständen, Kunst, Schmuck und anderen Gegenständen verschiedener Art.

Beispiele für hier zugeordnete Berufe:

Auktionator

Werbungsverkäufer

Literaturagent

Musikagent

Sportagent

Theateragent

Anmerkungen

4223 Telefonisten

422 Berufe im Bereich Kundeninformation

42 Bürokräfte mit Kundenkontakt

4 Bürokräfte und verwandte Berufe

Definition:

Telefonisten bedienen Telefonanlagen und -konsolen, um Telefonverbindungen herzustellen, nehmen Anfragen von Anrufern und Problemberichte entgegen und zeichnen Nachrichten an Personal oder Kunden auf und leiten diese weiter.

Aufgaben umfassen -

(a) Bedienung von Telefonanlagen und -konsolen, um Telefonanrufe zu verbinden, zu halten, weiterzuleiten und zu beenden;
(b) Herstellung von Verbindungen für ausgehende Telefonanrufe;
(c) Bearbeitung telefonischer Anfragen und Aufzeichnungen von Nachrichten;
(d) Weiterleitung von Nachrichten an Personal oder Kunden;
(e) Untersuchung von Problemen des Betriebssystems und Information des Reparaturdienstes.

Beispiele für hier zugeordnete Berufe:

Telefonauftragsdienstfachkraft

Telefonist

Anmerkungen

4227 Interviewer im Bereich Umfragen und Marktforschung

422 Berufe im Bereich Kundeninformation

42 Bürokräfte mit Kundenkontakt

4 Bürokräfte und verwandte Berufe

Definition:

Interviewer im Bereich Umfragen und Marktforschung befragen Personen und zeichnen deren Antworten auf Fragen im Bereich Umfragen und Marktforschung zu einer Vielzahl von Themen auf.

Aufgaben umfassen -

(a) telefonische oder persönliche Kontaktaufnahme zu Personen und Erläuterung des Zwecks des Interviews;
(b) Stellung von Fragen anhand von Fragebögen und Umfragen;
(c) Aufzeichnung der Antworten auf Papier oder direkte Eingabe der Antworten in eine Computer-Datenbank mithilfe computergestützter Befragungssysteme;
(d) Feststellung und Beseitigung von Inkonsistenzen in den Antworten;
(e) Feedback an die Sponsoren der Umfrage über Probleme bei der Einholung stichhaltiger Daten.

Beispiele für hier zugeordnete Berufe:

Interviewer in der Marktforschung

Interviewer in der Meinungsforschung

Interviewer im Bereich Umfragen

Anmerkungen

4419 Bürokräfte und verwandte Berufe, anderweitig nicht genannt

441 Sonstige Bürokräfte und verwandte Berufe

44 Sonstige Bürokräfte und verwandte Berufe

4 Bürokräfte und verwandte Berufe

Definition:

Diese Berufsgattung umfasst Bürokräfte, die anderweitig in Hauptgruppe 4, Bürokräfte und verwandte Berufe, nicht genannt sind. Die Berufsgattung umfasst beispielsweise Fremdsprachenkorrespondenten, Zeitungsausschneider und Redaktionsassistenten.

Aufgaben umfassen -

(a) Entgegennahme der Aufträge von Kunden für bestimmte Werbeschaltungen, Schreiben und Bearbeitung, Berechnung von Werbetarifen und Fakturierung;
(b) Schreiben von Korrespondenz für Unternehmen und staatliche Stellen wie Antworten auf Informations- und Unterstützungsanfragen, Schadensersatzforderungen, Gutschrifts- und Fakturierungsanfragen sowie Servicebeanstandungen;

(c) Hilfestellung bei der Herstellung von Zeitschriften, Werbeschaltungen, Katalogen, Verzeichnissen und sonstigen zur Veröffentlichung bestimmten Materialien;
(d) Lesen von Zeitungen, Magazinen, Pressemitteilungen und sonstigen Publikationen, um Artikel zu finden und abzulegen, die für Personal und Kunden von Interesse sind.

Beispiele für hier zugeordnete Berufe:

Anzeigenverkäufer

Fremdsprachenkorrespondent

Bürokraft für das Erstellen von Verzeichnissen

Redaktionsassistent

Zeitungsausschneider

Anmerkungen

Kategoriebeschreibungen (ISCO-08, englisch)

Wenn einzelne Formulierungen relevant sind, ist die englische Originalfassung zu konsultieren.

Zugeordnete Kategorien

4222 Contact centre information clerks

422 Client information workers

42 Customer services clerks

4 Clerical support workers

Definition:

Contact centre information clerks provide advice and information to clients, respond to queries regarding a company?s or an organization?s goods, services or policies, and process financial transactions using the telephone or electronic communications media, such as email. They are located in premises that may be remote from clients or other operations of the organizations or companies about whom information is provided.

Tasks include -

(a) dealing with incoming calls and messages from clients, whether to answer queries, handle calls for service or sort out complaints;
(b) identifying requirements and entering events into a computer system;
(c) dispatching tasks to other units, when relevant;
(d) invoicing or handling payments, where necessary;
(e) sending letters, information sheets and other documents to clients;
(f) advising clients of additional products or services.

Examples of the occupations classified here:

Customer contact centre information clerk

Some related occupations classified elsewhere:

- Telemarketing salesperson ? s. 5244
- Call centre salesperson ? s. 5244
- Customer contact centre salesperson ? s. 5244
- Telephone operator ? s. 4223
- Market research interviewer ? s. 4227

Notes

Only workers who respond to requests for information and/or handle straightforward transactions are classified in Unit group 4222, Contact Centre Information Clerks. Those who provide specialized services, such as travel consultants, are classified in the relevant specialized group whether or not they are located in customer contact centres.

5244 Contact centre salespersons

524 Other sales workers

52 Sales workers

5 Service and sales workers

Definition:

Contact centre salespersons contact existing and prospective customers, using the telephone or other electronic communications media, to promote goods and services, obtain sales and arrange sales visits. They may work from a customer contact centre or from non-centralised premises.

Tasks include -

(a) promoting goods and services by telephone or electronic mail, following scripts and working from lists of contacts;
(b) creating interest in goods and services, and seeking a sale or agreement to see sales representatives;
(c) arranging processing and despatch of goods and services, information kits and brochures to customers;
(d) arranging appointments for sales representatives;
(e) recording notes for follow-up action and updating marketing databases to reflect changes to the status of each customer;
(f) reporting competitor activities and issues raised by contacts for attention by managers;
(g) maintaining statistics of calls made and successes achieved;

(h) submitting periodic reports on telemarketing activities and results.

Examples of the occupations classified here:

Telemarketing salesperson

Call centre salesperson

Customer contact centre salesperson

Internet salesperson

Telemarketer

Some related occupations classified elsewhere:

- Contact centre information clerk ? s. 4222

3341 Office supervisors

334 Administrative and specialized secretaries

33 Business and administration associate professionals

3 Technicians and associate professionals

Definition:

Office supervisors supervise and co-ordinate the activities of workers in Major group 4, Clerical support workers.

Tasks include -

(a) coordinating, assigning and reviewing the work of clerks engaged in the following duties: word processing, record keeping and filing, operating telephones and switchboards; data entry, desktop publishing and other activities involving general office and administrative skills;
(b) establishing work schedules and procedures and co-coordinating activities with other work units or departments;
(c) resolving work-related problems and preparing and submitting progress and other reports;
(d) training and instructing employees in job duties, safety procedures and company policies, or arranging for training to be provided;
(e) evaluating employees? job performance and conformance to regulations, and recommending appropriate personnel action;
(f) assisting in recruitment, interviewing, and selection of employees.

Examples of the occupations classified here:

Clerical supervisor

Data entry supervisor

Filing clerks supervisor

Personnel clerks supervisor

Some related occupations classified elsewhere:

- Medical records unit supervisor ? s. 3252

Notes

Workers who supervise the activities of clerical support workers in law offices and legal departments are classified in Unit group 3342, Legal secretaries. Those who supervise the activities of clerical support workers in health facilities where the work requires specialist knowledge related to health and medicine, such as processing medical records and hospital admission details, are classified in Unit group 3344, Medical secretaries. Those who provide direct secretarial and administrative support to a manager or professional, (except legal and health professionals) and also supervise the activities of clerical support workers are classified in Unit group 3343, Administrative and executive secretaries.

Verwandte Berufskategorien ISCO-08

1221 Sales and marketing managers

122 Sales, marketing and development managers

12 Administrative and commercial managers

1 Managers

Definition:

Sales and marketing managers plan, direct and coordinate the sales and marketing activities of an enterprise or organization, or of enterprises that provide sales and marketing services to other enterprises and organizations.

Tasks include -

(a) planning and organizing special sales and marketing programmes based on sales records and market assessments;
(b) determining price lists, discount and delivery terms, sales promotion budgets, sales methods, special incentives and campaigns;
(c) establishing and directing operational and administrative procedures related to sales and marketing activities;
(d) leading and managing the activities of sales and marketing staff;
(e) planning and directing daily operations;
(f) establishing and managing budgets and controlling expenditure to ensure the efficient use of resources;
(g) overseeing the selection, training and performance of staff;
(h) representing the enterprise or organization at sales and marketing conventions, trade exhibitions and other forums.

Examples of the occupations classified here:

Marketing manager

Sales manager

1222 Advertising and public relations managers

122 Sales, marketing and development managers

12 Administrative and commercial managers

1 Managers

Definition:

Advertising and public relations managers plan, direct and coordinate the advertising, public relations and public information activities of enterprises and organizations or of enterprises that provide related services to other enterprises and organizations.

Tasks include -

(a) planning, directing and coordinating the advertising and public relations activities of an enterprise or organization;
(b) negotiating advertising contracts with clients or with newspapers, radio and television stations, sports and cultural organizations and advertising agencies;
(c) planning and managing information programmes to inform legislators, the mass media and the general public about the plans, accomplishments and points of view of the enterprise or organization;
(d) leading and managing the activities of advertising and public relations staff;
(e) establishing and managing budgets and controlling expenditure and ensuring the efficient use of resources;
(f) establishing and directing operational and administrative procedures;
(g) planning and directing daily operations;
(h) overseeing the selection, training and performance of staff;

Examples of the occupations classified here:

Advertising manager

Public relations manager

2320 Vocational education teachers

232 Vocational education teachers

23 Teaching professionals

2 Professionals

Definition:

Vocational education teachers teach or instruct vocational or occupational subjects in adult and further education institutions and to senior students in secondary schools and colleges.

They prepare students for employment in specific occupations or occupational areas for which university or higher education is not normally required.

Tasks include -

(a) developing curricula and planning course content and methods of instruction;
(b) determining training needs of students or workers and liaising with individuals, industry and other education sectors to ensure provision of relevant education and training programs;
(c) presenting lectures and conducting discussions to increase students' knowledge and competence;
(d) instructing and monitoring students in the use of tools, equipment and materials and the prevention of injury and damage;
(e) observing and evaluating students' work to determine progress, provide feedback, and make suggestions for improvement;
(f) administering oral, written or performance tests to measure progress, evaluate training effectiveness and assess competency;
(g) preparing reports and maintaining records such as student grades, attendance rolls, and training activity details;
(h) supervising independent or group projects, field placements, laboratory work, or other training;
(i) providing individualized instruction and tutorial or remedial instruction;
(j) conducting on-the-job training sessions to teach and demonstrate principles, techniques, procedures, or methods of designated subjects.

Examples of the occupations classified here:

Automotive technology instructor

Cosmetology instructor

Vocational education teacher

Some related occupations classified elsewhere:

- School Principal ? s. 1345
- Secondary education teacher ? s. 2330

Notes

Those who teach vocational subjects that are intended to prepare students for employment in a particular occupational group should be classified in Unit group 2320, Vocational education teachers, whether they work in a general secondary school or in a vocational or technical school or college. Those who teach, at secondary education level, subjects such as mathematics that do not aim to prepare students for employment in a specific occupational area, should be classified in Unit group 2330, Secondary education teachers, even if they are employed in a vocational or technical college.

2431 Advertising and marketing professionals

243 Sales, marketing and public relations professionals

24 Business and administration professionals

2 Professionals

Definition:

Advertising and marketing professionals develop and coordinate advertising strategies and campaigns, determine the market for new goods and services, and identify and develop market opportunities for new and existing goods and services.

Tasks include -

(a) planning, developing and organizing advertising policies and campaigns to support sales objectives;
(b) advising managers and clients on strategies and campaigns to reach target markets, creating consumer awareness and effectively promoting the attributes of goods and services;
(c) writing advertising copy and media scripts, and arranging television and film production and media placement;
(d) collecting and analyzing data regarding consumer patterns and preferences;
(e) interpreting and predicting current and future consumer trends;
(f) researching potential demand and market characteristics for new goods and services;
(g) supporting business growth and development through the preparation and execution of marketing objectives, policies and programs;
(h) commissioning and undertaking market research to identify market opportunities for new and existing goods and services;
(i) advising on all elements of marketing such as product mix, pricing, advertising and sales promotion, selling, and distribution channels.

Examples of the occupations classified here:

Advertising specialist

Market research analyst

Marketing specialist

3252 Medical records and health information technicians

325 Other health associate professionals

32 Health associate professionals

3 Technicians and associate professionals

Definition:

Medical records and health information technicians develop, maintain and implement health records processing, storage and retrieval systems in medical facilities and other health care settings to meet the legal professional, ethical and administrative records-keeping requirements of health services delivery.

Tasks include -

(a) planning, developing, maintaining and operating a variety of health record indexes and storage and retrieval systems to collect, classify, store and analyze information;
(b) transcribing, compiling and processing patient medical records, admission and discharge documents, and other medical reports into records-keeping systems to provide data for patient monitoring and referral, epidemiological monitoring, research, billing, cost control and care improvement;
(c) reviewing records for completeness, accuracy and compliance with regulations;
(d) translating narrative descriptions and numeric information from medical records and other documents on health services delivery into codes associated with standard classification systems;
(e) protecting the security of medical records to ensure that confidentiality is maintained and releasing information to authorized persons and agencies in accordance with regulations;
(f) supervising clerical and administrative workers involved in the maintenance of medical records.

Examples of the occupations classified here:

Clinical coder

Disease registry technician

Health information clerk

Medical records analyst

Medical records clerk

Medical records technician

Some related occupations classified elsewhere:

- Data entry clerk ? s. 4132
- Filing clerk ? s. 4415
- Medical secretary ? s. 3344

Notes

Occupations included in this unit group normally require knowledge of medical terminology, legal aspects of health information, health data standards, and computer- or paper-based data management as obtained through formal education and/or on-the-job training.

3339 Business services agents not elsewhere classified

333 Business services agents

33 Business and administration associate professionals

3 Technicians and associate professionals

Definition:

This unit group covers business services agents not classified elsewhere in Minor group 333, Business services agents. For instance, the group includes those who establish business contacts, sell business services such as advertising space in the media, arrange contracts for performances of athletes, entertainers and artists, for the publication of books, the production of plays, or the recording, performance and sale of music, sell property and goods by auction and who design and organize package and group tours.

In such cases tasks would include -

(a) obtaining information about services to be sold and needs of prospective buyers;
(b) negotiating contracts on behalf of seller or buyer and explaining terms of sale and payment to client;
(c) signing agreements on behalf of seller or buyer and ensuring that contract is honoured;
(d) making sure that the business service purchased is made available to the buyer in the agreed format at the agreed time;
(e) selling by auction various kinds of property, cars, commodities, livestock, art, jewellery and other objects.
(f) organizing group tours for business or vacation travel and making bulk travel and accommodation bookings.

Examples of the occupations classified here:

Auctioneer

Advertising salesperson

Literary agent

Musical performance agent

Sports agent

Theatrical agent

Tour operator

4223 Telephone switchboard operators

422 Client information workers

42 Customer services clerks

4 Clerical support workers

Definition:

Telephone switchboard operators operate telephone communications switchboards and consoles to establish telephone connections, receive caller inquiries and service problem reports, and record and relay messages to staff or clients.

Tasks include -

(a) operating switchboards and consoles to connect, hold, transfer, and disconnect telephone calls;
(b) making connections for outgoing calls;
(c) dealing with telephone inquiries and recording messages;
(d) forwarding messages to staff or clients;
(e) investigating operating system problems and informing repair services.

Examples of the occupations classified here:

Answering service operator

Telephone switchboard-operator

4227 Survey and market research interviewers

422 Client information workers

42 Customer services clerks

4 Clerical support workers

Definition:

Survey and market research interviewers interview people and record their responses to survey and market research questions on a range of topics.

Tasks include -

(a) contacting individuals by telephone or in person and explaining the purpose of the interview;
(b) asking questions following the outlines of questionnaires and surveys;
(c) recording responses on paper or entering responses directly into a computer database through computer-assisted interviewing systems;
(d) identifying and resolving inconsistencies in responses;
(e) providing feedback to survey sponsors concerning problems in obtaining valid data.

Examples of the occupations classified here:

Market research interviewer

Public opinion interviewer

Survey interviewer

4419 Clerical support workers not elsewhere classified

441 Other clerical support workers

44 Other clerical support workers

4 Clerical support workers

Definition:

This unit group covers clerical support workers not classified elsewhere in Major group 4, Clerical support workers. For instance, the group includes, correspondence clerks, press clippers and publication clerks.

In such cases tasks would include -

(a) receiving customers' orders for classified advertising, writing and editing copy, calculating advertising rates and billing customers;
(b) writing business and government correspondence such as replies to requests for information and assistance, damage claims, credit and billing enquiries and service complaints.;
(c) assisting in the preparation of periodicals, advertisements, catalogues, directories and other material for publication;
(d) reading newspapers, magazines, press releases and other publications to locate and file articles of interest to staff and clients.

Examples of the occupations classified here:

Advertising clerk

Correspondence clerk

Directory compiler

Publication clerk

Press clipper

Excursion: Extensional Definition versus Prototypical Definition

(This excursion is not part of any contribution in the thesis.)

As noted in the previous contribution, there exists a key difference between the 2010 GCO and the 2008 ISCO. The traditional purpose of German classifications was to systemize large numbers of job titles under a common label. By contrast, ISCO aims for international comparability, requiring high-quality definitions of occupational categories. This difference is further explored in this excursion. It is linked to two possible ways how categories can be defined. Some kind of definition is necessary because categories from occupational classifications are artificial constructions, nonexistent in reality or outside the classification, that have no meaning without a definition.

- *Extensional definition*: This definition specifies categories using a complete list of all elements in it. The alphabetical directory in the 2010 GCO, consisting of more than 20.000 job titles, provides such a list of job titles linked with categories. In this view, each category consists of nothing but its associated job titles; that is, only the job titles mentioned in the directory belong to a category and nothing else. If one follows this logic strictly, the only way to determine a category is to first select a job title from the alphabetical directory, and, having done so, the appropriate category follows automatically.

- *Prototypical definition* (preferred throughout this thesis): This definition is based on descriptions/definitions of each category. The descriptions highlight its prototypical elements and features, thereby characterizing the content of a category.

Since the 2010 GCO gives more emphasis to job titles than the 2008 ISCO, the subsequent discussion focuses on the 2010 GCO. The German classification contains an alphabetical directory, but also descriptions for each category. This allows using either type of definition, but the 2010 GCO does not write explicitly which type of definition should be used. Depending on the definition used, coders may find different categories most appropriate, leading to disagreement.

Most predecessors of the current German classification appear to have defined categories via extensional definitions, as can be seen from the fact that their categories consisted of lists of job titles grouped together under a common label. Organizing the vast amount of different job titles is still the focus of the current classification: for its development, a clustering algorithm has been used to group job titles into categories of similar occupational expertise. Moreover, for the creation of the alphabetical job title directory, experts coded job titles into the 2010 GCO, respecting the two dimensions occupational expertise ("Berufsfachlichkeit") and requirement level ("Anforderungsniveau"). This shows that its

development was centered around job titles, implicitly suggesting that the alphabetical directory is the most appropriate way to code job titles from respondents. In fact, Paulus and Matthes (2013), who were involved during the development, recommended to code survey answers by first choosing a job title and then translating the job title into a classification code.

However, only relying on the job title directory has some serious drawbacks for coding:

1. Job titles have their own challenges. Some job titles may be misunderstood since they are not vivid descriptions of underlying occupational activities (e.g., "Agrarlaborant"), they can be imprecise (e.g., "teacher" "machine operator") , or may change their meaning if connected with another word (e.g., "Käsebäcker" = "Molkereifachmann" ≠ "Bäcker"). Job titles are invented for communicative purposes but not designed for scientific measurement. Due to such difficulties, several job titles in the 2010 GCO have clarifying words in parenthesis (e.g., "Abdämmer/in (Bergbau)") and predecessors like the 1988 German classification (Bundesanstalt für Arbeit 1988) even included some instructions on the correct formatting and usage of the alphabetical directory.

2. Many respondents do not provide a job title when asked in a survey. Instead, answers may consist of job descriptions, arguably because the recommended German question wording asks about occupational activities, which is in contrast to the title-centered view from German classifications (Geis and Hoffmeyer-Zlotnik 2000). Besides the job descriptions, additional occupational information is collected with closed questions. If coders are provided with a job description and additional information and must select a job title, many different job titles translating to various codes can seem appropriate.

3. The measurement of the requirement level, the second dimension within the 2010 GCO, comes with additional complications. Paulus and Matthes (2013) and Müller (2014) recommend asking for it with a separate question, thus obtaining two values about the requirement level, one from this separate question and another one from the coded job title. Both values can be in conflict to another and arbitrary priority decisions are required.

These issues with job titles are spelled out in more detail in the previous contribution.

Given a set of similar job titles grouped together to form a category (the traditional standard in German classifications), one may wish to describe and define the common characteristics of all elements within this category. Thus, the 1961 German Classification of Occupations (Statistisches Bundesamt 1961) introduced category descriptions, stating conditions that an element should have if it belongs to a category (intensional definition, Sperling 1961). These conditions provide a formalism for classifying new jobs into the classification, as one simply needs to check if the job fulfills all conditions needed to belong to a

category. The formalism also allows to classify respondents' verbal answers even if no matching entries are provided in the job title directory, something that would be impossible if the extensional definition was applied strictly. This implies that category descriptions broaden the content of occupational categories as compared with a pure extensional definition.

Like the 1961 GCO, the 2010 GCO provides descriptions of each category, although the focus is no longer on necessary conditions an occupation must have to be part of a category. The descriptions now have a more illustrative character and exemplify what kinds of jobs belong to a category. For jobs that are too different from examples in any category, the illustrative character can make it difficult to determine the correctness of a category, arguably speaking against defining 2010 GCO categories through their illustrative descriptions. The extensional definition, in contrast, has a clear-cut inclusion criterion, pretending that the same problem is nonexistent here. Rosch (1978), however, argued against such in-or-out definitions to determine category membership. According to her, "categories do not have clear-cut boundaries" (p. 35) that could be meaningfully defined in real-world settings, but people can judge how typical an element is. Because "prototypicality [i.e. the degree of centrality in categories] is reliably rated and correlated with category structure" (p. 38), prototype theory provides a theoretical basis to define occupational categories not by in-or-out inclusion criteria that focus on category boundaries, but through a description of their prototypical features that focus on category centers. Since the category descriptions from the 2010 GCO illustrate the characteristic content of each category, they are well-suited as prototypical definitions.

To sum up, at least two types of definitions are possible. The extensional definition complies better with the reasoning that was used to develop the 2010 GCO. Coding with the alphabetical job title directory (i.e., algorithm 1 in the third contribution) is in line with this type of definition, suggesting that it should be the preferred way. However, coding with the alphabetical directory comes with its own difficulties. For an alternative way of coding, one may rely on the prototypical definition[1].

The third contribution provides evidence that coders do not always use the extensional definition, even if they could have found a job title in the alphabetical directory (see Table 3 on page 218). Similarly, researchers working with occupational data are more likely to look at the category descriptions and not at the alphabetical directory to learn about the content of a category. Both observations suggest that coding with the alphabetical directory according to the extensional definition is not considered a gold standard in scientific research. Throughout this thesis the prototypical view is preferred.

(The literature cited in this excursion is included in the main bibliography (see page 20)).

[1]The category descriptions at www.klassifikationsserver.de may be useful for this.

Machine Learning for Occupation Coding - A Comparison Study

Author: Malte Schierholz

June 18, 2019

Abstract

Occupation coding—the assignment of verbal responses to an occupation question into an official classification—can be time-consuming and expensive if done manually. Its automation has been a longstanding goal and numerous researchers have developed different algorithms for this purpose. After reviewing the solutions proposed so far, we select six highly promising algorithms for our comparison. The algorithms are tested with data from six German surveys. A tree boosting algorithm (`XGBoost`), not used in the occupation coding literature before, proves to be comparable or better than other algorithms if sufficient training observations are available. To improve results further, we develop a novel algorithm that enriches the training data with additional information from a coding index. Different practical applications are distinguished and evaluated separately. An R-package `occupationCoding` implementing the various algorithms is available on-line.

1 Introduction

Occupation coding—the assignment of verbal responses to an occupation question into an official classification—has many faces. For example, people make implicit use of occupational coding algorithms when comparing their income online to others in similar jobs (Tijdens et al., 2010). In job recruitment, the matching between job seekers and vacant positions is facilitated if both are coded to the same occupational group (Bekkerman and Gavish, 2011, Javed et al., 2015). Prices for car insurances can differ depending on a customer's occupational group (Scism, 2016). Countless scientific studies

exist, mainly from sociology, economics, and epidemiology, which use occupation as part of their statistical analyses. As the examples show, capturing individuals' occupations is relevant in many fields.

The various applications are connected by a data collection process, in which individuals need to describe their occupation, often using their own words, which are entered in a text field. For digital processing and analysis it is then almost inevitable to categorize (code) the responses into systematic units. If thousands or millions of answers need to be coded, the costs to employ human coders are substantial.

To save costs, high efficiency is essential. Machine learning (Bishop, 2006, Hastie et al., 2009) promises to increase the degree of automation, reducing costs. This academic field has mastered the challenge to detect some kind of signal from unstructured data (e.g., detecting words in audio recordings, recognizing handwritten digits from images, ...), proving its high accuracy in academic competitions (e.g., Schmidhuber, 2015). Algorithms from machine learning have been applied on textual data, often with the intention to classify large amounts of text while employing small resources for coding (Sebastiani, 2002, Grimmer and Stewart, 2013, Gentzkow et al., 2017). On its face, occupation coding is similar: One needs to detect the signals (i.e., occupations) and assign appropriate labels for a large number of textual responses. Indeed, various algorithms from machine learning and beyond have been examined to make occupation coding more efficient (see below).

Yet, occupation coding is different from other tasks in signal detection and text classification for a number of reasons.

1. The classification problem is high-dimensional with several hundreds of outcome categories and 10,000s of words used as predictors. On the one hand this poses challenges to the speed and efficiency of computation. On the other hand the ideal training data would need to contain every possible input text more than once, requiring millions of observations and more, a number rarely collected in typical surveys.

2. The textual input is often very short and can include misspellings, making it difficult to automatically extract the signal from the textual response. Automatic spelling correction is challenging because job titles are a rather particular vocabulary.

3. Machine learning is often used to predict rather manifest concepts of which the 'true' value is rarely contested. In contrast, human cod-

ing decisions are more debatable and low agreement among coders is a concern if the concept of interest is theoretical in nature or socially constructed (e.g., populism, rationality, inequality). Social scientists analyzing such topics usually rely on human coding as their gold standard, possibly with computer-assisted techniques having a complementary role (Nelson et al., 2018). The social science perspective may be more appropriate for 'occupation', due to the importance that human imagination and interpretation have at the coding stage.

In our subsequent analysis we react to the first point by pooling data from multiple surveys and from a coding index, alleviating the concerns about small training data. The second point, short and possibly misspelled text, is countered by using customized algorithms for occupation coding. The third point shifts the attention to computer-assisted coding; this paper is written with an eye on how machine learning can be used for such purposes.

Our area of application is the measurement of occupation in scientific surveys and statistical data collections. Statistical agencies have a long-standing expertise in this field—the first German *Occupation and Business Census* was conducted in 1882 (Rauchberg, 1888)—, and agencies around the world now maintain and update occupational classifications of their own. One well-known classification is the *International Standard Classification of Occupation* (ISCO) (International Labour Office, 2012), a hierarchical classification with 436 categories at its most detailed level. Equally important in Germany is the *2010 German Classification of Occupation* (GCO) (developed by the Federal Employment Agency (2011) in cooperation with the Federal Statistical Office), which we use in our subsequent analysis. The GCO was created to depict the special structure of the German labor market, aiming for high compatibility with the 2008 ISCO as a secondary goal (see the extensive GCO documentation and Schierholz, Brenner, Cohausz, Damminger, Fast, Hörig, Huber, Ludwig, Petry and Tschischka (2018) for a comparison with ISCO). At its most detailed level, the GCO has 1286 occupational categories, hierarchically systematized by two dimensions, 'occupational expertise' and 'requirement level', both being defined in the GCO documentation. Besides this general systematology, specific categories exist concerning the military, helpers, supervisors, managers, jobs "without specialization", and jobs "with specialization, [but] not elsewhere classified". A dictionary is part of the GCO, listing nearly 28,000 job titles and their associated categories. One example is the job title "Media-Designer/in" →

23224, a German loan word in its male and female spelling. It is mapped to category 23224, titled "Occupations in graphic, communication, and photo design—highly complex tasks", of which the GCO describes typical tasks and duties in detail. The reasoning behind this mapping is that the 'occupational expertise' of a "Media-Designer/in" is similar to other occupations in graphic, communication, and photo design and the 'requirement level' is such that the work often consists of highly complex tasks, usually requiring a university degree. This cursory inspection demonstrates that occupational classifications are based on a theoretical conception, shaping the content of each category. Importantly, the categories are artificial units that have no counterpart in the real world. Meaning and content of each category are defined by the respective official classification and depend on human interpretation.

Occupational classifications provide few recommendations about best practices for coding. In principle, human coders need to read and interpret the verbal job descriptions from respondents in order to find the most appropriate category for each. There is a strong feeling that identical answers should be coded to identical categories, implying that coders should follow previous decisions from their coworkers. However, if coders do not know the described job, if occupational categories are poorly defined in the official classification, or if the verbal answer matches poorly with categories from the classification, coders may disagree about the best-fitting category (Conrad et al., 2016). In fact, several studies show that the agreement among coders is often rather low (Elias, 1997, Bound et al., 2001, Mannetje and Kromhout, 2003); in our data it is between 50% and 80.2%, depending on the subset of data used and other factors (see on-line appendix). To resolve such difficulties and increase inter-coder reliability, coders often follow a set of rules and conventions (Geis and Hoffmeyer-Zlotnik, 2000). For example, if someone performs a wide range of tasks connected with different categories, the 2008 ISCO specifies that tasks related to the production of goods take precedence over tasks related to the distribution of goods. Additional information, often used in the German context, may also help. Paulus and Matthes (2013), for example, mention their rule of thumb that self-employed craftsmen are coded as managers if they have at least ten employees, otherwise they are coded as supervisors. Despite all these efforts, some cases still require subjective interpretations and human judgment.

The tools and technologies that coders use in their daily work are another important factor in the coding process. Historically, coders often consulted

alphabetic dictionaries to find, for any job title in it, the desired code. Nowadays, coding is usually done with computers using a number of different software products. A general distinction can be made between *computer-assisted coding* and fully *automated coding* (Riviere, 1997, Speizer and Buckley, 1998). In computer-assisted coding, coders enter the text into a search mask and select a code from a list of suggestions. It has been hypothesized that using a computer-assisted coding software could change coding decisions, either introducing a systematic bias or leading to improved inter-coder reliability, but effects were small (Campanelli et al., 1997, Bushnell, 1998). Automated coding refers to situations in which the software selects the final code without human intervention, reducing the workload for human coders at the risk of introducing coding errors. In both modes of operation it is common practice to calculate a score, which allows ordering the different categories by their relevance. For automated coding, the computer would simply choose the highest-scored category.

This paper reviews and compares various techniques to calculate such scores, focusing on algorithms from machine learning. The algorithms under comparison are carefully selected and several less-promising ones have been discarded. A novel algorithm is introduced that may stimulate future research. For the evaluation we take an applied perspective, asking how many answers could be coded automatically and how often the assigned categories would be identical with manual coding. Using five data sets we have at our disposal, we showcase how much the results depend on the respective choice of training and test data. This allows us to identify the most competitive algorithms. The algorithms can be applied in different environments (1. computer-assisted coding after data collection, 2. computer-assisted coding during data collection, 3. automated coding) and we describe for each environment the prerequisites an algorithm should meet to be useful.

This research is part of our larger research agenda, which focuses on computer-assisted coding at the time of a survey interview (Schierholz, Gensicke, Tschersich and Kreuter, 2018). The idea is to present possible job categories directly to respondents, who in turn can choose the most appropriate occupation, increasing data quality and reducing the workload for coders. Our research design and analysis is influenced by this application. Other researchers possibly have their own application in mind or wish to replicate our analysis with their own data. For such purposes we provide a software package written in R (R Core Team, 2016) at `https://github.com/malsch/occupationCoding`, which implements the various

algorithms, including the ones newly developed.

Section 2 ('Algorithms') reviews the major approaches that have been used to automate occupation coding and describes the rationale behind our selection. Section 3 ('Research Questions') motivates why we analyze the data the way we do. Section 4 ('Data') gives key information about the data sets originating from six German surveys, their differences, and how we harmonized them. Section 5 ('Analytical Strategy') describes the basic preprocessing steps to make use of textual data, the mapping between research questions and data sets, and our approach to model tuning. Section 5 ('Results') summarizes the results and Section 6 concludes. An extensive appendix is available on-line.

2 Algorithms

We tested ten different algorithms for our comparison. The first two algorithms use a coding index and no training data. By contrast, algorithms three to six use training data only. The final four algorithms are variants of a strategy we developed to combine both approaches, exploiting the respective strengths of using a coding index and training data. Table 1 summarizes the key features of the algorithms.

This section serves three purposes. It reviews the literature, highlights the key ideas of the different algorithms, and explains the reasoning behind our selection.

2.1 Algorithms without training data

2.1.1 Coding Index (Exact Matching)

Provided that a coding index containing entries like "Media-Designer" $\rightarrow$ 23224 exists, its application is straightforward. If the textual input is exactly identical with some entry from a coding index, the corresponding code is assigned. Lyberg and Andersson (1983) provided an early account of experiences when statistical agencies were just starting to use this algorithm. Bekkerman and Gavish (2011) argued that the development of a coding index was a highly efficient way to classify millions of users at LinkedIn.

The coding index used throughout this paper is, in principle, the alphabetic dictionary that is part of the GCO. Yet, most entries in the alphabetic

Table 1: Overview of algorithms

	Algorithm	Reference data	Inferential technique	Output
1.	Coding index (exact)	Coding Index	String identity	Code(s) (zero or more)
2.	Coding index (\w similars) (CASCOT)	Coding Index	String similarity	Code(s) (zero or more) \w confidence scores
3.	Memory-based Reasoning Creecy et al. (1992)	Previously coded	Bag-of-Words & Vector similarity	Code (zero or one) \w confidence score
4.	Adapted Nearest Neighbor Gweon et al. (2017)	Previously coded	Bag-of-Words & Vector similarity	Code(s) (zero or more) \w confidence scores $\in [0, 1)$
5.	Multinomial Regression (`glmnet`)	Previously coded	Bag-of-Words & Loss minimization	Codes (as in training data) \w predicted probabilities
6.	Tree Boosting (`XGBoost`)	Previously coded	Bag-of-Words & Loss minimization	All codes from classification \w predicted probabilities
7.	Similarity-based Reasoning (fulltext similarity)	Coding Index & Previously coded	String similarity & Bayes theorem	All codes from classification \w predicted probabilities
8.	Similarity-based Reasoning (substring similarity)	Coding Index & Previously coded	String similarity & Bayes theorem	All codes from classification \w predicted probabilities
9.	Similarity-based Reasoning (wordwise similarity)	Coding Index & Previously coded	String similarity & Bayes theorem	All codes from classification \w predicted probabilities
10.	Maximum Probability	Coding Index & Previously coded	Ensemble (Algorithms 6-9)	All codes from classification \w predicted probabilities

dictionary have a gender-neutral formatting (e.g., "Media-Designer/in"), unsuited for exact matching with verbal answers. We use a related dictionary, continuously updated by the Federal Employment Agency (2019) to include new jobs, because this file separates male and female job titles. To simplify exact matching further, we remove all text written in parentheses (e.g., "Receptionistin (Hotel)" → "Receptionistin"). The resulting coding index contains a male and a female entry for each job, totaling in more than 50,000 entries.

Despite this size, not every possible input text can be included in the coding index and exact matching will often fail.

2.1.2 Coding Index (Similarity Matching)

Even if the textual input and an entry from the coding index are not exactly identical, they may still be similar enough to be considered a match. This general approach—using a coding index and calculating similarities—is widely used and any comparison study would be incomplete without it. Yet, similarity calculations can differ in an infinite number of ways. Best practices do not exist and many software developers create their own variant of this approach. This means our result can only be indicative of what can be achieved with it and developers might wish to make their own comparisons.

Some examples demonstrate how coding indexes and similarity matching have been used. Ossiander and Milham (2006) preprocess the textual input (i.e., correct spelling errors, remove punctuation) before they search the coding index for similar entries. Their similarity score is the number of words that appear in both the textual input and in job titles from the coding index. The job title with the highest score is used for automated coding. Likewise, Russ et al. (2014) calculate a score for automated coding, but they divide the above score by the total number of different words from both texts (known as the Jaccard index), standardizing the score between zero and one. To be robust against misspellings, Damerau-Levenshtein distances between individual words are also taken into account. Yet another example is from Munz et al. (2016), who calculate the number of characters that would need to change before the textual input and the index entry are identical (generalized Levenshtein distance). As this is implemented in a computer-assisted coding application, coders see the most similar job titles to choose from.

Statistical agencies were early adopters of related approaches, which can be quite sophisticated. Algorithms developed at the US Census Bureau and

at Statistics Canada had routines to standardize the textual input (e.g., "Private Family Babysitter" → "PRIV FAMIL BABYSIT", an example taken from Appel and Hellerman (1983)) to eliminate and replace misspellings, abbreviations, trivial words, and other undesired characters, only keeping those terms that are useful for coding (Wenzowski, 1988). Rather complex scoring algorithms were developed to weigh the relative importance of various terms against each other and to calculate an overall score for each category (Knaus, 1987). Speizer and Buckley (1998) review this American line of research in more detail. Riviere (1997) collected descriptions about developments in a number of European statistical agencies, which used various similarity measures. These were often not based on splitting the text into words, but instead the text was split into short sequences of two or three characters.

One prominent and widely used computer program that is based on similarity calculations is the 'Computer-assisted structural coding tool' (CASCOT). Index editors can fine-tune the tool to achieve better performance by downgrading less important words, defining equivalent word endings, defining rules to replace abbreviations and other terms, among many other available options. This makes the tool very flexible, although the authors write that fine-tuning is a "resource-demanding, time-consuming" task. The tool has been developed to be used with different languages for a variety of classifications (Elias et al., 2014).

For our evaluation, we load the GCO and the same coding index as before into CASCOT, but make no further adjustments. Results obtained this way are certainly suboptimal and one might try an infinite number of ways to improve them. We intentionally keep it simple to see how out-of-the-box solutions for occupation coding perform. CASCOT has a mode for automated coding, which outputs the highest similarity score along with the predicted category. To evaluate computer-assisted coding with CASCOT, we captured the data from the computer screen with optical character recognition. The analysis will require that we rank different textual inputs by their likelihood that one of the top five categories will be selected. CASCOT outputs for every suggested category a score between 1 and 100. Summing these scores as we do for other algorithms would be inadequate here, because a single suggested category with a score of 100 (aggregate score $= 1 \cdot 100$ and very likely to be selected) would be lower-ranked than 5 suggested categories each having a score of 30 (aggregate score $= 5 \cdot 30$, but unlikely to be selected). Instead of a sum, we use the mean among the top five categories.

2.2 Algorithms based on Training Data

The algorithms above were designed to utilize coding indexes, emulating their usage by professional coders. If no such coding index exists, one may try to use previously coded answers to automatically build a coding index, an approach that reappears in work on automated occupation coding since its early days (O'Reagon, 1972, Thompson et al., 2014). However, one may also use previously coded answers directly in place of a carefully edited coding index. Using answers as they naturally occur might be beneficial, because it is impossible to consider all conceivable textual answers in an edited coding index. We now look at algorithms that were designed to learn prediction rules from previously coded answers.

A common type of text processing is applied for all algorithms that follow in this section. The idea is to reduce the dimensionality of text and to bring it in a numeric format, making it more suitable for analysis. Common steps include lower-casing all letters, removing punctuation marks, or substituting special characters (e.g., '€' → 'euro'). Stemming (e.g., 'designer' → 'design') and the removal of stop words (e.g., 'and', 'the') are also often used to make similar answers identical, reducing the dimensionality, at the risk that decisive information for coding is lost. A document-term matrix consists of one row per document and one column per word. It is created from the standardized texts by counting how often each word/term appears in each document. This disregards the word order. As a resort, one may add additional features to the document-term matrix, for example counting the frequency that all two-word sequences appear in each document. Additional information may be included in the predictor matrix as well, such as the number of words per document or other variables from a survey. The exact choices how predictor matrices are calculated vary widely as different authors try what works best for them, depending on the data and the algorithms they wish to use.

2.2.1 Memory-based Reasoning

Creecy et al. (1992) proposed a k-nearest neighbor algorithm, which aims to code answers in the same way as previous similar answers were coded. Their document-term matrix does not only include columns for every word in the training data, but also for all two-word co-occurances in the training data. Two metrics to calculate similarity between different answers were proposed. Both metrics take into account that words like "the" provide

little insight about an appropriate occupational category, whereas words like "weaver" are more meaningful and should have higher weights. To predict a category, one searches the training data for k nearest neighbors having highest similarity. Among these, the similarity scores are added per category and the category with the highest summed score is selected. If one wishes to use this algorithm to separate easy-to-code from hard-to-code answers, the authors recommended calculating confidence scores and defining thresholds category-wise. However, since Creecy et al. (1992) determined thresholds using the same test data that was used again to evaluate the performance, they risked that their performance metrics are biased upwards. We do not implement this problematic procedure in our software, but related approaches still might help to improve the performance at low production rates.

This algorithm contrasts with previous research at the US Census Bureau, which was based on coding indices and similarity matching. Creecy et al. (1992) argued that their new algorithm outperformed the previous system by wide margins. Moreover, they developed the new system in just "four person-months while the [previous] expert system required 192 person-months." With impressive results like this, confirmed by Gillman and Appel (1994) in a comparison with other supervised learning algorithms, it is worth including the Memory-based Reasoning algorithm in our comparison.

2.2.2 Adapted Nearest Neighbor

Gweon et al. (2017) proposed an adapted nearest neighbor algorithm for occupation coding. To classify a new response x, one searches in the training data for all responses v that have largest similarity. The cosine similarity

$$s(x, v) = \frac{\sum x_j v_j}{\sqrt{\sum x_j^2}\sqrt{\sum v_j^2}}$$

between vectors from the document-term matrix is used (the index j runs across columns/words in the document-term matrix). It ranges between 0 if both answers have no words in common and 1 if both texts (i.e., rows from the document-term matrix) are identical. Let $K(x)$ be the number of responses from the training data that are most similar and let $\hat{p}_{nn}(l|x)$ be the relative frequency of category l among those most similar responses. The category which appears most often ($\arg\max_l \hat{p}_{nn}(l|x)$) is predicted along with a score to express the certainty of correctness. This score is calculated as a

product of the relative frequency, the similarity, and another multiplier,

$$\gamma(l|x) = \hat{p}_{nn}(l|x)s(x)\frac{K(x)}{K(x)+0.1}$$

The motivation is that relative frequencies alone are poorly suited as an expression of certainty. Therefore $s(x)$ shrinks the relative frequency towards 0 depending on how similar the most similar training cases are. Likewise, $\frac{K(x)}{K(x)+0.1}$ shrinks the relative frequencies towards zero if the prediction is based on very few (e.g., $K(x) = 1$) training observations.

Gweon et al. (2017) developed the adapted Nearest Neighbor algorithm specifically for occupation coding and compared it with a number of other algorithms. It outperformed support vector machines with linear kernel, a duplicate algorithm, a hybrid combination of support vector machines with the duplicate algorithm, and other combinations of algorithms that exploit the hierarchical structure of occupational classifications. This superior performance leads us to include the adapted nearest neighbor algorithm in our comparison.

2.2.3 Multinomial Logistic Regression

We now turn to algorithms that are popular not only for occupation coding but for a wide range of supervised learning applications. The first approach is *multinomial logistic regression with elastic net regularization* as implemented in the R-package `glmnet` (Friedman et al., 2010, Hastie et al., 2015). Let $Y = (0, ..., 0, 1, 0, ..., 0)$ be a coordinate vector of length K with a 1 at the lth position indicating that the lth category has been selected. Given a p-dimensional feature vector x, the probability for category l, $l = 1, ..., K$, is modeled as $\mathbb{P}(Y_l = 1|x) = \frac{\exp f_l^{\text{reg}}(x)}{\sum_{k=1}^{K} \exp f_k^{\text{reg}}(x)}$ with $f_l^{\text{reg}}(x) = \beta_{0l} + x^T\beta_l$. The parameter matrix $\beta \in \mathbb{R}^{K\times(p+1)}$ contains several million parameters in our application. To deal with the high dimensionality we employ elastic net regularization, $P_\alpha(\beta_l) = \sum_{j=1}^{p}(1-\alpha)\frac{1}{2}\beta_{jl}^2 + \alpha|\beta_{jl}|$. Since α is typically chosen in the model tuning phase, the elastic net is more flexible than ridge regularization ($\alpha = 0$) and lasso regularization ($\alpha = 1$). Parameters are estimated using a pathwise coordinate descent algorithm that maximizes the regularized multinomial log-likelihood

$$\max_{\beta} \frac{1}{N}\sum_{i=1}^{N}(\sum_{l=1}^{K} y_{il} \cdot f_l^{\text{reg}}(x_i) - \log(\sum_{l=1}^{K} \exp f_l^{\text{reg}}(x_i))) - \lambda\sum_{l=1}^{K} P_\alpha(\beta_l)$$

In practice, `glmnet` does not support very rare categories. We were forced to remove between 61 and 178 observations, depending on the training data used, whose categories appear less than two times in the training data. Rare categories will never be predicted. We set their predicted probabilities to zero, which is needed for some of our subsequent evaluations.

Various authors have used logistic regression (Measure, 2014), support vector machines with linear kernels (Takahashi et al., 2005, 2014, Westermark et al., 2015, Gweon et al., 2017), or maximum entropy classifiers (Jung et al., 2008, Russ et al., 2016) for automated occupation coding. These algorithms share the common feature that they search the function space of linear functions (e.g., $f_l^{\text{reg}}(x)$) to maximize some objective. Theoretical results summarized in Hastie et al. (2009) prove a high similarity between regularized logistic regression and support vector machines with linear kernels. This makes us believe that the different algorithms will achive similar performance, confirmed by an empirical result from Measure (2014). We therefore include only one of these algorithms in our comparison and prefer regularized multinomial logistic regression because, unlike support vector machines, it returns predictions on a probability scale and it avoids estimating a large number of binary classifiers.

2.2.4 Gradient Tree Boosting

Gradient Tree Boosting (Friedman, 2001) is another well-known supervised learning technique. We use the speed-optimized implementation from the R-package `XGBoost`, which enhances Friedman's proposal with some additional tweaks (Chen and Guestrin, 2016). Like in logistic regression, the goal is to obtain functions $f_l(x)$ that minimize the negative multinomial log-likelihood; only the regularization term is different. However, logistic regression is rather restrictive and inflexible as it forces the functions $f_l^{\text{reg}}(x) = \beta_{0l} + x^T\beta_l$ to be linear in their parameters. In contrast, gradient boosting can learn more flexible functions $f_l^{\text{boost}}(x) = \sum_{t=1}^{T} f_l^{(t)}(x)$. Thus, one learns an ensemble of base learners $f_l^{(t)}$ and the flexibility of the algorithm is determined by the class of functions that $f_l^{(t)}$ comes from. As is common, we choose decision trees as base learners. Decision trees learn step functions in a high-dimensional input space and allow for high-order interactions. The base learners are trained iteratively in T rounds, with each iteration focusing on examples that the previous iterations got wrong. The higher flexibility comes at a cost. There exist a dozen tuning parameters in `XGBoost`, some of them relating

to the overall gradient boosting algorithm, others relating to the algorithm that grows the individual trees, making a careful tuning phase mandatory. This can be time-consuming. Developers should have a sufficient number of CPUs in their computing environment and a tuning strategy to achieve good-enough results within a limited number of iterations.

We include `XGBoost` in our comparison because the algorithm has an impressive track-record in winning machine learning competitions, showing that it is state-of-the-art when excellent predictions are desired. Moreover, tree-based approaches appear very promising for occupation coding, because trees are very fast in detecting relevant features and interactions and because split point selection is obvious for binary features (i.e., the (non-)occurrence of words in our situation). Unlike regression, `XGBoost` outputs positive predicted probabilities for all categories, even for those categories not observed in the training data.

2.2.5 Other Algorithms

We believe the algorithms mentioned above are most promising for occupation coding. There exist, of course, additional studies that have explored other directions. For example, Ikudo et al. (2018) trained a Random Forest and Nahoomi (2018) tried whether Convolutional Neural Networks could outperform classic approaches, albeit both studies do not use survey data. Besides, the large number of categories in occupational classifications led Javed et al. (2015), Gweon et al. (2017), and Nahoomi (2018) to consider hierarchical approaches, which exploit that classifications aggregate the bottom-level categories into a smaller number of medium- or top-level categories. The empirical evidence from this literature gives no reason to expect large gains in performance by using any of these algorithms. Therefore, they are not part of our comparison.

Several authors have combined various algorithms in a number of ways. Jung et al. (2008) and Westermark et al. (2015) followed a sequential approach, in which matching with a coding index is tried first and only if this is unsuccessful a support vector machine (or a maximum entropy model) is used for prediction. Likewise, Takahashi et al. (2005, 2014) used exact matching first (in fact, their matching process is more complicated due to Japanese grammar), but its agreement with human coders was unsatisfactory. Thus, the resulting category is not coded directly, but it is added as a feature to a document-term matrix and used in a support vector machine.

This strategy relates to 'stacking'-based ensemble classifiers, implemented for occupation coding by Russ et al. (2016) and Schierholz, Gensicke, Tschersich and Kreuter (2018). The idea of 'stacking' is to predict categories using various classification algorithms and to use the predictions as features in a final adjudication algorithm. The verbal answers enter the adjudication algorithm not directly, but only through the predictions from the other algorithms. Thompson et al. (2014) described another related algorithm. They automatically built a coding index from the data, a first-level classifier, but its predictions were adjudicated with a logistic regression model that takes several other variables into account. Taken together, several independent research projects have combined different algorithms in a number of ways but there is no consensus about best practices. This paper focuses on isolated algorithms, which can be used as a stand-alone module or combined with one another to form an integrated system.

2.3 Similarity-based Reasoning: Combining the Coding Index and Training Data

Algorithms that make predictions from training data often fail if the textual input is a rare job title (e.g., ornithologist, farrier) or if it includes misspellings. This happens if no identical words are found in the training data and, thus, these words cannot be used for prediction. In contrast, algorithms that rely on the coding index have solved this issue. Rare job titles can be found in the coding index and string similarity calculations cope with misspellings; both are desirable features we wish to exploit.

Yet, coding-index-based algorithms are usually developed to find one or more possible categories; there is no intention to predict probabilities. We now describe an algorithm we developed to combine the strengths from using both a coding index and training data, aiming to predict probabilities along with each category.

At its core, the algorithm computes string similarities (e.g., van der Loo, 2014), comparing verbal inputs with entries from a coding index. Three variants of the algorithm can be distinguished, depending on different definitions of what is "similar".

- *Fulltext Similarity*: An entry from the coding index is similar if the verbal answer differs by at most one character.

- *Substring Similarity*: An entry from the coding index is similar if it is a substring of the verbal answer.

- *Wordwise Similarity*: An entry from the coding index is similar if there exists a word in the verbal answer that differs by at most one character.

Given a verbal answer, we can use any definition of similarity to find a set of similar entries in the coding index. A range of different similarity calculations were tested and those were the most promising ones.

We also added a few hundred entries to the coding index (e.g., 'student assistant', 'sales', 'worker') that appear several times in our data but were not yet included in the index described in section 2.1.1. Since they can refer to different categories, the new entries are left without a category assignment. Importantly, we do not require that the coding index makes correct category assignments—unlike current systems that rely on a coding index. Instead, training data is used to learn for each index entry about the possible category assignment.

In the training phase we count for every entry in the coding index how many verbal answers from the training data are similar and how often they were assigned into the different categories. Generally speaking, to make a prediction for a new verbal answer, we first search the coding index for a similar entry and predict the category which was most often associated with it in the training data.

Three complexities arise because we do not want to predict a single category, but we want to estimate a posterior predictive distribution, predicting the probability of each category. Firstly, if the training data contains only a single observation similar to an index entry, this index entry is coded with 100% relative frequency into a single category. Yet, it would be wrong to expect that the same category will always get selected with similar future observations. The issue is addressed by combining the observed data with prior beliefs, in effect down-weighting high relative frequencies towards more reasonable probabilities. Secondly, to make use of category assignments from the coding index we need to find an appropriate balance between the evidence from the coding index and the evidence provided by the training data. Finally, a verbal answer can be similar to more than one entry from the coding index. To still output only a single predictive distribution, a weighted average over index entries is used, giving less weight to ambiguous index entries that were coded into many different categories. A hierarchical Bayesian

model was used to accomplish the three objectives. Its mathematics are described in the on-line appendix.

The idea of using a Bayesian approach for down-weighting is taken from our previous research (Schierholz, 2014, Bethmann et al., 2014). Our current work is an improvement since the choice of a prior used for down-weighting is now derived theoretically using a hierarchical Bayesian framework. Also, this proposal is not bound to any particular definition of similarity, making it more flexible than our previous work, in which we considered identical matching only.

2.4 Maximum Probability Algorithm

With three different ways to calculate similarities, we have three algorithms that output probabilities for every category. Tree boosting is yet another algorithm that outputs probabilities. Since each of the four algorithms has its own strengths and weaknesses, we looked for a way to combine these four algorithms. The result is the *Maximum Probability Algorithm*. For every input text to be coded, it runs the three similarity-based algorithms and tree boosting in parallel. To actually make a prediction, it selects the algorithm which outputs the highest probability with its highest-ranked category and ignores the other algorithms. The rationale is that with different ways to calculate probabilities, we would like to use the algorithm which is most certain to find the "correct" category. We admit that this is an ad-hoc approach.

3 Research Questions

Each algorithm has its own procedure to predict occupational categories for future data. To estimate the expected performance of each procedure, the predictions are compared with coded occupations from various test data sets.

We anticipate that some individuals will provide verbal answers that are difficult to code, e.g., unforeseen new jobs and spellings, and any automatic procedure will miscode them with near certainty. If high data quality is desired, a human expert will need to code the most difficult cases, i.e., the ones that the computer gets probably wrong.

To acknowledge this, we look at two key metrics that are easy to interpret for practitioners. The *production rate* is the proportion of responses for

which the automatic prediction procedure will be used to create some output. Since more automation reduces the amount of work for human coders, higher production rates can save costs for coding. The *agreement rate among top 1* is conditional on the production rate. Among the subset of responses coded automatically, it is the proportion of answers where the top-ranked category from the algorithm and the "true" category from the evaluation data set are identical. Since "truth" is difficult to establish in occupation coding, a term like "accuracy" would be a misnomer. Instead, our term highlights whether automatic procedures are in agreement with current coding procedures. Ideally, we would like an agreement rate of 100%, implying that the automated coding procedure will always output the same categories as hitherto. Yet, the low rates of agreement between human coders raise doubt whether human coding decisions are deterministic, a prerequisite to achieve perfect rates of agreement.

There exists an inherent trade-off between the production rate and the agreement rate. If high production rates are desired, a larger number of hard-to-code answers needs to be coded automatically, which are more likely to disagree. With increasing production rates we expect decreasing agreement rates.

In practical applications, users are advised to consider all possible combinations of production rate and agreement rate among top 1. If they feel that cost reductions are possible while maintaining a sufficiently high agreement rate among top 1, it makes sense to integrate an automated coding procedure into their coding process.

Our first two research questions describe the performance of automated coding procedures in absolute and relative terms.

Question 1: For each algorithm, which performance (production rate, agreement rate among top 1) can be achieved?

Question 2: Which algorithm performs best?

Most of the algorithms described above need training data. Yet, creating training data requires expensive manual coding that we wish to avoid. A possible resort is to use coded data from previous studies.

A standard assumption in supervised learning is that training data and the future data to be predicted are drawn from the same distribution. This assumption is not necessarily true for occupation coding. Each study has its own data collection and coding processes, possibly leading to systematic dif-

ferences between studies. One study may collect more detailed occupational information than another study. If information is lacking for a well-founded decision, coders often follow some default coding conventions of their institute instead. Predictions based on training data will perpetuate the patterns and coding conventions found in the training data. Yet, if a study wants to code occupations that are described in detail, it is problematic to just replicate coding decisions from previous studies for which less information was available. Only if studies aim for high consistency with previous coding decisions, this gives a reason to ignore the above argument about sub-optimal data quality. Trust in the quality of the training data is thus the decisive criterion whether to use training data or not. If one believes the data quality is high and consistency is desired, replicating the same coding decisions with automated algorithms seems like a good idea. If one believes the quality in the training data is low and coders should use additional information not used there, the reliance on training data is less promising.

To judge whether the training data is useful for the desired application, we need to make predictions and evaluate their performance. Importantly, this should be done with test data that are representative of the future application. We will use a number of different training data sets, but compare their predictions on a single test data set. We hypothesize that the data generating process is quite different in this test data, making it difficult to achieve excellent coding decisions with automated procedures.

Question 3: By how much does the agreement rate degrade if training data and test data have been created in different ways?

It is well recognized that supervised learning usually improves with larger training data. Especially for high-dimensional and local prediction problems like occupation coding we expect large gains because the ideal training data should contain several codings for every possible input text. Yet, there exist few studies that have coded more than 50,000 occupations in Germany. Increasing the size of the training data is only possible if we pool data from various studies. This leads once again to concerns that studies differ, contradicting the assumption that all observations are drawn from the same distribution. If observations are added to the training data which differ from coding decisions made in the test data, the performance in the test data may deteriorate. Whether and by how much the performance can improve with larger training data is thus an empirical question.

Question 4: Can we improve predictions by pooling different data sets to form larger training data?

As described, the agreement rate among top 1 is used to judge the data quality from fully automated coding. If it is low, automated coding may be prohibitive. Quality is less of a concern in computer-assisted coding, because in this mode a human coder would select a category from a list of suggestions. However, poor suggestions are useless for a human coder, who would waste time scanning the suggestions without finding an appropriate category. We anticipate that our system for computer-assisted coding (interview coding) will suggest the five highest-ranked categories to a human coder. This makes the *agreement rate among top 5* a relevant criterion for computer-assisted coding, which measures how often coders will find the "true" category within the list of suggestions.

By definition, the agreement rate among top 5 will always be at least as large as the agreement rate among top 1. If an answer is ambiguous and can be coded into multiple categories, the agreement rate among top 5 is more forgiving as it finds the "true" category even if it is not the top suggestion. If the goal is to find all possible categories for a given answer, pooling various data sets may prove useful because more data sets will cover a wider range of contingencies what might be considered "correct".

Question 5: How do the results change if we look at the performance of computer-assisted coding (agreement rate among top 5)?

The reported low rates of agreement between human coders raise doubt whether there exists a single appropriate occupational category for every individual. To increase the inter-coder reliability, it certainly helps if coders receive the same training, know the same conventions, and interpret ambiguous texts and additional information in the same way. Yet, these measures impair the objectivity of coding results because they depend on the organization of the coding process. Even if coders use the same coding procedures, there is, for some individuals, still human judgment involved and coders may disagree about the correct code. Thus, we suggest that ambiguity is inherent in occupation coding and one should account for the accompanying uncertainty.

Point forecasts, in our case this could be the single most probable category

predicted by an algorithm, cannot account for this uncertainty. To acknowledge the uncertainty better, several authors from different domains (Dawid, 1984, Gneiting and Katzfuss, 2014, among others) have called for issuing probabilistic forecasts instead. Some algorithms proposed earlier provide such probabilistic forecasts $\hat{p}_{n_f} = (\hat{p}_{n_f 1}, ..., \hat{p}_{n_f K})$ for $n_f = 1, ..., N_f$ future individuals. It is desirable to evaluate these probabilistic forecasts directly and not just the derived point forecasts (e.g., agreement rate among top 1).

From the applied interview coding perspective it is also desirable to evaluate probabilistic forecasts. If respondents have a low probability to find an appropriate category among a list of category suggestions, little is gained by asking respondents an extra question. Schierholz, Gensicke, Tschersich and Kreuter (2018) thus suggest that the production rate of interview coding needs to be chosen carefully. Ideally, we would like to know the probability that a respondent chooses a category. Using Kolmogorov's axiom of additivity, we could then sum over the probabilities of each category to obtain the probability that the respondent finds his occupation among the list of suggestions. Within a decision-theoretic framework one might further define losses of asking or not asking respondents an extra question. For example, the loss of asking a question that the respondent cannot answer is small and positive, because the extra question lengthens the interview without any gain. Based on these losses and probabilities, decision theory provides a framework to calculate whether or not the extra question should be asked. It would give a mathematical solution to the aforementioned problem.

Various metrics to evaluate probabilistic forecasts exist, outlined in the the on-line appendix. A common requirement is that probabilities are calibrated, i.e., the predicted probabilities and the observed relative frequencies should be approximately equal.

Question 6: Are the predicted probabilities calibrated?

Of course, all our results will depend on the choices made within this research project and may not always generalize to other situations. As a resort, we run robustness checks on several data sets and publish our algorithms on-line to simplify replication.

Importantly, this research focuses on algorithms that could be used for interview coding. This implies that predictions will be generated immediately after respondents enter their first verbal response and without using any other information. This is an important distinction, different from current post-

interview coding situations, in which more information is usually available. Better results could be achieved if the algorithms would base their decision on more information, imitating professional coders. While possible, this is out of scope for this research.

4 Data

Our data stem from six different surveys and, after pooling data from two highly similar health-related studies, we have five data sets at our disposal. In particular, our data are from the survey '*Working and learning in a changing world*' (data set A) (Antoni et al., 2010), the '*BIBB/BAuA Employment Survey 2012*' (data set B) (Rohrbach-Schmidt and Hall, 2013), the '*German Health Update 2014/2015*' (Lange et al., 2017), the second wave of the '*German Health Interview and Examination Survey for Children and Adolescents*' (Hoffmann et al., 2018) (both pooled in data set C), the tenth wave of the panel study '*Labour Market and Social Security*' (data set D) (Trappmann et al., 2013), and the survey '*Selectivity effects in address handling*' (data set E) (Schierholz, Gensicke, Tschersich and Kreuter, 2018). We describe all data sets and their respective coding procedures in more detail in the on-line appendix.

The surveys differ to a considerable degree and we highlight three such differences that might be most relevant with respect to our application. Firstly, respondents were asked about *different types of jobs*, i.e. about current jobs, past jobs, their parents' jobs, and so on. Secondly, the *question wording and its presentation* in each study are often similar to another, but not identical. Finally, the verbal answers in each study were coded by *different coding agencies* and, following Massing et al. (2019), we expect that each agency has their own coding conventions, implying that coding decisions will systematically differ if the same answers are coded by different agencies.

Table 2 provides a descriptive summary of the five data sets. The number of observations in each data set is calculated after excluding some verbal answers that we do not use in our analysis. The first reason for exclusion is if a verbal answers was coded into some category for 'not employed'. This mostly happens if filters in the questionnaire were not used as intended and respondents answered occupational questions although they should not have done so. These respondents did not report an occupation but something else (unemployed, retired, student, Housewife, don't know, ...). The second

reason for exclusion is if answers are after removal of punctuation marks and empty spaces less than two characters long and therefore uncodable.

The amount of information available from each respondent, measured as the average number of words in each description, differs substantially between data sets. We believe the main drivers for this are, firstly, the questionnaire design, which differs in terms of positioning and technical implementation of occupational questions, secondly, the interviewers, who may have different standards about the level of detail needed, and, thirdly, the respondents, whose motivation and ability may be low to precisely describe some specific occupation they or their parents had a long time ago. All respondents in data sets A, C, and E were asked a follow-up question to collect further details about their occupations. In contrast, the vast majority of respondents in data sets B and D was not asked a follow-up question.

Verbal answers in all data sets were coded according to the 2010 GCO. Some occupations are very uncommon and, thus, no study used the complete range of 1,286 5-digit categories from the classification. Moreover, it is impossible to find for every verbal answer a single appropriate category in the 2010 GCO. To solve this issue, all coding agencies introduced some additional, special codes that are not part of the official classification. Since we want to have codes that are comparable across data sets, we recode and harmonize the special codes. We only keep two extra categories for 'Student assistant' and 'Federal volunteer service / Voluntary social service / Civil service', both arguably very job-like although not included in the 2010 GCO, and three broader, poorly defined categories 'Multiple jobs', 'Job description not precise enough / Uncodeable', and 'Blue-collar worker' for hard-to-code answers. Table 2 provides frequencies of the special codes in our data sets after harmonization.

5 Analytical Strategy

For our current study, verbal answers from all data collections are preprocessed in a basic manner to make very similar strings identical to another, thus reducing the dimensionality of the data. Letters are capitalized, special characters are replaced (e.g. 'Ä' → 'AE', '€' → 'EURO'), punctuation is removed, and space characters at the start and the end of all strings are trimmed. At this point, we do not use other preprocessing steps like the

Table 2: Data Summary

	(A)	*(B)*	*(C)*	*(D)*	*(E)*
Mode	CATI	CATI	Paper & Web	CATI & CAPI	CATI
Coding Agency	IAB	Kantar	RKI	infas	Kantar
First answer					
No. of job descriptions	32931	55944	48994	7639	1064
No. of words					
.... Median	1	1	1	2	1
.... Mean	1.6	1.6	1.8	2.9	2.1
.... Maximum	9	34	26	31	13
Second answer (if informative)					
No. job descriptions	19330	-	39389	-	1013
No. of words					
.... Median	2	-	5	-	5
.... Mean	2.7	-	5.5	-	6.7
.... Maximum	10	-	45	-	33
Job codes					
No. of GCO codes used†	853	1057	1106	685	351
Freq. of special codes					
. Student Assistant	308	60	124	-	-
. Social Service	-	15	9	-	-
. Multiple Jobs	41	-	-	-	-
. Blue-collar worker	50	797	-	-	1
. Not precise enough/uncodeable	152	1805	311	186	-

†The 2010 GCO consists of 1286 5-digit job categories.

removal of stop words or stemming that are commonly used for text mining. Although this could improve the results further (see Gweon et al., 2017), one would risk losing relevant information during data preparation. Since the optimal type of preprocessing may depend on the algorithm, we regard other options for preprocessing as tuning parameters that are to be tested in the model tuning phase.

For each algorithm we would like to know how well it predicts categories from data set E and similar future data sets (questions 1+2). This test data was chosen because it represents our population of interest (adults were asked about their own occupation) and because more information than usual about each person's job was available to coders, which we take as a sign of high data quality. Schierholz, Gensicke, Tschersich and Kreuter (2018) analyzed the data quality in this data set. For these reasons we have highest confidence in our evaluations when data set E is used. Results obtained from other data sets are a by-product of our analysis. They provide a robustness check and allow us to analyze whether particularities of the test data matter (question 3).

Making predictions is straightforward when using the coding index (Exact Matching and CASCOT). In contrast, the choice of the training data matters for all other algorithms. To demonstrate this, we use data sets A, B, C, and D as training data and run separate analyses. In addition, we pool the four data sets to create a fifth data set, called (A,B,C,D). This last training data set allows us to analyze whether predictions improve when more training data from various studies are available (question 4).

For all algorithms that use training data we need to carefully select appropriate parameters in a tuning phase. Tuning can be time-intensive when a dense grid of possible values is searched via cross-validation in large data sets. Our approach is more pragmatic as we do not aim to find optimal values for the tuning parameters in a continuous parameter space. Unlike typical cross-validation approaches (Pers et al., 2009), our goal is also not to control for randomness in the selection of the evaluation data. Instead, our goal is to obtain tuning parameters that are not too far off the optimal value. We only argue that the predictions in data set E could not be dramatically improved with any other choice of tuning parameters.

All data sets except E are split at random, resulting in 1,064 test observations for each (selected to be identical in size as data set E) and all left-overs are used as training observations. Since data set D is smallest and the algorithms run fastest here, this data set is used to test and select possible

tuning parameters. For each algorithm, a grid search is performed in data set D. The same procedure (random splitting and grid search) is also run in data set C, showing that identical tuning parameters are still reasonable in this larger data set. The on-line appendix provides complete results from our grid search in both data sets. Thus, all tuning parameters were selected using data sets D and C only, with two exceptions. The penalty parameter λ needed for logistic regression is dependent on the respective data set. Likewise, the number of iterations in `XGBoost` is dependent on the respective data set. As a result, evaluations with test data from data sets A, B, C, D, and (A,B,C,D) need to be taken with a grain of salt since we peeked at those data sets to make the predictions. Yet, we do not believe that this peeking leads to highly different results—and data set E remained completely untouched.

While we can only measure agreement rates in our test data, we wish to make inferences about the population it was drawn from. By how much might the agreement rate vary, had we observed different persons in our test data? Since automatic coding procedures are deterministic (conditional on the verbal answer and after deciding which data are used for training) the number of agreements with professional coding decisions can be modeled as a Binomial random variable. Standard errors of agreement rates are therefore estimated using the formula $\sigma(\text{Agreement rate}) = \sqrt{\frac{1}{N}\text{Agreement rate}(1 - \text{Agreement rate})}$.

Mathematical definitions of all metrics used for evaluation are available in the on-line appendix.

6 Results

Preliminaries

The simplest algorithm for automated coding is exact matching with a coding index (algorithm 1). Its results are shown in Table 3. Production rates vary between 28.7% and 48.8%. The production rate in data set D is exceptionally low because many verbal answers are more than one word long, longer than usual job titles in German language, making exact matching with job titles from the coding index all but impossible.

Looking at the subset of answers that have matching entries in the coding index, the agreement rates between codes from the coding index and from manual codings range between 72.7% and 93.1%. For comparison, Westermark et al. (2015) reported an ideal 100% agreement rate in a Swedish survey

Table 3: Production Rates and Agreement Rates Among Top 1 from Coding with the Alphabetic Dictionary (algorithm 1) for each data set

	(A)	*(B)*	*(C)*	*(D)*	*(E)*
Production Rate [%]	45.96	48.78	41.82	28.67	42.01
Agreement Rate [%]	77.91	89.21	75.73	93.11	72.71

and Takahashi et al. (2005) were dissatisfied with agreement rates around 80% in Japanese surveys. Since our coding index can be understood as an 'official' mapping between job titles and categories, implying that agreement rates should be 100%, the low agreement rates we find here may come as a surprise. Yet, coders have reason to deviate from the coding index, because, if the first verbal answer is ambiguous, they may rely on other survey variables, e.g., industry, supervisory status, occupational status, or the education usually required for a job. In addition, respondents were prompted to describe their jobs in detail. Since questionnaires and their implementation in the interviewing software differed, a second text field containing responses from all respondents is available only in data sets A, C, and E. If available, coders consider this second text field for their decisions, but the exact matching algorithm does not use this information. This second text field provides a plausible explanation for the lower agreement rates in data sets A, C, and E.

Exact Matching with a coding index yields unsatisfactory results, but it provides a baseline that more advanced algorithms should meet.

All other algorithms are more flexible than exact matching because they output a relevance score (or a predicted probability) associated with the predicted category. We sort different verbal answers by this score. If the score is above a user-defined threshold, the textual input is coded automatically. This allows users to select their preferred production rate and its corresponding agreement rate.

Diagrams like the one in Figure 1 are a helpful tool to explore the trade-off between production rates and agreement rates. Likewise, they can be used to compare the agreement rates of different algorithms at any fixed production rate. Although this type of visualization is our favorite way to compare the various algorithms, the space in this article is limited, forcing us to provide the complete set of diagrams only in the on-line appendix. Some key results and important patterns from all diagrams are also visible in Table 4, which

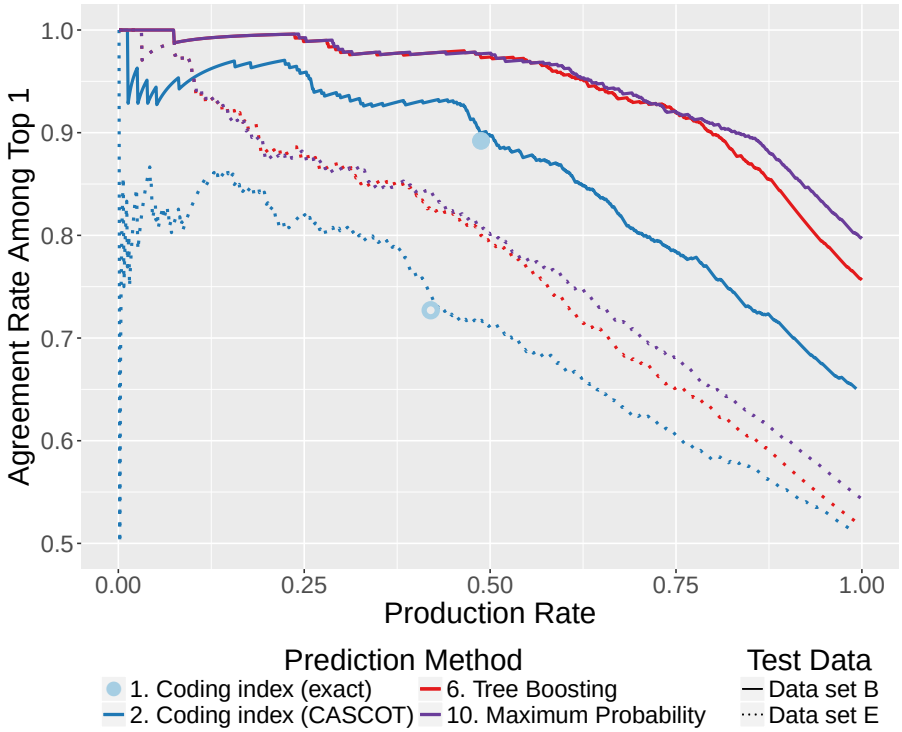


Figure 1: Agreement rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases). Example: 48.8% (production rate) of the responses in test data set B have an exact match with the coding index (algorithm 1). Within this subset, 89.2% (agreement rate) of the assigned categories agree with evaluation data. The production rates and their respective agreement rates of any other algorithm are controlled by the user choosing a threshold.

contains agreement rates when the production rates are fixed at 100%. If an algorithm does not achieve a 100% production rate, it will misclassify all remaining answers, implying that the *agreement rate at 100% production rate* = *production rate* $\cdot$ *agreement rate* + (1 − *production rate*) $\cdot$ 0.

Question 1: For each algorithm, which performance (production rate, agreement rate among top 1) can be achieved?

To provide a preliminary answer we first look at results when the training data are from data set A, B, C, or D and the test data are from the same data set (upper left in Table 4). Since the table is limited to agreement rates at 100% production rate, we consult diagrams from the on-line appendix for

Table 4: Agreement rate among top 1 [%] at 100 percent production rate ± standard errors (appropriate for automated coding).

	Training data				
	(A)	*(B)*	*(C)*	*(D)*	*(A,B,C,D)*
	Test data are from the same data set as the training data.				
1. Coding Index (exact)	35.81 ± 1.47	43.52 ± 1.52	31.67 ± 1.43	26.69 ± 1.36	35.06 ± 1.46
2. Coding Index (CASCOT)	56.58 ± 1.52	64.57 ± 1.47	52.82 ± 1.53	55.64 ± 1.52	57.05 ± 1.52
3. Mem-based Reasoning	67.95 ± 1.43	73.50 ± 1.35	58.18 ± 1.51	57.61 ± 1.51	64.47 ± 1.47
4. Adapt. Nearest Neighbor	65.60 ± 1.46	74.15 ± 1.34	57.05 ± 1.52	57.52 ± 1.52	63.82 ± 1.47
5. Multinomial Regression	68.14 ± 1.43	76.41 ± 1.30	59.02 ± 1.51	60.34 ± 1.50	†58.93 ± 1.51
6. Tree Boosting (XGBoost)	66.92 ± 1.44	75.66 ± 1.32	59.68 ± 1.50	59.12 ± 1.51	63.82 ± 1.47
7. Fulltext Similarity	53.95 ± 1.53	66.17 ± 1.45	44.17 ± 1.52	35.43 ± 1.47	49.53 ± 1.53
8. Substring Similarity	62.31 ± 1.49	70.49 ± 1.40	55.55 ± 1.52	55.73 ± 1.52	57.89 ± 1.51
9. Wordwise Similarity	62.31 ± 1.49	70.96 ± 1.39	54.79 ± 1.53	53.29 ± 1.53	57.99 ± 1.51
10. Maximum Probability	**70.02** ± 1.40	**79.70** ± 1.23	**63.16** ± 1.48	**66.82** ± 1.44	**66.82** ± 1.44
	Test data are from data set E.				
1. Coding Index (exact)	30.55 ± 1.41	30.55 ± 1.41	30.55 ± 1.41	30.55 ± 1.41	30.55 ± 1.41
2. Coding Index (CASCOT)	**50.66** ± 1.53	50.66 ± 1.53	50.66 ± 1.53	**50.66** ± 1.53	50.66 ± 1.53
3. Mem-based Reasoning	43.52 ± 1.52	49.62 ± 1.53	47.84 ± 1.53	39.10 ± 1.50	53.95 ± 1.53
4. Adapt. Nearest Neighbor	42.86 ± 1.52	49.34 ± 1.53	45.77 ± 1.53	38.16 ± 1.49	53.29 ± 1.53
5. Multinomial Regression	45.02 ± 1.53	51.50 ± 1.53	49.72 ± 1.53	40.23 ± 1.50	†49.25 ± 1.53
6. Tree Boosting (XGBoost)	44.17 ± 1.52	51.69 ± 1.53	49.15 ± 1.53	39.85 ± 1.50	54.70 ± 1.53
7. Fulltext Similarity	36.18 ± 1.47	39.10 ± 1.50	37.31 ± 1.48	37.50 ± 1.48	39.94 ± 1.50
8. Substring Similarity	45.30 ± 1.53	48.31 ± 1.53	47.65 ± 1.53	44.17 ± 1.52	50.56 ± 1.53
9. Wordwise Similarity	45.11 ± 1.53	46.52 ± 1.53	46.52 ± 1.53	43.52 ± 1.52	48.31 ± 1.53
10. Maximum Probability	**50.66** ± 1.53	**54.32** ± 1.53	**51.97** ± 1.53	50.47 ± 1.53	**56.77** ± 1.52

†Regularized Multinomial Regression does not converge in data set (A,B,C,D), leading to poor performance.

complementary results at lower production rates.

The algorithms' performance varies widely among different data sets. For example, at 100% production rate the Maximum Probability algorithm (algorithm 10) achieves 66.8% agreement rate in data set D and 79.7% in data set B. At 50% production rate it achieves 89.8% agreement rate in data set D and 97.7% in data set B. The absolute performance thus depends on the specific data set in use. Generalizations for new data sets are hard to obtain.

Question 2: Which algorithm performs best?

Looking at the production rates achieved by the exact matching algorithm, Figures A2 to A5 in the on-line appendix show that in each data set exact matching (algorithm 1) and CASCOT (algorithm 2) exhibit similar agreement rates if production rates are fixed by algorithm 1. Yet, as a major improvement, the CASCOT algorithm provides greater flexibility, allowing users to choose any production rate they like. If a 100% production rate is desired, the differences between both algorithms are more pronounced (see Table 4), reflecting that this metric, the agreement rate at 100% production rate, is ill-suited to describe the performance of the exact matching algorithm.

Table 4 and Figures A2 to A5 in the on-line appendix show that all algorithms outperform the CASCOT algorithm, in some data sets by wide margins of more than ten percentage points. Only Memory-based Reasoning (algorithm 3) is usually worse at low production rates.

Comparing the four algorithms that rely on training data only (algorithms 3-6) against each other, the agreement rates at 100% production rate are quite similar within each data set (see Table 4). The ranges between the worst-performing algorithm and the best-performing algorithm are usually small, never exceeding three percentage points. These small differences may well be irrelevant to a practitioner. If we still wish to rank the different algorithms by their relative performance in each data set, multinomial regression is often best. XGBoost is on all data sets a close competitor. Memory-based Reasoning and the Adapted Nearest Neighbor algorithm often come in third and forth.

This picture changes if we look at agreement rates at low and medium production rates (see Figures A2 to A5 in the on-line appendix). All algorithms outperform Memory-based Reasoning by wide margins. Similarly, but only at very low production rates, the agreement rates from multinomial

regression decline rapidly before improving again. The reason is that both algorithms are overconfident and output scores that are too high for answers that are, in fact, difficult to code. Only XGBoost and the Adapted Nearest Neighbor algorithm achieve agreement rates that are close to optimal at low and medium production rates.

We now turn to the novel similarity-based algorithms that use both the coding index and the training data (algorithms 7-9). As these algorithms were not designed to achieve 100% production rate, the agreement rates at 100% production rate are often rather low (see Table 4). The performance is better understood from Figures A2 to A5 in the on-line appendix. Using fulltext similarity (algorithm 7), the maximal production rates are in all data sets between 61 and 73%, except in data set D, which has a production rate as low as 37% due to the unusual length of the answers. If the criterion what is considered as similar is relaxed, the production rates improve (90% to 95% with algorithm 9 (substring similarity); 85% to 91% with algorithm 8 (word-wise similarity)). Looking at lower production rates, the similarity-based algorithms frequently outperform CASCOT, but rarely achieve the performance of XGBoost. A notable exception is the fulltext similarity algorithm applied on data set B, which achieves a 99% agreement rate at 50% production rate. This means if circumstances are the same as they were when data set B was coded, we could completely automate half the work and still obtain basically identical codings.

Unlike the supervised learning algorithms from above, the similarity-based algorithms do not use all words from each answer, but only the subset of words that are similar to entries from the coding index. Since this ignores possibly relevant information, similarity-based algorithms and supervised learning algorithms may be complementary and combining them may lead to better results. To test this hypothesis, the Maximum Probability algorithm is an ensemble of XGBoost and the three similarity-based algorithms. Figures A2 to A5 in the on-line appendix show that the performance of the Maximum Probability algorithm (algorithm 10) is very similar to XGBoost at low production rates, but at higher production rates it is often better than XGBoost and, in fact, at 100% production rate it outperforms all other algorithms.

These general patterns are quite similar across the four different data sets. This provides a robustness check for the relative comparison of algorithms presented above.

Question 3: By how much does the agreement rate degrade if training data and test data have been created in different ways?

The data set that best represents our future application is data set E. Respondents in this data set were asked several questions about their own job to obtain as much details as possible, admitting high-quality coding. The results obtained from this data set are therefore of particular interest.

The agreement rates drop dramatically, in some cases by more than 25 percentage points, if one uses data set E for evaluation and not data from the same data set that was used for training (see bottom in Table 4). At least two reasons exist. Firstly, automatic coding of data set E appears to be particularly hard. Evidence for this comes from the exact matching algorithm and from the CASCOT algorithm (algorithms 1 and 2), which perform, among all data sets, poorest in data set E (see Figures A2 to A5). An explanation is that we used only the first verbal answer for automatic coding, as opposed to human coders who had access to more information. Secondly, supervised learning algorithms may find patterns in the training data that they extrapolate to the test data. Since different coding procedures and conventions were used in data set E, this will lead to higher disagreement. The exact matching algorithm and the CASCOT algorithm is not harmed by this effect, because they do not rely on training data.

Looking at the relative performance of each algorithm in data set E, we find most results from the comparison above confirmed. In particular, the Maximum Probability algorithm (algorithm 10) performs similar or better than all other algorithms. Also, multinomial regression and XGBoost (algorithms 5 and 6) are at 100% production rate better than Memory-based Reasoning and the Adapted Nearest Neighbor algorithm (algorithms 3 and 4). An exception is the CASCOT algorithm (algorithm 2), which previously performed worse than most other algorithms. In contrast, its agreement rates at low and medium production rates are more competitive in data set E, especially if other algorithms are trained with data set A or D. At 100% production rate, the agreement rates of CASCOT are similar to the Maximum Probability algorithm if trained with data set A, or D, outperforming all other algorithms. Another point to note is the poor performance of algorithms 3 to 6 if trained with data set D to predict data set E (see Figure A5). The similarity-based algorithms 8 and 9 outperform in this small training data set the aforementioned algorithms, which depend solely on training data.

Question 4: Can we improve predictions by pooling different data sets to form larger training data?

Data sets A, B, C, and D are pooled to form a combined data set (A,B,C,D), which is then split at random to obtain 144,444 training observations and 1,064 test observations. We also note that the sizes of the training data sets differ. Data set D is smallest with 6,575 training observations; data set A has 31,867 training observations; data set C has 47,930 training observations; and data set B has 54,880 training observations.

The last column in Table 4 shows agreement rates for the pooled data set and Figure A6 shows results at lower production rates. The test observations in data set (A,B,C,D) have been coded inconsistently by different coding agencies, making them hard to predict without using information about the coding agency. Predicting test data from data sets A or B is simpler. This explains why agreement rates when using test data (A,B,C,D) are lower than agreement rates when using test data from data sets A or B, despite the larger size of the training data (A,B,C,D).

Data set E provides a better benchmark that avoids comparing apples with oranges. For about any algorithm that uses training data we find improving agreement rates at 100% production rate as the size of the training data increases (D < A < C < B < (A,B,C,D)). Thus, the best predictions for data set E are obtained when using the pooled data set (A,B,C,D) for training.

Question 5: How do the results change if we look the at performance of computer-assisted coding (agreement rate among top 5)?

To achieve high data quality, computers should make the same coding decisions as human coders would. However, agreement rates close to 100% are only achieved at very low production rates and may drop dramatically as the production rates increase. While the exact numbers depend on the data sets used for training and testing, we worry that the agreement rates will usually not suffice for high quality coding. This means fully automated coding will only be useful for a smallish proportion of answers. For most answers we will need computer-assisted coding, i.e., the computer suggests some categories and a human chooses the most appropriate category.

For this purpose we calculate agreement rates among top five. While the

agreement rate among top one, used above, evaluates whether the highest-ranked category is in agreement with a human coder, the agreement rates among top five evaluates whether the human-coded category is among the top five suggestions.

Using this alternative way to calculate the agreement rate, the agreement rates increase by several percentage points. Still, the patterns described above remain essentially similar. Detailed results are provided in Table 5 and in Figures A17 to A21 in the on-line appendix. Only the key differences are highlighted in this section.

Interestingly and unlike before, the similarity-based algorithm with substring similarity (algorithm 8) becomes a close competitor at high production rates. Especially with smaller training data sets the algorithm is as good as XGBoost (algorithm 6) at 100% production rate, in data set D even better. The reason is that this algorithm exploits the coding index to find additional categories, improving the top-ranked set of suggestions.

If data set E is used for testing, the Maximum Probability algorithm (algorithm 10) is not any longer the top performer it was previously. Instead, it depends on the available training data and the desired production rate which algorithm is best. For production rates around 50% the similarity-based algorithm with substring similarity (algorithm 8) is very powerful (see Figures A17 to A21). In practice, computer-assisted coding is more often used at 100% production rate. In this situation, CASCOT (algorithm 2) outperforms all other algorithms if only few training observations are available (training data sets D and A). With more training observations the algorithms that rely on training data become stronger. Thus, XGBoost (algorithm 6) and the similarity-based algorithms (algorithms 8 and 9) achieve their highest agreement rates among top 5 if data set (A,B,C,D) is available for training. Combining tree boosting and the similarity-based algorithms (algorithm 10) is even better (71.5% agreement rate at 100 % production rate), once again outperforming all other algorithms.

Question 6: Are the predicted probabilities calibrated?

To check whether our predicted probabilities are calibrated, we provide various reliability diagrams in Figures A7 to A16 and Figures A27 to A39 in the on-line appendix. Tables A13 and A14 show results on sharpness and logloss. No probabilistic analysis is carried out for algorithms 1-3 because they are not described in probabilistic terms.

Table 5: Agreement rate among top 5 [%] at 100 percent production rate ± standard errors (appropriate for computer-assisted coding).

	Training data				
	(A)	*(B)*	*(C)*	*(D)*	*(A,B,C,D)*
	Test data are from the same data set as the training data.				
2. Coding Index (CASCOT)	66.82 ± 1.44	76.69 ± 1.30	70.39 ± 1.40	70.30 ± 1.40	69.92 ± 1.41
4. Adapt. Nearest Neighbor	72.27 ± 1.37	80.08 ± 1.22	69.45 ± 1.41	61.47 ± 1.49	74.62 ± 1.33
5. Multinomial Regression	76.41 ± 1.30	83.65 ± 1.13	76.32 ± 1.30	68.98 ± 1.42	†67.39 ± 1.44
6. Tree Boosting (XGBoost)	76.88 ± 1.29	83.27 ± 1.14	76.60 ± 1.30	69.17 ± 1.42	79.61 ± 1.24
7. Fulltext Similarity	59.02 ± 1.51	69.27 ± 1.41	53.20 ± 1.53	35.71 ± 1.47	57.52 ± 1.52
8. Substring Similarity	75.94 ± 1.31	82.52 ± 1.16	73.59 ± 1.35	71.43 ± 1.38	77.16 ± 1.29
9. Wordwise Similarity	73.21 ± 1.36	81.30 ± 1.20	70.96 ± 1.39	64.94 ± 1.46	73.68 ± 1.35
10. Maximum Probability	**79.42** ± 1.24	**87.69** ± 1.01	**79.79** ± 1.23	**77.26** ± 1.29	**83.18** ± 1.15
	Test data are from data set E.				
2. Coding Index (CASCOT)	**63.82** ± 1.47	63.82 ± 1.47	63.82 ± 1.47	**63.82** ± 1.47	63.82 ± 1.47
4. Adapt. Nearest Neighbor	50.66 ± 1.53	56.77 ± 1.52	57.80 ± 1.51	42.67 ± 1.52	63.44 ± 1.48
5. Multinomial Regression	55.64 ± 1.52	62.59 ± 1.48	64.85 ± 1.46	48.12 ± 1.53	†56.95 ± 1.52
6. Tree Boosting (XGBoost)	55.08 ± 1.52	60.53 ± 1.50	63.72 ± 1.47	48.40 ± 1.53	68.05 ± 1.43
7. Fulltext Similarity	41.92 ± 1.51	42.58 ± 1.52	43.80 ± 1.52	38.82 ± 1.49	45.21 ± 1.53
8. Substring Similarity	60.06 ± 1.50	63.35 ± 1.48	65.51 ± 1.46	58.36 ± 1.51	67.86 ± 1.43
9. Wordwise Similarity	56.67 ± 1.52	59.21 ± 1.51	60.81 ± 1.50	53.85 ± 1.53	63.53 ± 1.48
10. Maximum Probability	61.75 ± 1.49	**65.41** ± 1.46	**68.33** ± 1.43	60.90 ± 1.50	**71.90** ± 1.38

†Regularized Multinomial Regression does not converge in data set (A,B,C,D), leading to poor performance.

Algorithms 1 and 3 are not included because they output only a single category.

Looking at diagrams where training data and test data are from the same data set, we find that, in general, predicted probabilities from XGBoost (algorithm 6) and multinomial logistic regression (algorithm 5) are usually well calibrated. The Maximum Probability Algorithm (algorithm 10) resembles algorithm 6 closely if the forecasted probabilities are high, unsurprisingly, because the predictions are basically identical. The adapted nearest neighbor algorithm (algorithm 4) outputs scores that are frequently too high, whereas the predicted probabilities from similarity-based algorithms (especially substring similarity and wordwise similarity, algorithms 7-9) are often lower than the observed relative frequencies.

The picture changes if data set E is used as test data. While the algorithms output the same probabilities as before, the observed relative frequencies of agreement decrease. Thus, the predicted probabilities from most algorithms are usually higher than their observed counterparts. This makes all predicted probabilities over-optimistic, caused by systematic differences across data sets. Only the similarity-based algorithms 8 and 9, found to be under-pessimistic before, happen to output probabilities that are often close to the observed relative frequencies.

In the absence of calibrated probabilities, decision theory should not be used to find appropriate thresholds. Instead, we propose plotting the number of true positives against the number of false positives for different thresholds (see Figures A22 to A26 in the on-line appendix) to make this decision.

7 Discussion

Manual occupation coding is time-consuming and its automation could be useful in many fields. For this purpose we compared ten algorithms that help selecting occupational categories, focusing on algorithms that use previously coded data. After reviewing available algorithms for occupation coding, we compare the algorithms using five data sets. Since each study has its own processes for data collection and coding, we find large differences across studies. It thus depends on the study and the research question which algorithm is best.

If only a coding index is available but no training data, users will need to use exact matching (algorithm 1), similarity matching (algorithms 2), or variants thereof. Exact matching might be most useful for automated coding if consistent codings are required and the coding index is viewed as a gold

standard. We reported agreement rates far off from ideal values at 100%, indicating that this view is not widespread in Germany. Instead, human coding is widely considered the gold standard and computer-assisted coding systems are common. In this situation, the similarity matching algorithm as implemented in CASCOT shines. It makes decent category suggestions for almost all individuals without being influenced by debatable coding decisions made in the training data.

If there are training data available but no coding index, users can choose between various machine learning algorithms and we tested four of them. These algorithms allow imitating coding decisions found in the training data and perpetuating them to future data sets. They achieve higher agreement rates than algorithms 1 and 2 if training data and test data have been created through identical coding procedures or if the training data are sufficiently large. Regularized multinomial regression (algorithm 5) and tree boosting (algorithm 6) usually outperform Memory-based Reasoning (algorithm 3) and the Adapted Nearest Neighbor algorithm (algorithm 4), with the exception that multinomial regression performs poorly at very low production rates. Yet, differences among the four machine learning algorithms are rather small and possibly irrelevant in practice, especially when considering the agreement rates of the most probable categories at 100% production rate. Practitioners might worry more about computational resources needed for training or hassles with parameter tuning when using the more complex algorithms 5 and 6. Thus, the Adapted Nearest Neighbor algorithm (algorithm 4) is a good choice to find out quickly what is achievable, whereas tree boosting (algorithm 6) allows users to get the most out of their training data.

If there are both a coding index and training data available, the results from algorithms 3 to 6 can be improved further. These algorithms fail to use important words, leading to poor predictions, if such words have never been recorded in the training data (e.g., misspellings, infrequent job titles). We develop Similarity-based Reasoning (algorithms 7 to 9) to counter this weakness, combining information from both the coding index and from training data. This culminates in the Maximum Probability algorithm (algorithm 10), an ensemble of tree boosting (algorithm 6) and Similarity-based Reasoning, which is usually among the best algorithms under many different circumstances. It is the exception that other algorithms are better, observed only if both the number of observations in the training data is low and the test data have been created through a different coding procedure (data set E).

It is notable how different the results are for each data set. Imagine a new sample from the same population as in data set B and we want to use the same coding process used before. More than 50% of the verbal answers from this sample could be coded completely automatically with near perfect agreement. In contrast, with data sets C and E the same agreement rate (called the agreement rate among top 1 at 50% production rate above) is rarely above 80%, insufficient for high-quality automated coding. The lower agreement rates are found in data sets for which coders had access to more additional information from respondents (unused by our algorithm), whereas the higher agreement rates are found in data sets which were created with more extensive usage of an exact matching algorithm. Since the coding processes to create these data sets were already partially automated, it is unclear whether the algorithms proposed in this paper would bring additional benefits.

Having found that improvements in automated coding are difficult to achieve, we explore the usefulness of machine learning for computer-assisted coding and, more specifically, interview coding. In the planned application, not expert coders but respondents themselves will be asked during an interview to select one out of five occupational categories. In this situation the agreement rate among top 5 (i.e., the relative frequency how often coders selected one of the five highest ranked categories) is a good proxy for how often respondents will find their own occupational category in the list of suggestions. To reduce respondent burden, we would not want that categories are suggested if all suggestions are wrong. Data set E is used to evaluate the results as this represents our population of interest and coders from the study had access to the most detailed information to make their decisions. The results suggest that the Maximum Probability algorithm (algorithm 10) is best for this situation and should be trained using all available training data. It achieves a 84% agreement rate among top 5 at 0.75% production rate, making this algorithm highly promising for our future application.

Many possibilities exist that might improve the results further and could be tested. Additional training observations, if available, should lead to improvements. In the same vein, job titles from the coding index might be viewed as additional training data. Coders had access to additional information from other variables and it is straightforward to make algorithms 5 and 6 use the same information. One may also wish to improve the algorithms further and we believe two directions are most promising. Firstly, algorithms 3 to 6 split text into words. This is not optimal because every character is

relevant if texts are short. Compound words, often found in German job titles (e.g., "Alarmanlagenmonteur" = "alarm system fitter"), are not split at all. Alternative techniques—like string distance calculations or splitting texts into short character sequences (n-grams)—might be fruitful but have not been fully exploited yet. Secondly, our literature review shows that combining different algorithms is common for occupation coding, although there is no consensus how this should be done. We achieved very promising results with the Maximum Probability algorithm (algorithm 10), which is yet another ad-hoc approach to combine results from different algorithms. A more principled method could prove useful.

Our data descriptions show that there exist subtle differences between occupational data sets, even if all verbal answers are written in German language and coded into the German Classification of Occupation. Thus, patterns observed in one data set may be different in another data set. Whether and how to use previously coded data to improve current coding processes depends on the available data and the desired application. Data scientists may wish to explore the possibilities they have with their own data. To facilitate this, we make our code available in an R-package, downloadable at `https://github.com/malsch/occupationCoding`.

Acknowledgements

Funding for this work has been provided by the German Institute for Employment Research and the Mannheim Centre for European Social Research, and by grant KR 2211/3-1 from the German Research Foundation to Frauke Kreuter. We are grateful for the support from the Federal Institute for Vocational Education and Training (BIBB) and from the Robert Koch Institute (RKI), who provided two of the five data sets used in this study. We sincerely thank our colleagues at the German Institute for Employment Research and at the University of Mannheim for their valuable comments.

References

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2010). Arbeiten und Lernen im Wandel * Teil 1: Überblick über die Studie, *FDZ-Methodenreport 05/2010*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Appel, M. V. and Hellerman, E. (1983). Census Bureau Experience with Automated Industry and Occupation Coding, *Proceedings of the Survey Research Methods Section: American Statistical Association*, pp. 32–40.

Bekkerman, R. and Gavish, M. (2011). High-Precision Phrase-Based Document Classification on a Modern Scale, *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, ACM, New York, NY, USA, pp. 231–239.
URL: *http://doi.acm.org/10.1145/2020408.2020449*

Bethmann, A., Schierholz, M., Wenzig, K. and Zielonka, M. (2014). Automatic coding of occupations, *Proceedings of Statistics Canada Symposium 2014*, Statistics Canada.
URL: *https://www.statcan.gc.ca/eng/conferences/symposium2014/program/14291 eng.pdf*

Bishop, Christopher, M. (2006). *Pattern Recognition and Machine Learning*, Springer, New York.

Bound, J., Brown, C. and Mathiowetz, N. (2001). Chapter 59 - Measurement Error in Survey Data, Vol. 5 of *Handbook of Econometrics*, Elsevier, pp. 3705 – 3843.
URL: *http://www.sciencedirect.com/science/article/pii/S1573441201050127*

Bushnell, D. (1998). An evaluation of computer-assisted occupation coding, *in* A. Westlake, J. Martin, M. Rigg and C. Skinner (eds), *New Methods for Survey Research, Proceedings of the International Conference*, Association for Survey Computing, Southampton, pp. pp. 23–36.

Campanelli, P., Thomson, K., Moon, N. and Staples, T. (1997). The quality of occupational coding in the United Kingdom, *in* L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz and D. Trewin (eds),

Survey Measurement and Process Quality, Wiley, New York, pp. pp. 437–453.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, ACM, New York, NY, USA, pp. 785–794.
URL: *http://doi.acm.org/10.1145/2939672.2939785*

Conrad, F. G., Couper, M. P. and Sakshaug, J. W. (2016). Classifying open-ended reports: Factors affecting the reliability of occupation codes, *Journal of Official Statistics* **32**(1): 75–92.

Creecy, R. H., Masand, B. M., Smith, S. J. and Waltz, D. L. (1992). Trading mips and memory for knowledge engineering, *Commun. ACM* **35**(8): 48–64.
URL: *http://doi.acm.org/10.1145/135226.135228*

Dawid, A. P. (1984). Statistical theory: The prequential approach (with discussion), *Journal of the Royal Statistical Society. Series A (General)* **147**(2): 278–292.
URL: *http://www.jstor.org/stable/2981683*

Elias, P. (1997). Occupational classification (ISCO-88): Concepts, methods, reliability, validity and cross-national comparability, *OECD Labour Market and Social Policy Occasional Papers 20*, OECD Publishing, Paris. (Available from http://dx.doi.org/10.1787/304441717388.).

Elias, P., Birch, M. and Ellison, R. (2014). CASCOT International version 5, *User Guide*, Institute for Employment Research, University of Warwick, Coventry. (Available from http://www2.warwick.ac.uk/fac/soc/ier/software/cascot/internat/ .).

Federal Employment Agency (2011). *Klassifikation der Berufe 2010*, Bundesagentur für Arbeit, Nuremberg.

Federal Employment Agency (2019). Gesamtberufsliste der Bundesagentur für Arbeit (Stand: 03.01.2019), *Gesamtberufsliste_der_BA.xlsx*, Bundesagentur für Arbeit, Nuremberg. (Available from http://download-portal.arbeitsagentur.de/files/.).

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine, *Ann. Statist.* **29**(5): 1189–1232.
URL: *https://doi.org/10.1214/aos/1013203451*

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* **33**(1): 1–22.

Geis, A. and Hoffmeyer-Zlotnik, J. H. (2000). Stand der Berufsvercodung, *ZUMA-Nachrichten* **24**(47): 103–128.

Gentzkow, M., Kelly, B. T. and Taddy, M. (2017). Text as Data, *NBER Working Paper No. 23276*, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w23276*

Gillman, D. W. and Appel, M. V. (1994). Automated Coding Research at the Census Bureau, *Statistical Research Report Series 94/04*, United States Census Bureau, Suitland.
URL: *https://www.census.gov/srd/papers/pdf/rr94-4.pdf*

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting, *Annual Review of Statistics and Its Application* **1**(1): 125–151.
URL: *https://doi.org/10.1146/annurev-statistics-062713-085831*

Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts, *Political Analysis* **21**(3): 267–297.

Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M. and Steiner, S. (2017). Three Methods for Occupation Coding Based on Statistical Learning, *Journal of Official Statistics* **33**(1): 101–122.

Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, 2 edn, Springer.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC Press, Boca Raton.

Hoffmann, R., Lange, M., Butschalowsky, H., Houben, R., Schmich, P., Allen, J., Kuhnert, R., Schaffrath Rosario, A. and Gößwald, A. (2018). KiGGS Wave 2 cross-sectional study – participant acquisition, response rates and representativeness, *Journal of Health Monitoring* **3**(1): 78–91.

Ikudo, A., Lane, J., Staudt, J. and Weinberg, B. (2018). Occupational Classifications: A Machine Learning Approach, *Working Paper 24951*, National Bureau of Economic Research, Cambridge, MA.

International Labour Office (2012). *International Standard Classification of Occupations: ISCO-08*, International Labour Organization, Geneva.

Javed, F., Luo, Q., McNair, M., Jacob, F., Zhao, M. and Kang, T. S. (2015). Carotene: A Job Title Classification System for the Online Recruitment Domain, *Proceedings of the IEEE International Conference on Big Data*, Institute of Electrical and Electronics Engineers (IEEE), Redwood City, pp. pp. 286–293.

Jung, Y., Yoo, J., Myaeng, S.-H. and Han, D.-C. (2008). A Web-Based Automated System for Industry and Occupation Coding, *in* J. Bailey, D. Maier, K.-D. Schewe, B. Thalheim and X. Wang (eds), *Web Information Systems Engineering - WISE 2008*, Vol. 5175 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 443–457.
URL: *http://dx.doi.org/10.1007/978-3-540-85481-4_33*

Knaus, R. (1987). Methods and Problems in Coding Natural Language Survey Data, *Journal of Official Statistics* **3**(1): 45–67.

Lange, C., Finger, J., Allen, J., Born, S., Hoebel, J., Kuhnert, R., Müters, S., Thelen, J., Schmich, P., Varga, M., von der Lippe, E., Wetzstein, M. and Ziese, T. (2017). Implementation of the European health interview survey (EHIS) into the German health update (GEDA), *Archives of Public Health* **75**(1): 40.
URL: *https://doi.org/10.1186/s13690-017-0208-6*

Lyberg, L. and Andersson, R. (1983). Automated Coding at Statistics Sweden, *Proceedings of the Survey Research Methods Section: American Statistical Association*, pp. 41–50.

Mannetje, A. t. and Kromhout, H. (2003). The use of occupation and industry classifications in general population studies, *International Journal of Epidemiology* **32**(3): 419–428.

Massing, N., Wasmer, M., Wolf, C. and Zuell, C. (2019). How standardized is occupational coding? A comparison of results from different coding agencies in germany, *Journal of Official Statistics* **35**(1): 167–187.
URL: *http://dx.doi.org/10.2478/JOS-2019-0008*

Measure, A. (2014). Automated coding of worker injury narratives, *Proceedings of the Government Statistics Section: American Statistical Association*, pp. 2124–2133.

Munz, M., Wenzig, K. and Bela, D. (2016). String coding in a generic framework, *in* H.-P. Blossfeld, J. von Maurice, M. Bayer and J. Skopek (eds), *Methodological Issues of Longitudinal Surveys: The Example of the National Educational Panel Study*, Springer Fachmedien Wiesbaden, Wiesbaden, pp. 709–726.
URL: *https://doi.org/10.1007/978-3-658-11994-2_39*

Nahoomi, N. (2018). *Automatically Coding Occupation Titles to a Standard Occupation Classification*, Master's thesis, University of Guelph, Guelph.
URL: *http://hdl.handle.net/10214/14251*

Nelson, L. K., Burk, D., Knudsen, M. and McCall, L. (2018). The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods, *Sociological Methods & Research* **Online First**: 0049124118769114.
URL: *https://doi.org/10.1177/0049124118769114*

O'Reagon, R. T. (1972). Computer-assigned codes from verbal responses, *Commun. ACM* **15**(6): 455–459.

Ossiander, E. M. and Milham, S. (2006). A computer system for coding occupation, *American Journal of Industrial Medicine* **49**(10): 854–857.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/ajim.20355*

Paulus, W. and Matthes, B. (2013). Klassifikation der Berufe * Struktur, Codierung und Umsteigeschlüssel, *FDZ-Methodenreport 08/2013*,

Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Pers, T. H., Albrechtsen, A., Holst, C., Sørensen, T. I. A. and Gerds, T. A. (2009). The validation and assessment of machine learning: A game of prediction from high-dimensional data, *PLOS ONE* **4**(8): 1–8.
URL: *https://doi.org/10.1371/journal.pone.0006287*

R Core Team (2016). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Rauchberg, H. (1888). Die deutsche Berufs- und Betriebszählung vom 5. Juni 1882, *Statistische Monatsschrift* **14**: 569–603.

Riviere, P. (1997). Automated Coding - Foreword, *in* United Nations Statistical Commission and Economic Commission for Europe (ed.), *Statistical Data Editing Volume No. 2*, United Nations, New York.

Rohrbach-Schmidt, D. and Hall, A. (2013). BIBB/BAuA Employment Survey 2012, *BIBB-FDZ Data and Methodological Reports Nr. 1/2013. Version 4.1*, Federal Institute for Vocational Education and Training, Bonn.

Russ, D. E., Ho, K.-Y., Colt, J. S., Armenti, K. R., Baris, D., Chow, W.-H., Davis, F., Johnson, A., Purdue, M. P., Karagas, M. R., Schwartz, K., Schwenn, M., Silverman, D. T., Johnson, C. A. and Friesen, M. C. (2016). Computer-based coding of free-text job descriptions to efficiently identify occupations in epidemiological studies, *Occupational and Environmental Medicine* **73**(6): 417–424.
URL: *https://oem.bmj.com/content/73/6/417*

Russ, D. E., Ho, K.-Y., Johnson, C. A. and Friesen, M. C. (2014). Computer-based coding of occupation codes for epidemiological analyses, *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pp. 347–350.

Schierholz, M. (2014). Automating survey coding for occupation, *FDZ-Methodenreport 10/2014*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Schierholz, M., Brenner, L., Cohausz, L., Damminger, L., Fast, L., Hörig, A.-K., Huber, A.-L., Ludwig, T., Petry, A. and Tschischka, L. (2018). Eine Hilfsklassifikation mit Tätigkeitsbeschreibungen für Zwecke der Berufskodierung * Leitgedanken und Dokumentation, *IAB-Discussion Paper 13/2018*, Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Schierholz, M., Gensicke, M., Tschersich, N. and Kreuter, F. (2018). Occupation coding during the interview, *Journal of the Royal Statistical Society: Series A* **181**(2): 379–407.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview, *Neural Networks* **61**: 85–117.

Scism, L. (2016). Car insurance firms could be banned from asking what you do for a living, *The Wall Street Journal* .
URL: *https://www.wsj.com/articles/car-insurance-firms-could-be-banned-from-asking-what-you-do-for-a-living-1479308820*

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization, *ACM Comput. Surv.* **34**(1): 1–47.
URL: *http://doi.acm.org/10.1145/505282.505283*

Speizer, H. and Buckley, P. (1998). Automated coding of survey data, *in* M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls II and J. M. O'Reilly (eds), *Computer Assisted Survey Information Collection*, Wiley, New York, pp. pp. 223–243.

Takahashi, K., Takamura, H. and Okumura, M. (2005). Automatic Occupation Coding with Combination of Machine Learning and Hand-Crafted Rules, *in* T. B. Ho, D. Cheung and H. Liu (eds), *PAKDD 2005. LNCS, vol. 3518*, Springer, Berlin, pp. 269–279.

Takahashi, K., Taki, H., Tanabe, S. and Li, W. (2014). An Automatic Coding System with a Three-Grade Confidence Level Corresponding to the National/International Occupation and Industry Standard, *Proceedings of the International Conference on Knowledge Engineering and Ontology Development - Volume 1: KEOD, (IC3K 2014)*, INSTICC, SciTePress, pp. 369–375.

Thompson, M., Kornbau, M. E. and Vesely, J. (2014). Creating an automated industry and occupation coding process for the American Community

Survey, *Background Material for a meeting of the Federal Economic Statistics Advisory Committee*, U.S. Census Bureau, Suitland. (Available from http://www.census.gov/about/adrm/fesac/meetings/june-13-2014-meeting.html.).

Tijdens, K., Zijl, S. v., Hughie-Williams, M., Kaveren, M. v. and Steinmetz, S. (2010). Codebook and explanatory note on the WageIndicator dataset, *AIAS Working Paper 10-102*, Universiteit van Amsterdam, Amsterdam.

Trappmann, M., Beste, J., Bethmann, A. and Müller, G. (2013). The pass panel survey after six waves, *Journal for Labour Market Research* **46**(4): 275–281.
URL: *https://doi.org/10.1007/s12651-013-0150-1*

van der Loo, M. (2014). The stringdist package for approximate string matching, *The R Journal* **6**: 111–122.
URL: *https://CRAN.R-project.org/package=stringdist*

Wenzowski, M. (1988). Actr - a generalized automated coding system, *Survey Methodology* **14**(2): 299–307.

Westermark, M., Franzen, M. and Kraft, K. (2015). Automatic Coding of Occupation Using Spell Checking and Machine Learning, *30th JOS Anniversary Conference*.

Appendix to: *Machine Learning for Occupation Coding - A Comparison Study*

Malte Schierholz
Institute for Employment Research, Nuremberg, Germany

Contents

1 Part A: Study Descriptions and Data **3**
1.1 Question Wording 7

2 Part B: Evaluation Metrics **9**
2.1 Production Rates and Agreement Rates 9
2.2 Probabilistic Forecast Evaluation: Calibration, Sharpness, and Scoring Rules 10

3 Part C: Model Tuning **13**
3.1 Logistic Regression 13
3.2 Adapted Nearest Neighbor (Gweon et al. 2017) 19
3.3 Memory-based Reasoning (Creecy et al. 1992) 22
3.4 Tree Boosting (XGBoost) 22
3.5 Similarity-based reasoning 30

4 Part D: Detailed Results **33**
4.1 Agreement Rate Among Top 1 vs. Production Rate (Automated Coding) 33
4.2 Reliability Diagrams (k = 1) 38
4.3 Agreement Rate Among Top 5 vs. Production Rate (Computer-Assisted Coding) 48
4.4 True Positives among Top 5 vs. False Positives (Computer-Assisted Coding) 53
4.5 Reliability Diagrams (k = 5) 58
4.6 Sharpness 68
4.7 Logloss 69

5 Part E: Similarity-based Reasoning: Connecting Approximate String Matching and a Hierarchical Bayesian Model **70**
5.1 A Bayesian hierarchical model if verbal answers and entries from the coding index are identical 70
5.2 Balancing evidence from the coding index and evidence from training data 74

5.3 Verbal answers and entries from the coding index are similar 75
5.4 Derived formulas 78

6 References **83**

1 Part A: Study Descriptions and Data

Data set A.
The survey '*Working and learning in a changing world*' was conducted by the German Institute for Employment Research (IAB) to study the pathways how informal competencies and knowledge support professional careers. A clustered sample of persons living in Germany and born between 1956 and 1988 was surveyed in 2007/2008 with computer-assisted telephone interviews (Antoni et al., 2010). Among other themes, the questionnaire contained questions about the employment biography, i.e., about all the jobs that each person has held during her lifetime. Two questions were asked to collect detailed information about each job. The first question was about the 'berufliche Tätigkeit' (occupational activity) and the second question asked for a more precise activity description. For the first question, a total of 32,931 verbal answers from 9,230 persons is available. For the second question, many respondents were less motivated and their answers are either identical to their first answer or not meaningful for coding (e.g., "don't know"). After replacing such non-informative answers with empty strings, we are left with 19,330 answers for the second question. For both text fields, the inputs were restricted to have at most 50 characters and additional text beyond this threshold was discarded.

Drasch et al. (2012) described the subsequent coding process: Instead of using the KldB 2010 for coding, answers were assigned to the *Dokumentationskennziffer*, an internal list of job titles that is continuously updated by the German Federal Employment Agency. Transition tables are available to convert the assigned job titles into classification codes. To select appropriate job titles from the *Dokumentationskennziffer*, a three-step process was developed, involving (1) automatic coding, (2) manual coding by specially trained coders and (3) manual coding by supervisors. For automatic coding, the respondents' verbal answers were compared with the job titles from the *Dokumentationskennziffer* and with another database of search words. A job title was selected automatically if an entry was identical with the verbal answer, applicable for 39% of the answers. Human coders were involved in this automatic process only seldom, that is, only when the automatic text matching suggested more than a single job title and a human decision was required. The residual answers (61%) were coded by human coders and their supervisors who were trained to follow a set of rules. Their task was to enter the verbal answers from respondents in a web mask and to choose an appropriate job title from a list of suggestions. Additional answers from respondents about each job were available to coders and, as a fundamental rule, coders were required to select only job titles, if the job titles' usual educational requirement level is appropriate when compared to the differentiated occupational status as reported by the respondents. Final quality assurance showed that the inter-coder reliability of manual coding is 50% for job titles and improves when these job titles are grouped together into higher-level units of occupational classifications (65% inter-coder reliability for the 4-digit KldB 2010).

Data set B.
The '*BIBB/BAuA Employment Survey 2012*' was conducted by the German Federal Institute for Vocational Education and Training (BIBB) in cooperation with the Ger-

man Federal Institute for Occupational Safety and Health (BAuA) to collect data about employees and workplaces, focusing on working conditions and qualifications. The population consisted of persons aged 15 or older in paid employment for at least ten hours per week. Random digit dialing was used to select respondents and, if a short screening interview proved their eligibility, their data were collected in computer-assisted telephone interviews. 20.036 persons participated in the survey (Rohrbach-Schmidt and Hall, 2013, Hall et al., 2015). A single open-ended question asked for a 'Tätigkeitsbezeichnung' (occupational activity title). Implementing an automatic filtering mechanism, the answer was then compared with a list of general job titles to decide if asking for further specification with a second open-ended question was necessary. Respondents did not only answer a question about their current occupation (20.031 verbal answers), but also about the first occupation they had in their life (17.379 answers) and about the occupation of a parent when they were 15 years old (16.130 answers about fathers and 2.404 answers about mothers), yielding a total of 55.944 answers that we pool for our analysis.

Hartmann et al. (2012) described the rules and procedures applied for coding the verbal answers into the 2010 GCO. If possible, answers were coded automatically using a coding index or other hand-crafted rules, which are based on information from other survey variables. Answers that were not codable automatically were coded manually by professionals, who followed a set of general and specific rules. Quality measures of the coding process are not published.

In addition to the categories from the 2010 GCO, 22 special categories were used for coding. 18 of those special codes (e.g., 'Technician', 'Engineer', 'Kaufmann (=Merchant / Business Administration)', ...) comprising 1.346 verbal answers are not directly translatable to any of the special codes that we use for our analysis. In order to still obtain comparable coding schemes among our data sets, we aggregate those codes and move the answers into the special category 'Job description not precise enough / Uncodeable'. This recoding procedure partly explains the high proportion of uncodables in this data set.

Data set C.

The '*German Health Update 2014/2015*', harmonized with the '*European Health Interview Survey*' (GEDA 2014/2015-EHIS), was carried out by the German Robert Koch Institute (RKI) to obtain representative statistics related to public health. The population consisted of all persons aged 15 or older with permanent residence in Germany. Based on a two-stage stratified cluster sampling approach using official population registries, selected residents were invited per mail to fill out self-administered questionnaires, either online or paper-based. In total, 24.824 individuals responded, for a response rate (AAPOR RR 1) of 27.6% (Lange et al., 2017). Persons who were employed at the time of the interview were asked in detail about their current occupation.

A related study, the '*German Health Interview and Examination Survey for Children and Adolescents*' (KiGGS) was designed to monitor public health in the next generation. We use data from wave two, conducted between September 2014 and June 2017 by the Robert Koch Institute (RKI), which has both a cross-sectional and a longitudinal component. The population from the cross-sectional component consists of all children aged 0 to 17 with permanent residence in Germany, sampled with a two-stage

stratified cluster approach. 15.023 children participated, for a response rate of 40.1% (AAPOR RR 2) (Hoffmann et al., 2018). The longitudinal component is a follow-up of a similar study that took place between 2003 and 2006. In wave two, grown-up children of age 10 to 31 were recontacted and invited to participate again. 10.853 children participated (61.5% of the baseline study). Data for both components were collected in temporary medical examination centers and with self-administered questionnaires. As the unit of interest is a child, whole families were actually involved if children were younger than the legal age (18). Most families and grown-up children received a printed version of the questionnaire, but a few children aged 18 and older were also invited to participate in a web survey (Lange et al., 2018). For children aged 17 and younger, both parents were asked about their respective current occupations in a background questionnaire. Children aged 18 and older were asked about their own current occupation (Mauz et al., 2017).

Both studies used (mostly) self-administered paper questionnaires with question wordings that are nearly identical regarding the occupational questions. All currently employed persons were asked for a precise job title in a first open-ended question. A second open-ended questions was then asked to obtain further details about the occupational activity. From the German Health Update (GEDA) we obtain verbal answers from 14.143 currently employed respondents and from the KiGGS we obtain current occupations from 15.425 fathers, 15.507 occupations of mothers, and 3.919 occupations of children aged 18 to 31. Since a single coding unit at the Robert Koch Institute coded all answers from both studies and the data are also in other relevant aspects quite similar, we pool verbal job descriptions from both studies, yielding a total of 48.994 answers.

Albrecht et al. (2017) described the coding process and its evaluation. The first step necessary for coding was to capture the input texts from paper questionnaires in a digital format. In a second step, answers were coded according to the 2010 GCO. This was done automatically using a coding index whenever possible (approx. 20% of all answers). Remaining cases were coded manually by trained coders. A computer-assisted coding software was developed to assist coders and supervisors. Regular meetings between coders and supervisors took place to discuss difficult decisions and to improve the process. To evaluate the quality of coding, approx. 20% of the answers from the German Health Update (GEDA) were double-coded by a second independent coder and, subsequently, an expert decided which of the two codes was more appropriate. For 77.7% of the cases both codes were equally correct (either being identical or according to expert judgment), for 6.7% of the cases only the first code was considered correct, for 3.4% the first code was better, but the second code was considered possible, for 4.1% the second code was better, but the first code was considered possible, and for 8.1% only the second code was considered correct, implying that the decision from the first coder was wrong.

Data set D.

The study '*Labour Market and Social Security*' is an ongoing panel study at the German Institute for Employment Research (IAB), designed to analyze long-term unemployment and accompanying welfare benefits in Germany (Trappmann et al., 2013). Computer-assisted personal interviews and telephone interviews are employed for data

collection. For our study, we use all job descriptions that were collected during the 10th wave of the panel, carried out in 2016 (Berg et al., 2017). Since the occupation is often already known from previous waves, not everyone was asked about it. 4309 respondents, some of them living together in the same households, replied to occupational questions. The extensive questionnaire contained filter questions to ensure that respondents were asked only questions relevant to them. Occupational questions about the current and all recent occupations since 2014 (2411 answers and additional 64 answers from elderly people), about current minijobs (1365 answers), about the respondents' first and last job in their lifetime (938/449 answers) and about the occupations of mothers and fathers when the respondent was 15 years old (846/1566 answers) were asked, yielding a total of 7639 occupational descriptions that we pool for our analysis. The question wording was about the 'berufliche Tätigkeit' (occupational activity). Interviewers were instructed to probe for a job title when the first answer was not precise enough. Unlike other studies used here, the verbal answers were recorded in a single text field.

A team of trained coders coded the verbal answers into the KldB 2010. Coders had access to answers from additional questions providing background information about the respondents' jobs. For assistance, code suggestions were sometimes available from an automatic textual comparison with a coding index. Two coders worked independently on each answer. The second coder did not know the code from the first coder, but an automatic indicator of agreement was available, such that the second coder could reconsider and correct his decision. The agreement rates between both coders range from 67.2% to 80.2%, depending on the specific persons involved. If no agreement was reached, the final decision was made by a third person or, for the most difficult cases, in a group discussion. Only the final codes from this process are used in our analysis.

Data set E.

The survey '*Selectivity effects in address handling*' was conducted in 2014 by the Institute for Employment Research (IAB) to analyze a number of methodological issues. Individuals were drawn at random from a German federal database used in the social security administration. Valid computer-assisted telephone interviews were conducted with 1.208 persons (Sakshaug et al., 2016). Respondents were asked about their current occupation or—for the unemployed—about their most recent occupation and, as a result, verbal answers on occupation were collected from 1.064 individuals.

A central element of the study was testing a new instrument for occupation coding, which applied supervised learning techniques to predict possible job titles at the time of the interview. The most probable job titles were suggested to the respondents who in turn could choose the most appropriate occupation. To evaluate the new instrument and assess its quality, all answers were coded into the KldB 2010 by two independent coding professionals. They were given access to several occupation-related background variables. Only codes from one of the two coders are used in our subsequent analysis. To obtain those codes, the professional coder used automatic coding as a first step, that is, the verbal answers were compared textually with a coding index. The correctness of the so-found codes was checked manually. In a second step, leftover cases were coded manually, hereby following a few priority rules concerning how to interpret answers

with conflicting information and how to code ambiguous cases. A separate indicator variable is part of the data, stating for three observations that the "level of qualification [is] unknown, lowest is coded" and for another 19 observations that "multiple codes [are] possible, decision made". 1042 observations have no such cautionary note in the data. The quality from this coder is as follows: The second professional coder, who worked independently from the first and applied different coding procedures and rules, obtains for 60.7% of the answers identical 5-digit KldB codes. Moreover, the quality was assessed by two student assistants, who were given access to the complete information and all available codes. Their respective findings are that 89.6%/93.2% of the assigned codes are "acceptable" (different codes were allowed to be acceptable for the same individual), 6.4%/1.1% are "uncertain" and 4.0%/5.7% of the codes are considered "wrong". Schierholz et al. (2018) described the study and the complete evaluation in more detail.

1.1 Question Wording

The general recommendation to collect occupational data in Germany is to ask three open-ended questions, the first asking for a 'berufliche Tätigkeit' (occupational activity), the second question should ask for a more precise occupational activity description, and the third question should ask if this occupation has yet another job title (Statistisches Bundesamt, 2016). No study that we use fully implemented this time-demanding protocol requiring three questions, but all studies ask for either a job title, or an occupational activity, or both. Some questionnaires explicitly ask respondents to provide a precise answer, whereas others mention this in the instructions given to the interviewers.

The following surveys have published their questionnaires on-line:

- Working and learning in a changing world (data set A): `http://doku.iab.de/fdz/reporte/2010/DR_02-10.pdf`
- BIBB/BAuA Employment Survey 2012 (data set B): `https://metadaten.bibb.de/download/732`
- Labour Market and Social Security (data set D)[1]: `https://fdz.iab.de/de/FDZ_Individual_Data/PASS/PASS-SUF0617.aspx`
- Selectivity effects in address handling (data set E): `https://rss.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111\%2Frssa.12297&attachmentId=2205795635`

The questionnaires of the '*German Health Update 2014/2015*' and of the '*German Health Interview and Examination Survey for Children and Adolescents*' (data set C) are not available on-line. We report the question wording of the two open-ended questions in the following. Additional job-related variables were available for coding.

[1] Answers from both open-ended questions are written down in a single text field. The second question is only asked if the first question is not precise enough.

What is your current occupation?
Please specify the precise job title, not the vocational degree or the rank.
For example:

- Flower seller (not seller)
- Bricklayer (not construction worker)
- Management consultant (not business administration graduate)

To facilitate the classification of your job, please add further explanations in keywords.
For example:

- Customer service, sales, packaging of plants (as a flower seller)
- Customs investigation, operational planning, public relations (as a customs officer)
- Maintenance, repair, equipment of vehicles, workshop management (as an automotive mechanic)

If you have management responsibilities, please mention this.

Question Wording in the '*German Health Update 2014/2015*' and in the '*German Health Interview and Examination Survey for Children and Adolescents*'

2 Part B: Evaluation Metrics

To describe the metrics, we first clarify our notation. $y_{n_f} = (y_{n_f 1}, ..., y_{n_f K})$ is a 0-1-vector (the element at the k-th position equals 1 and all other elements are 0), indicating the 'true' category of individual $n_f = 1, ..., N_f$ from the test data. If an algorithm outputs probabilities, the probabilities predicted for individual n_f are denoted by $\hat{p}_{n_f} = (\hat{p}_{n_f 1}, ..., \hat{p}_{n_f K})$. The predicted probabilities can be transformed into category suggestions $\hat{y}_{n_f}^{(m)} = (\hat{y}_{n_f 1}, ..., \hat{y}_{n_f K})$ by setting the m highest-ranked elements from $\hat{p}_{n_f}$ to 1 (m is a predefined threshold). Using observed categories y_{n_f} and suggested categories $\hat{y}_{n_f}^{(m)}$, we create a scalar indicator of agreement. $a_{n_f}^{(m)} := \hat{y}_{n_f}^{(m)} y_{n_f}^T$ is 1 for individual n_f only iff the 'true' category is among the m highest-ranked categories. The predicted probability that the 'true' category is among the top-m-suggestions, $\hat{p}_{n_f}^{(m)}$, is calculated as a sum over the m most probable categories, $\hat{p}_{n_f}^{(m)} := \hat{y}_{n_f}^{(m)} \hat{p}_{n_f}^T$. In the main text we use $a_{n_f}^{(1)}$ and $\hat{p}_{n_f}^{(1)}$ to calculate agreement rates among top 1 and $a_{n_f}^{(5)}$ and $\hat{p}_{n_f}^{(5)}$ to calculate agreement rates among top 5.

2.1 Production Rates and Agreement Rates

Production rates and agreement rates are described in the main text. For completeness we provide the mathematical definitions here. m and t are two values a user needs to select based on their particular application. For example, we would use $m = 1$ and $t = 0$ to obtain measures of performance for fully automated coding at 100% production rate. *Agreement rate*$_0^{(1)}$ is sometimes also called the *accuracy*.

The number of individuals that will be coded automatically and in agreement with the test data is called the

$$\textit{Number of true positives}_t^{(m)} = \sum_{n_f : \hat{p}_{n_f}^{(m)} > t} a_{n_f}^{(m)}$$

The number of individuals that will be coded automatically but not in agreement with the test data is called the

$$\textit{Number of false positives}_t^{(m)} = \sum_{n_f : \hat{p}_{n_f}^{(m)} > t} 1 - a_{n_f}^{(m)}$$

The proportion of cases that go into automatic coding is called the

$$\textit{Production rate}_t^{(m)} = \frac{\text{Number of true positives}_t^{(m)} + \text{Number of false positives}_t^{(m)}}{N_f}$$

The proportion of successes among all automatic codings is called the

$$\textit{Agreement rate}_t^{(m)} = \frac{\text{Number of true positives}_t^{(m)}}{\text{Number of true positives}_t^{(m)} + \text{Number of false positives}_t^{(m)}}$$

Our use of production rates and agreement rates is inspired by Chen et al. (1993), who used similar definitions to calculate separate thresholds for each category.

Agreement rates and production rates are meaningful and easily interpretable terms in our context. Yet, we caution that the agreement rate is an estimator having high variance at low production rates, making it prone to over-interpretation. In addition, since the agreement rate averages over easier-to-code and harder-to-code individuals, this number obscures for the harder-to-code individuals how successful an automated coding procedure is and if it is useful at all.

For this reason we prefer for our own inspection sometimes a different type of presentation, plotting the number of true positives versus the number of false positives as the threshold to use automatic coding increases. Formally, one can calculate the Number of true positives$_t^{(m)}$ and the Number of false positives$_t^{(m)}$ if only the Production rate$_t^{(m)}$ and the Agreement rate$_t^{(m)}$ are known (a bijective transformation exists between both diagrams).

$$\text{Number of true positives}_t^{(m)} = N_f \cdot \text{Production rate}_t^{(m)} \cdot \text{Agreement rate}_t^{(m)}$$
$$\text{Number of false positives}_t^{(m)} = N_f \cdot \text{Production rate}_t^{(m)} \cdot (1 - \text{Agreement rate}_t^{(m)})$$

Thus, both diagrams are equivalent in a mathematical sense. Yet, both diagrams have their role and we will depict a few diagrams in Section 4.4 so that readers get an impression which type of diagram is more useful for their own purposes.

2.2 Probabilistic Forecast Evaluation: Calibration, Sharpness, and Scoring Rules

We evaluate probabilistic predictions with respect to three criteria: (1) calibration (= reliability), (2) sharpness, and (3) log loss. Calibration refers to the desired attribute that observed relative frequencies should realize as often as the predicted probabilities suggest. Sharpness refers to the degree of certainty in predictions about future values. Log loss is an overall score, incorporating both calibration and sharpness, that has been shown to be ideal under certain reasonable conditions. Jollife and Stephenson (2012) provide an introduction to (probabilistic) forecast verification.

We assess *calibration* using reliability diagrams that compare predicted probabilities with their observed counterparts (Hsu and Murphy, 1986, Bröcker and Smith, 2007). Usually, reliability diagrams are used for binary outcomes. Since our application has K outcome categories, we extend this as follows. The probabilities to be accurate among top m, $\hat{p}_{n_f}^{(m)}$, are partitioned into ten equal-length segments $[0, 0.1), ..., [0.9, 1]$. We then average the probabilities within each segment and plot them against their observed averaged accuracies $a_{n_f}^{(m)}$. We choose $m = 1$ and $m = 5$ since these values appear meaningful for automated and computer-assisted coding, respectively.

Sharpness is another desired property of good predictions. The idea is that with little background information the forecast about a future event should be rather vague, i.e., the forecast should allow for many different outcomes, but as more background

information becomes available (e.g., more information about a respondent's job), predictions should become sharper, i.e., the certainty about the outcome should increase. As a result, one can obtain better and worse forecasts, all being calibrated, which means that calibration alone is not sufficient to evaluate probabilistic forecasts. Gneiting et al. (2007) argued in the context of forecasting continuous variables that the optimal forecast should maximize sharpness subject to calibration. Ideally, we would like that probabilistic predictions are certain, i.e., they should predict an outcome with probability 1.

Following Murphy and Epstein (1967) and Potts (2012), we use the information-theoretic measure of entropy, H, and average over it to determine sharpness $\bar{H}$,

$$\bar{H} = \frac{1}{N_f} \sum_{n_f=1}^{N_f} H_{n_f} = \frac{1}{N_f} \sum_{n_f=1}^{N_f} (-\sum_{k=1}^{K} \hat{p}_{n_f k} \log_2 \hat{p}_{n_f k}) \tag{1}$$

If $\hat{p}_{n_f k} = 0$, the term $\hat{p}_{n_f k} \log_2 \hat{p}_{n_f k}$ is not defined and, as a workaround, we will set it to zero.

The minimal value is zero (optimal sharpness) if all forecasts predict a single outcome with probability 1. The maximum (no sharpness) occurs if all categories have equal probability, $\hat{p}_{n_f k} = \frac{1}{K}$ for all individuals and all k. Note that sharpness is a property of the predictive distribution only and is not related to values that will eventually realize. Each component H_{n_f} may be interpreted as the average information content, i.e., the expected minimal number of bits that would need to be transferred between two persons if both knew $\hat{p}_{n_f}$ and one person wants to inform the other about an expected realization drawn from this distribution. To acknowledge sample variation in the test data, we calculate standard errors using the formula $\sigma(\bar{H}) = \sqrt{\frac{1}{N_f} \hat{\mathbb{V}ar}(H_{n_f})} = \sqrt{\frac{1}{N_f(N_f-1)} \sum_{n_f=1}^{N_f} (H_{n_f} - \bar{H})^2}$

A general score to evaluate probabilistic predictions is the *log loss* (or ignorance score). It is the average negative $\log$ probability of the categories that will actually realize in the test data,

$$\log_2 \text{loss} = \frac{1}{N_f} \sum_{n_f=1}^{N_f} \log_2 \text{loss}_{n_f} = \frac{1}{N_f} \sum_{n_f=1}^{N_f} \sum_{k=1}^{K} -y_{n_f k} \log_2 \hat{p}_{n_f k} \tag{2}$$

The optimal $\log_2$ loss equals 0 and occurs if all observed categories $y_{n_f k}$ have been predicted correctly and with probability one. The worst case is ∞, which happens if at least one category realizes although it was predicted with probability zero. To acknowledge sample variation in the test data, we calculate standard errors using the formula $\sigma(\log_2 \text{loss}) = \sqrt{\frac{1}{N_f(N_f-1)} \sum_{n_f=1}^{N_f} (\log_2 \text{loss}_{n_f} - \log_2 \text{loss})^2}$.

Scoring rules like the logarithmic loss have been developed in the context to evaluate an expert's probabilistic judgment. If his prediction is in line with the reality as observed later, the expert should be rewarded, otherwise he should get punished. A good scoring rule should ensure that the expert thinks carefully about his judgment and answers as best as he can. In our case, we regard the algorithms' predicted probabilities $\hat{p}$ as expert judgments. The scoring rule $\log \hat{p}_{n_f k}$ is, apart from linear transformations,

the *unique* scoring rule that is proper, symmetric and impartial (O'Hagan and Forster, 2004, 55ff.), three conditions that are desirable for our application. A scoring rule is called proper, if it encourages the expert to report his actual beliefs. It is symmetric if the rule induces no preference for a specific outcome category that might realize. It is impartial if calculations for the final reward are only based on the expert's prediction of the category that actually realizes and on no other categories. We use the linear transformation $-\log_2 \hat{p}_{n_f k} = -\log \hat{p}_{n_f k} / \log 2$ because it has an information-theoretic interpretation: It is the minimal number of bits that need to be transferred between two persons if both knew $\hat{p}_{n_f k}$ and one person wants to inform the other about a realized outcome y_{n_f} (Roulston and Smith, 2002).

3 Part C: Model Tuning

The results from any algorithm depend on the choice of tuning parameters. This appendix demonstrates the performance of different parameter configurations and explores the importance of the parameters. For each algorithm, we need to select a final parameter configuration that we use to report the results in the main text. For this purpose, a grid search was performed and the best configuration among the ones tested was chosen. In general, we will see in this appendix that the performances of different parameter configurations are often very similar to another. This suggests that our chosen parameters are close to optimal and a more profound grid search over additional values will not improve the predictions in a practically meaningful way.

3.1 Logistic Regression

The following parameters are varied:

- whether to exclude stopwords or not (while creating the document-term matrix);
- whether to use stemming or not (while creating the document-term matrix);
- whether to add an additional column to the predictor matrix counting the number of words in a document;
- choice of the regularization parameter λ (see below);
- choice of the elastic-net regularization parameter α on a grid $(0, 0.05, 0.2)$ (see below).

The default in the `glmnet`-package is to estimate a sequence of models for 100 values of λ using a highly efficient cyclical coordinate descent algorithm. The largest value of λ is chosen such that all coefficients in the linear term are shrunk to zero. The smallest value of λ is 0.0001 times the largest value. A regular grid (equidistant on a logarithmic scale of λ) is spanned in between and we report predictions for values of λ that are at the 50-th, 70-th, 80-th, and 100-th position in this grid.

In our experience with occupational data, the algorithm in `glmnet` often does not converge during a limited number of iterations if values of α are substantially larger than zero. To counter this problem, we increase the maximum number of iterations `maxit` to 10^6 (the default is 10^5) and reduce the convergence threshold `thresh` to 10^{-4} (the default is 10^{-7}). If the threshold is still not reached within the maximum number of iterations, we report the performance for the smallest value of λ that is available. Default-values are used for all other parameters in the `glmnet`-package. The low threshold can potentially have a negative impact on model performance. To rule out this possibility, the $\texttt{thresh} = 10^{-4}$ is only used in this appendix to select the best-performing parameter configuration. The threshold 10^{-7} was used to estimate all models in the main paper.

What are good tuning parameters to use? Table A1 and Table A2 show the predictive performance for various combinations of α and λ in data set D (small) and data set C (larger). Both tables indicate that the choice $\alpha = 0.05$ is close to optimal

with respect to the criteria accuracy, logloss and sharpness. In Tables A3 and A4 we show more results for various parameter configurations from both data sets when α is fixed at 0.05. We observe that accuracy is almost constant, sharpness improves as λ decreases, and logloss is optimal at mean values of λ. No meaningful differences are found whether or not stopwords and counting of words are used, a result that also holds for other choices of α (not shown). The performance appears to improve marginally if the German Porter stemmer is used.

Based on these results, almost all models in the main paper are estimated with $\alpha = 0.05$, $maxit = 10^6$, $thresh = 10^{-7}$, with stemming, without excluding stopwords and without counting words. We look at λ-values at the 50-th, 70-th, 80-th, and 100-th position in the provided grid and find in all data sets that the λ-value at the 70-th position is close to optimal, minimizing logloss. An exception is the model that we trained on the complete data set (A,B,C,D). Due to the large numbers of outcome categories and predictor variables, computational limitations forced us to build a grid of 60 values of λ and the algorithm stopped with an error (convergence not reached) at the 26-th position. The 25-th λ-value is thus reported, although smaller values of λ, if calculable, would certainly improve the results.

The role of λ is better understood in the context of a reliability diagram as shown in Figure A1. If λ is too large, categories are predicted for many observations with a probability that is too low. Thus, the predicted categories will realize more often than expected (underfitting with $\lambda = 0.02499$). Alternatively, if λ is too small, for many observations a single category is predicted with high probability (overfitting with $\lambda = 0.00024$). Overfitting improves sharpness, which would be optimal if all probabilities are either exactly one or zero. However, one would like that forecasted probabilities reflect the observed relative frequencies in the test data. This requires a value for λ that is neither too large nor too small.

Table A1: Performance measures of logistic regression in data set D without stopwords, without stemming and without counting words. Selected combinations of α and λ are shown.†

α	λ	$accuracy$	$logloss$	$sharpness$
0.00	1.7540	0.04 ± 0.01	7.12 ± 0.06	7.65 ± 0.00
0.00	0.2729	0.43 ± 0.02	5.17 ± 0.09	7.43 ± 0.02
0.00	0.1076	0.56 ± 0.02	4.07 ± 0.11	6.52 ± 0.04
0.00	0.0167	0.59 ± 0.02	3.30 ± 0.13	4.14 ± 0.08
0.05	0.0351	0.58 ± 0.02	3.55 ± 0.12	5.27 ± 0.07
0.05	0.0055	0.59 ± 0.02	3.26 ± 0.13	3.37 ± 0.09
0.05	0.0022	0.59 ± 0.02	3.29 ± 0.14	2.84 ± 0.09
0.05	0.0003	0.59 ± 0.02	3.54 ± 0.16	2.19 ± 0.09
0.20††	0.0244	0.03 ± 0.00	147.91 ± 0.99	0.00 ± 0.00

†Data set D was split into 6575 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on a subset of 991 test observations.
††With $\alpha = 0.20$ results for $\lambda < 0.0244$ are not available since the algorithm did not converge after 10^6 iterations.

Table A2: Performance measures of logistic regression in data set C without stopwords, without stemming and with counting words. Selected combinations of α and λ are shown.†

α	λ	$accuracy$	$logloss$	$sharpness$
0.00	1.2495	0.11 ± 0.01	7.86 ± 0.07	8.45 ± 0.00
0.00	0.1944	0.51 ± 0.02	5.61 ± 0.10	8.05 ± 0.03
0.00	0.0767	0.58 ± 0.02	4.41 ± 0.11	6.76 ± 0.05
0.00	0.0119	0.60 ± 0.02	3.57 ± 0.13	3.95 ± 0.07
0.05	0.0250	0.59 ± 0.02	3.95 ± 0.12	5.53 ± 0.07
0.05	0.0039	0.59 ± 0.02	3.50 ± 0.14	3.23 ± 0.08
0.05	0.0015	0.59 ± 0.02	3.56 ± 0.15	2.61 ± 0.07
0.05	0.0002	0.59 ± 0.02	3.92 ± 0.17	1.92 ± 0.07
0.20††	0.0132	0.58 ± 0.02	4.03 ± 0.13	5.47 ± 0.07

†Data set C was split into 47,930 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on a subset of 1061 test observations.
††With $\alpha = 0.20$ results for $\lambda < 0.0132$ are not available since the algorithm did not converge after 10^6 iterations.

Table A3: Performance measures of logistic regression in data set D with $\alpha = 0.05$. Selected λ and all combinations of stopwords, stemming and word counting are shown.†

λ	*stem*	*stop*	*count*	*accuracy*	*logloss*	*sharpness*
With threshold $= 10^{-4}$*:*						
0.0055	NO	YES	YES	0.59 ± 0.02	3.26 ± 0.13	3.37 ± 0.09
0.0055	NO	NO	YES	0.59 ± 0.02	3.26 ± 0.13	3.38 ± 0.09
0.0055	NO	YES	NO	0.59 ± 0.02	3.26 ± 0.13	3.37 ± 0.09
0.0055	NO	NO	NO	0.59 ± 0.02	3.26 ± 0.13	3.37 ± 0.09
0.0055	YES	YES	YES	0.60 ± 0.02	3.15 ± 0.13	3.23 ± 0.09
0.0055	YES	NO	YES	0.60 ± 0.01	3.15 ± 0.13	3.23 ± 0.09
0.0055	YES	YES	NO	0.60 ± 0.01	3.15 ± 0.13	3.22 ± 0.09
0.0055	YES	NO	NO	0.60 ± 0.01	3.15 ± 0.13	3.23 ± 0.09
0.0022	NO	YES	YES	0.59 ± 0.02	3.30 ± 0.14	2.84 ± 0.09
0.0022	NO	NO	YES	0.59 ± 0.02	3.29 ± 0.14	2.85 ± 0.09
0.0022	NO	YES	NO	0.59 ± 0.02	3.30 ± 0.14	2.83 ± 0.09
0.0022	NO	NO	NO	0.59 ± 0.02	3.29 ± 0.14	2.84 ± 0.09
0.0022	YES	YES	YES	0.61 ± 0.01	3.21 ± 0.14	2.66 ± 0.09
0.0022	YES	NO	YES	0.60 ± 0.01	3.19 ± 0.14	2.67 ± 0.09
0.0022	YES	YES	NO	0.61 ± 0.01	3.21 ± 0.14	2.66 ± 0.09
0.0022	YES	NO	NO	0.61 ± 0.01	3.20 ± 0.14	2.67 ± 0.09
0.0003	NO	YES	YES	0.59 ± 0.02	3.55 ± 0.16	2.19 ± 0.09
0.0003	NO	NO	YES	0.59 ± 0.02	3.54 ± 0.16	2.20 ± 0.09
0.0003	NO	YES	NO	0.59 ± 0.02	3.55 ± 0.16	2.18 ± 0.09
0.0003	NO	NO	NO	0.59 ± 0.02	3.54 ± 0.16	2.19 ± 0.09
0.0003	YES	YES	YES	0.60 ± 0.02	3.51 ± 0.16	1.98 ± 0.08
0.0003	YES	NO	YES	0.60 ± 0.02	3.48 ± 0.16	2.00 ± 0.09
0.0003	YES	YES	NO	0.60 ± 0.02	3.51 ± 0.16	1.97 ± 0.08
0.0003	YES	NO	NO	0.60 ± 0.02	3.48 ± 0.16	1.99 ± 0.08
With threshold $= 10^{-7}$ *(Improvements expected):*						
0.0055	YES	NO	NO	0.60 ± 0.01	3.15 ± 0.13	3.19 ± 0.09
0.0022	YES	NO	NO	0.61 ± 0.01	3.21 ± 0.14	2.64 ± 0.09
0.0003	YES	NO	NO	0.60 ± 0.02	3.50 ± 0.16	1.99 ± 0.09

†Data set D was split into 6575 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on a subset of 991 test observations.

Table A4: Performance measures of logistic regression in data set C with $\alpha = 0.05$. Selected λ and all combinations of stopwords, stemming and word counting are shown.†

λ	*stem*	*stop*	*count*	*accuracy*	*logloss*	*sharpness*
0.0250	NO	YES	YES	0.59 ± 0.02	3.95 ± 0.12	5.53 ± 0.07
0.0250	NO	NO	YES	0.59 ± 0.02	3.95 ± 0.12	5.53 ± 0.07
0.0250	NO	YES	NO	0.59 ± 0.02	3.95 ± 0.12	5.53 ± 0.07
0.0250	NO	NO	NO	0.59 ± 0.02	3.95 ± 0.12	5.53 ± 0.07
0.0250	YES	YES	YES	0.58 ± 0.02	3.96 ± 0.12	5.58 ± 0.07
0.0250	YES	NO	YES	0.58 ± 0.02	3.96 ± 0.12	5.57 ± 0.07
0.0250	YES	YES	NO	0.58 ± 0.02	3.96 ± 0.12	5.58 ± 0.07
0.0250	YES	NO	NO	0.58 ± 0.02	3.96 ± 0.12	5.57 ± 0.07
0.0039	NO	YES	YES	0.59 ± 0.02	3.50 ± 0.14	3.24 ± 0.08
0.0039	NO	NO	YES	0.59 ± 0.02	3.50 ± 0.14	3.23 ± 0.08
0.0039	NO	YES	NO	0.59 ± 0.02	3.50 ± 0.14	3.24 ± 0.08
0.0039	NO	NO	NO	0.59 ± 0.02	3.50 ± 0.14	3.23 ± 0.08
0.0039	YES	YES	YES	0.59 ± 0.02	3.45 ± 0.13	3.27 ± 0.08
0.0039	YES	NO	YES	0.59 ± 0.02	3.45 ± 0.13	3.27 ± 0.08
0.0039	YES	YES	NO	0.59 ± 0.02	3.45 ± 0.13	3.27 ± 0.08
0.0039	YES	NO	NO	0.59 ± 0.02	3.45 ± 0.13	3.27 ± 0.08
0.0015	NO	YES	YES	0.59 ± 0.02	3.56 ± 0.15	2.61 ± 0.07
0.0015	NO	NO	YES	0.59 ± 0.02	3.56 ± 0.15	2.61 ± 0.07
0.0015	NO	YES	NO	0.59 ± 0.02	3.56 ± 0.15	2.61 ± 0.07
0.0015	NO	NO	NO	0.59 ± 0.02	3.56 ± 0.15	2.61 ± 0.07
0.0015	YES	YES	YES	0.59 ± 0.02	3.49 ± 0.14	2.63 ± 0.07
0.0015	YES	NO	YES	0.59 ± 0.02	3.49 ± 0.14	2.63 ± 0.07
0.0015	YES	YES	NO	0.59 ± 0.02	3.49 ± 0.14	2.63 ± 0.07
0.0015	YES	NO	NO	0.59 ± 0.02	3.49 ± 0.14	2.63 ± 0.07
0.0002	NO	YES	YES	0.59 ± 0.02	3.92 ± 0.17	1.93 ± 0.07
0.0002	NO	NO	YES	0.59 ± 0.02	3.92 ± 0.17	1.92 ± 0.07
0.0002	NO	YES	NO	0.59 ± 0.02	3.93 ± 0.17	1.93 ± 0.07
0.0002	NO	NO	NO	0.59 ± 0.02	3.93 ± 0.17	1.92 ± 0.07
0.0002	YES	YES	YES	0.59 ± 0.02	3.83 ± 0.17	1.93 ± 0.07
0.0002	YES	NO	YES	0.59 ± 0.02	3.84 ± 0.17	1.92 ± 0.07
0.0002	YES	YES	NO	0.59 ± 0.02	3.84 ± 0.17	1.93 ± 0.07
0.0002	YES	NO	NO	0.59 ± 0.02	3.84 ± 0.17	1.92 ± 0.07

†Data set C was split into 47,930 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on a subset of 1061 test observations.

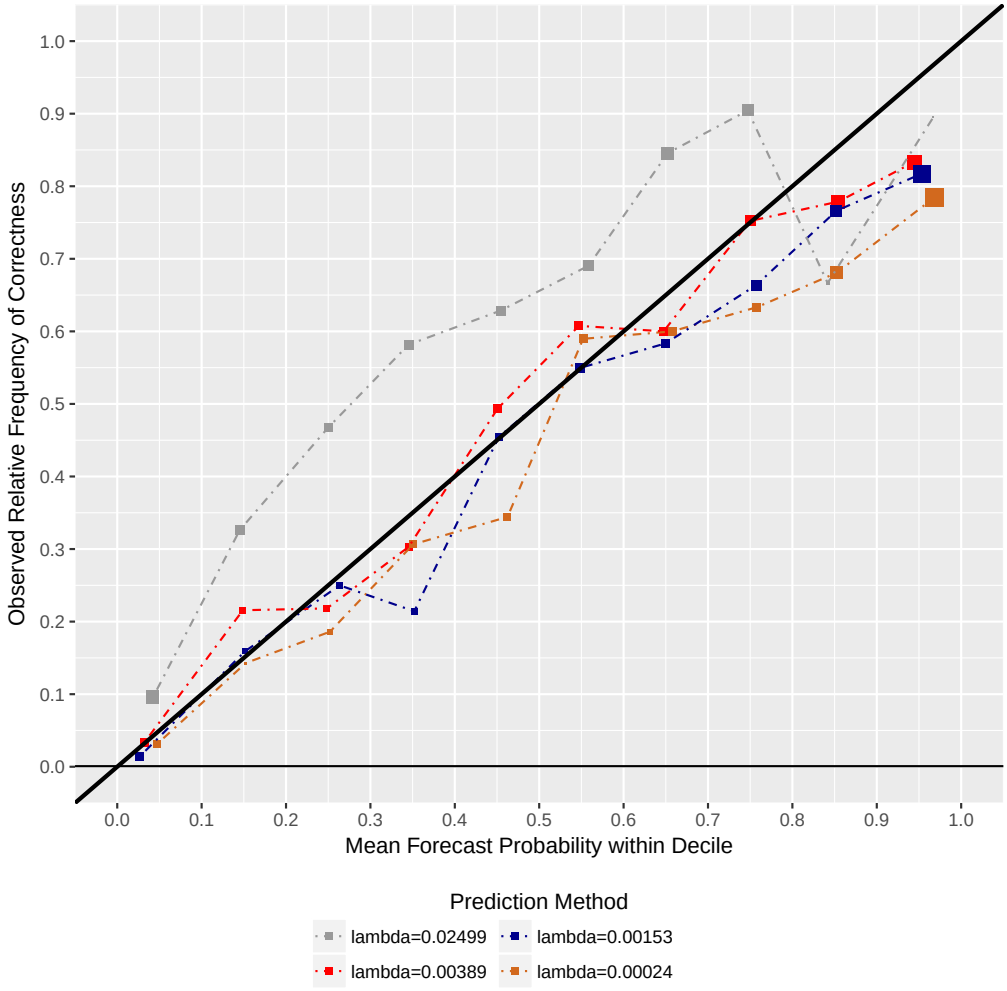


Figure A1: Reliability Diagram of most probable category (k=1); Ideal probabilistic predictions should match the observed relative frequencies, the diagonal; The horizontal line at $\frac{1}{1291}$ represents random guessing among all 1291 categories; Point size is proportional to the number of observations within each bin. Data set C was split into 47,930 training observations and 1064 test observations and a regularized logistic regression with parameters $\alpha = 0.05$, $maxit = 10^6$, $thresh = 10^{-7}$, with stemming, without excluding stopwords and without counting words was estimated.

3.2 Adapted Nearest Neighbor (Gweon et al. 2017)

The following parameters are varied:

- whether to exclude stopwords or not (while creating the document-term matrix);
- whether to use stemming or not (while creating the document-term matrix);
- whether to do extended preprocessing (YES) or just removal of punctuation marks (NO);
- choice of multiplier $mult$.

We can judge the performance of different parameter configurations with respect to accuracy or with respect to N, the number of observations in the test data that had the true category predicted with non-zero relevance score. The logloss itself is less suited for an overall comparison because it is evaluated on subsets of test data that have different size N. Also, the logloss is derived from probabilistic theory, but it is doubtful if a probabilistic framework is appropriate to judge the output from this algorithm.

The different parameter configurations are evaluated on data set D (Table A5) and data set C (Table A6). The highest accuracies and the largest N are achieved if stopwords are removed. Wether stemming or the type of preprocessing have a positive impact is more speculative and additional analyses would be needed. The exact choice of the multiplier $mult$ should not be judged based on this result since it was developed for a different purpose. We still see that it has only a small effect on the evaluation metrics provided here.

This analysis confirms that Gweon et al. (2017) made reasonable choices. We follow them and exclude stopwords, use stemming, remove punctuation marks, and set the multiplier to 0.1 for all analyses in the main paper.

Table A5: Performance measures of Adapted Nearest Neighbor in data set D under different configurations.†

stop	*stem*	*prep*	*mult*	*accuracy*	*logloss* (N)	*sharpness*
NO	NO	NO	0.05	0.50 ± 0.02	2.26 ± 0.14 (722)	1.67 ± 0.10
NO	NO	NO	0.10	0.50 ± 0.02	2.28 ± 0.14 (722)	1.68 ± 0.10
NO	NO	NO	0.20	0.50 ± 0.02	2.33 ± 0.14 (722)	1.70 ± 0.10
NO	NO	YES	0.05	0.51 ± 0.02	2.09 ± 0.14 (718)	1.55 ± 0.09
NO	NO	YES	0.10	0.51 ± 0.02	2.12 ± 0.14 (718)	1.56 ± 0.09
NO	NO	YES	0.20	0.51 ± 0.02	2.17 ± 0.14 (718)	1.58 ± 0.09
NO	YES	NO	0.05	0.51 ± 0.02	2.15 ± 0.14 (722)	1.60 ± 0.09
NO	YES	NO	0.10	0.51 ± 0.02	2.18 ± 0.14 (722)	1.62 ± 0.09
NO	YES	NO	0.20	0.51 ± 0.02	2.22 ± 0.14 (722)	1.64 ± 0.09
NO	YES	YES	0.05	0.52 ± 0.02	2.01 ± 0.13 (719)	1.50 ± 0.09
NO	YES	YES	0.10	0.52 ± 0.02	2.03 ± 0.13 (719)	1.51 ± 0.09
NO	YES	YES	0.20	0.52 ± 0.02	2.08 ± 0.13 (719)	1.53 ± 0.09
YES	NO	NO	0.05	0.56 ± 0.02	2.26 ± 0.14 (783)	1.74 ± 0.10
YES	NO	NO	0.10	0.56 ± 0.02	2.28 ± 0.14 (783)	1.75 ± 0.10
YES	NO	NO	0.20	0.56 ± 0.02	2.32 ± 0.14 (783)	1.77 ± 0.10
YES	NO	YES	0.05	0.56 ± 0.02	2.08 ± 0.13 (768)	1.59 ± 0.10
YES	NO	YES	0.10	0.56 ± 0.02	2.10 ± 0.13 (768)	1.60 ± 0.10
YES	NO	YES	0.20	0.56 ± 0.02	2.15 ± 0.13 (768)	1.62 ± 0.10
YES	YES	NO	0.05	0.58 ± 0.02	2.09 ± 0.13 (784)	1.63 ± 0.10
YES	YES	NO	0.10	0.58 ± 0.02	2.12 ± 0.13 (784)	1.64 ± 0.10
YES	YES	NO	0.20	0.58 ± 0.02	2.16 ± 0.13 (784)	1.66 ± 0.10
YES	YES	YES	0.05	0.58 ± 0.02	1.95 ± 0.13 (772)	1.51 ± 0.09
YES	YES	YES	0.10	0.58 ± 0.02	1.98 ± 0.13 (772)	1.52 ± 0.09
YES	YES	YES	0.20	0.58 ± 0.02	2.02 ± 0.13 (772)	1.54 ± 0.09

†Data set D was split into 6575 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on different, non-comparable subsets of size N.

Table A6: Performance measures of Adapted Nearest Neighbor in data set C under different configurations.†

$stop$	$stem$	$prep$	$mult$	$accuracy$	$logloss\ (N)$	$sharpness$
NO	NO	NO	0.05	0.55 ± 0.02	$1.98 \pm 0.11\ (820)$	1.70 ± 0.08
NO	NO	NO	0.10	0.55 ± 0.02	$1.99 \pm 0.11\ (820)$	1.71 ± 0.08
NO	NO	NO	0.20	0.55 ± 0.02	$2.01 \pm 0.11\ (820)$	1.72 ± 0.08
NO	NO	YES	0.05	0.56 ± 0.02	$1.78 \pm 0.10\ (818)$	1.53 ± 0.07
NO	NO	YES	0.10	0.56 ± 0.02	$1.79 \pm 0.10\ (818)$	1.53 ± 0.07
NO	NO	YES	0.20	0.56 ± 0.02	$1.81 \pm 0.10\ (818)$	1.55 ± 0.07
NO	YES	NO	0.05	0.55 ± 0.02	$1.92 \pm 0.11\ (821)$	1.66 ± 0.08
NO	YES	NO	0.10	0.55 ± 0.02	$1.93 \pm 0.11\ (821)$	1.67 ± 0.08
NO	YES	NO	0.20	0.55 ± 0.02	$1.95 \pm 0.11\ (821)$	1.69 ± 0.08
NO	YES	YES	0.05	0.56 ± 0.02	$1.75 \pm 0.10\ (819)$	1.52 ± 0.07
NO	YES	YES	0.10	0.56 ± 0.02	$1.76 \pm 0.10\ (819)$	1.53 ± 0.07
NO	YES	YES	0.20	0.56 ± 0.02	$1.78 \pm 0.10\ (819)$	1.54 ± 0.07
YES	NO	NO	0.05	0.56 ± 0.02	$1.99 \pm 0.11\ (837)$	1.74 ± 0.08
YES	NO	NO	0.10	0.56 ± 0.02	$2.00 \pm 0.11\ (837)$	1.75 ± 0.08
YES	NO	NO	0.20	0.56 ± 0.02	$2.02 \pm 0.11\ (837)$	1.77 ± 0.08
YES	NO	YES	0.05	0.58 ± 0.02	$1.77 \pm 0.10\ (832)$	1.56 ± 0.07
YES	NO	YES	0.10	0.58 ± 0.02	$1.78 \pm 0.10\ (832)$	1.56 ± 0.07
YES	NO	YES	0.20	0.58 ± 0.02	$1.80 \pm 0.10\ (832)$	1.58 ± 0.07
YES	YES	NO	0.05	0.57 ± 0.02	$1.92 \pm 0.11\ (839)$	1.69 ± 0.08
YES	YES	NO	0.10	0.57 ± 0.02	$1.93 \pm 0.11\ (839)$	1.70 ± 0.08
YES	YES	NO	0.20	0.57 ± 0.02	$1.95 \pm 0.11\ (839)$	1.72 ± 0.08
YES	YES	YES	0.05	0.58 ± 0.02	$1.74 \pm 0.10\ (835)$	1.54 ± 0.07
YES	YES	YES	0.10	0.58 ± 0.02	$1.75 \pm 0.10\ (835)$	1.55 ± 0.07
YES	YES	YES	0.20	0.58 ± 0.02	$1.77 \pm 0.10\ (835)$	1.56 ± 0.07

†Data set C was split into 47,930 training observations and 1064 test observations. Logloss = ∞ in the complete test set since the realized category of some cases was predicted with probability 0. Excluding those cases, we report logloss on different, non-comparable subsets of size N.

3.3 Memory-based Reasoning (Creecy et al. 1992)

The following parameters are varied:

- whether to exclude stopwords or not;
- whether to use stemming or not;
- choice of similarity metric to calculate 'nearness' between observations;
- choice of k, the number of nearest matches to consider.

In all trials we use extended preprocessing and not just removal of punctuation marks. The reason is that preprocessing standardizes the text better by replacing German umlauts and other special characters as well. The exact type of preprocessing should, however, not lead to very different results (and we did see an empirical proof of this when tuning the Adapted Nearest Neighbor algorithm).

The different parameter configurations are evaluated on data set D (Table A7) and data set C (Table A8). Accuracies are very similar in all situations unless k is chosen too small. This suggests that the exact parameter configuration is of minor relevance, mirroring a finding by (Creecy et al., 1992). Note, however, that the metric used for tuning, accuracy, relates to 100% production rate. We did not try if the agreement rate among top 1 could be improved at lower production rates using a different parameter configuration.

For all of our analyses reported in the main text we use extended preprocessing, exclude stopwords, apply the German Porter stemmer, and use the ERROR-metric with $k = 7$.

3.4 Tree Boosting (XGBoost)

Tuning an XGBoost-model is a demanding task because many different tuning parameters are available and interactions between them need to be considered. Due to the special characteristics of occupational data, we use a parameter configuration that is different from standard recommendations.

Parameter tuning of logistic regression (see above) suggested that the German Porter stemmer can improve performance. The exclusion of stopwords and counting of words did not have a large impact on the results. For simplicity we do not analyze this again, but use stemming and do not remove stopwords with XGBoost. A variable counting the number of words is included in the predictor matrix to allow for possible interactions with it.

Tree boosting is an iterative process, where the r-th tree aims to find a signal that the $r-1$ previous iterations have not fully detected. To save time, we aim to achieve close-to-optimal predictions after approximately 20 rounds and fix the maximal number of iterations at 40. η (and other parameters) control the learning rate and larger values of η will make sure that we obtain reasonable results by then. However, if η is too large or if we run too many iterations, overfitting becomes an issue and the r-th tree will pick up noise in the training data that will lead to poor predictions in the test data. In such

Table A7: Accuracy of Memory-based Reasoning in data set D under different configurations.†

metric	*k*	*stem*	*stop*	*accuracy*
SUM	2	NO	YES	0.508 ± 0.015
SUM	2	NO	NO	0.502 ± 0.015
SUM	2	YES	YES	0.497 ± 0.015
SUM	2	YES	NO	0.486 ± 0.015
SUM	7	NO	YES	0.567 ± 0.015
SUM	7	NO	NO	0.563 ± 0.015
SUM	7	YES	YES	0.568 ± 0.015
SUM	7	YES	NO	0.569 ± 0.015
SUM	12	NO	YES	0.573 ± 0.015
SUM	12	NO	NO	0.572 ± 0.015
SUM	12	YES	YES	0.571 ± 0.015
SUM	12	YES	NO	0.570 ± 0.015
SUM	17	NO	YES	0.573 ± 0.015
SUM	17	NO	NO	0.574 ± 0.015
SUM	17	YES	YES	0.576 ± 0.015
SUM	17	YES	NO	0.577 ± 0.015
ERROR	2	NO	YES	0.549 ± 0.015
ERROR	2	NO	NO	0.535 ± 0.015
ERROR	2	YES	YES	0.557 ± 0.015
ERROR	2	YES	NO	0.542 ± 0.015
ERROR	7	NO	YES	0.568 ± 0.015
ERROR	7	NO	NO	0.564 ± 0.015
ERROR	7	YES	YES	0.576 ± 0.015
ERROR	7	YES	NO	0.570 ± 0.015
ERROR	12	NO	YES	0.564 ± 0.015
ERROR	12	NO	NO	0.555 ± 0.015
ERROR	12	YES	YES	0.569 ± 0.015
ERROR	12	YES	NO	0.566 ± 0.015
ERROR	17	NO	YES	0.567 ± 0.015
ERROR	17	NO	NO	0.554 ± 0.015
ERROR	17	YES	YES	0.569 ± 0.015
ERROR	17	YES	NO	0.559 ± 0.015

†Data set D was split into 6575 training observations and 1064 test observations.

Table A8: Accuracy of Memory-based Reasoning in data set C under different configurations.†

metric	*k*	*stem*	*stop*	*accuracy*
SUM	2	NO	YES	0.486 ± 0.015
SUM	2	NO	NO	0.479 ± 0.015
SUM	2	YES	YES	0.482 ± 0.015
SUM	2	YES	NO	0.475 ± 0.015
SUM	7	NO	YES	0.575 ± 0.015
SUM	7	NO	NO	0.571 ± 0.015
SUM	7	YES	YES	0.569 ± 0.015
SUM	7	YES	NO	0.565 ± 0.015
SUM	12	NO	YES	0.573 ± 0.015
SUM	12	NO	NO	0.570 ± 0.015
SUM	12	YES	YES	0.566 ± 0.015
SUM	12	YES	NO	0.563 ± 0.015
SUM	17	NO	YES	0.576 ± 0.015
SUM	17	NO	NO	0.571 ± 0.015
SUM	17	YES	YES	0.576 ± 0.015
SUM	17	YES	NO	0.571 ± 0.015
ERROR	2	NO	YES	0.573 ± 0.015
ERROR	2	NO	NO	0.567 ± 0.015
ERROR	2	YES	YES	0.570 ± 0.015
ERROR	2	YES	NO	0.564 ± 0.015
ERROR	7	NO	YES	0.586 ± 0.015
ERROR	7	NO	NO	0.582 ± 0.015
ERROR	7	YES	YES	0.582 ± 0.015
ERROR	7	YES	NO	0.581 ± 0.015
ERROR	12	NO	YES	0.571 ± 0.015
ERROR	12	NO	NO	0.568 ± 0.015
ERROR	12	YES	YES	0.571 ± 0.015
ERROR	12	YES	NO	0.569 ± 0.015
ERROR	17	NO	YES	0.568 ± 0.015
ERROR	17	NO	NO	0.564 ± 0.015
ERROR	17	YES	YES	0.564 ± 0.015
ERROR	17	YES	NO	0.56 ± 0.015

†Data set C was split into 47,930 training observations and 1064 test observations.

cases we use early stopping and report results from earlier iterations before overfitting occurred.

In multiclass classification with K classes, XGBoost learns in each iteration K trees with binary outcome. The parameters max_depth, min_child_weight, γ, and $tree_method$ control the tree growing process.

- max_depth determines the maximal number of leaves in a tree, 2^{max_depth}. For example a tree of depth 2 can have a leaf for "egg" and another leaf for "chicken". These leaves should output high probabilities for chicken farming occupations. Trees with depth 2 also allow for a leaf that may predict a probability if the word from the first leaf ("egg") co-occurs with some other word. A final leaf will cover all other verbal answers that neither contain "egg" nor "chicken". More than four leaves are not possible with $max_depth = 2$. Thus, if max_depth is chosen too small, a single tree cannot detect all words (and possibly co-occurrences of words) that are predictive of a given category. A large number of iterations would then be needed before the boosting algorithm can make good predictions.

- min_child_weight controls that the tree building process is stopped whenever the sum of hessians in a leaf $\sum_{i \in \text{leaf}} \partial^2_{\hat{y}_i^{(r-1)}} l(y_i, \hat{y_i}^{(r-1)}) = \sum_{i \in \text{leaf}} p_i(1-p_i)$ is smaller than min_child_weight. We anticipate that tree leaves can be very pure (a leaf for the job title "beekeeper" perfectly predicts a single category, $p_i = 1$), which suggests that min_child_weight should be set to zero.

- γ determines the minimum loss reduction needed to split a leaf further. If it is too small, many words that are not job-related (e.g., "and", "am", ...) will be used to make predictions about job categories, which does not seem reasonable. If it is too large, the algorithm will ignore important words.

- $tree_method$ controls the tree construction algorithm, which mainly differ in how they enumerate possible splitting points. We use the exact greedy algorithm ($tree_method =' exact'$), reasoning that our predictor variables are mostly binary (either a word occurs in a verbal answer or not) and faster but more approximate algorithms for continuous predictors should not be necessary.

λ (L2 norm), α (L1 norm), and max_delta_step are shrinkage parameters that control by how much the weights from each leaf are shrunk towards zero. max_delta_step is recommended in situations as ours with high class imbalance and we find it to be one of the most important parameters. It defines a maximal weight that can be returned from a single leaf, making sure that the leaf does not overfit by too much. Because the returned value is then multiplied by η, the choice of max_delta_step has a direct impact on the learning rate.

The parameters $subsample$, $colsample_bytree$, and $colsample_bylevel$ let us specify if we do not want to train every tree with the complete data. $subsample$ draws a sample of training instances and uses only the selected cases to learn a tree. The $colsample$ parameters let us randomly sample a set of variables to be used for tree learning. The XGBoost default implementation fixes all three parameters at a proportion 1 (=use complete data), but lowering the proportions can speed up computation

of each iteration (although we will need more iterations) and may prevent overfitting. In our application we are skeptical whether drawing a subsample of variables that will be used to construct the complete tree ($colsample_bytree$) is reasonable. If two words interact, but the subsample contains only one of the two words, these interactions could not be detected during tree learning.

An important property of tree boosting must be observed. All categories are predicted with probability close to zero in the first few iterations and the probabilities will usually increase in each iteration if this improves the objective function. Consequently, the sum of probabilities of the most probable category ($\sum_{n_f=1}^{N_{test}} \hat{p}_{n_f}^{(1)}$ in the notation from Section 2) increases with the number of iterations. However, it should not overshoot the observed number of cases correctly predicted, $\sum_{n_f=1}^{N_{test}} a_{n_f}^{(1)}$, which would be hard to reverse and is a sign of overfitting. In our experience, increasing γ or λ helps to prevent overshooting behavior.

Our main objective is to minimize logloss. We thus stop iterating when logloss stays approximately constant. If we were to continue iterating for too long, overfitting would occur and logloss would deteriorate. However, we do not always iterate until logloss remains constant and report non-optimal logloss for some parameter configurations. Two reasons for this exist that must be kept in mind when interpreting the results.

- The first reason is that we run at most 40 iterations. If we had time for more iterations, logloss might improve further.
- The second reason is that we require calibrated probabilities. $\bar{\hat{p}}_{n_f}^{(1)} = \frac{1}{N_{test}} \sum_{n_f=1}^{N_{test}} \hat{p}_{n_f}^{(1)}$ increases with the number of iterations and it may overshoot the observed accuracy. We thus stop early if it is by 0.01 points larger than the accuracy. When this is the reason for stopping, logloss is often satisfactory, but further improvements might be possible with additional fine-tuning of parameters.

Many different parameter configurations were tested on data set D until we finally reached a default parameter configuration that is sufficient for our purpose. Results are shown in Table A9. To demonstrate the performance under different parameter configurations, all parameters except the one indicated are kept constant, which allows to compare deviating configurations with the default configuration. The optimal parameter configuration would minimize logloss and sharpness, the latter is only useful under the condition that probabilities are calibrated. High accuracy is also desirable, but note that most differences in accuracy are rather small and may be due to chance. In addition, we would like that the mean most probable probability over all test cases, $\bar{\hat{p}}_{n_f}^{(1)} = \frac{1}{N_{test}} \sum_{n_f=1}^{N_{test}} \hat{p}_{n_f}^{(1)}$, is close to the accuracy $acc = \frac{1}{N_{test}} \sum_{n_f=1}^{N_{test}} a_{n_f}^{(1)}$, an indicator of calibration.

The results in Table A9 confirm our theoretical reasoning. max_delta_step and η determine the rate of learning. If they are too small, results might still improve but the calculations will take too long. If they are too large, we obtain poor results within few iterations. The exact choice of max_depth makes little difference, but one may speculate that larger max_depth might achieve slightly better results in fewer iterations. With small γ we observe overshooting behavior and the algorithm stops early.

Fixing $\lambda = 0$ leads to poor learning behavior. However, if either γ or λ are set to overly large values, $\bar{\hat{p}}_{n_f}^{(1)}$ will underestimate the accuracy. Increasing α leads to additional iterations and makes the results worse. Similarly, the results deteriorate drastically if $min_child_weight > 0$. Furthermore, no improvements can be observed if $subsample$ is varied or if $colsample_bytree$ is reduced, but sampling smaller proportions will increase the number of iterations needed. Changing $colsample_bylevel$ to 0.6 leads to minimal improvements in logloss, but sharpness and accuracy are worse, a behavior that could be explored further. All in all, one could certainly try additional parameter configurations and/or run more iterations to achieve better results, though this time would be misspend. Results of most parameter configurations are already quite similar to another, suggesting that our default parameter configuration is not too far away from some hypothetical optimum.

Since the default parameter configuration was optimized in data set D, another question is if the results can be generalized to other occupational data. Table A10 shows performance measures in data set C using identical parameter configurations as before. The most salient patterns described above are confirmed (referring to max_delta_step, η, max_depth (partly confirmed), γ, λ, min_child_weight, $colsample_bytree$) except that changes in α may actually improve performance and lowering $subsample$ has in this larger data set no effect on the number of iterations needed. Among the reported parameter configurations, the logloss is minimal with $\lambda = 1$. With the default parameter configuration logloss is 0.083 bits away from this observed optimum, still a reasonable choice that will be used in the main paper for a comparison of XGBoost with other methods. Yet, if XGBoost predictions of data set C were used for practical applications, we would recommend trying a few more parameter configurations, maybe increasing γ and/or λ in a first step. This would prevent the omnipresent overshooting behavior in this table and reduce logloss further.

Table A9: Performance measures of XGBoost in data set D. All parameters except one are kept constant at the default configuration.†

Parameter configuration	$iter$††	acc	$\bar{\hat{p}}_{n_f}^{(1)}$	$logloss$	$sharp$
$default$	22 [1]	0.591	0.588	3.826	3.788
$max_delta_step = 0.33$	40 [3]	0.583	0.543	$\leq$3.878	4.210
$max_delta_step = 3$	15 [1]	0.590	0.591	3.862	3.849
$\eta = 0.25$	40 [3]	0.590	0.574	$\leq$3.810	4.019
$\eta = 1$	9 [2]	0.572	0.576	$\leq$3.915	3.944
$max_depth = 10$	23 [1]	0.583	0.580	3.885	3.690
$max_depth = 30$	22 [1]	0.582	0.589	3.818	3.806
$\gamma = 0.3$	16 [2]	0.583	0.592	$\leq$4.007	4.034
$\gamma = 0.9$	22 [1]	0.585	0.554	3.826	4.079
$\lambda = 0$	11 [2]	0.039	0.049	$\leq$8.128	8.228
$\lambda = 1$	26 [1]	0.576	0.525	3.931	4.505
$\alpha = 0.2$	26 [1]	0.583	0.545	3.865	4.346
$min_child_weight = 0.01$	22 [1]	0.555	0.543	4.122	4.213
$subsample = 0.5$	28 [1]	0.579	0.588	3.861	3.689
$subsample = 1$	20 [1]	0.583	0.589	3.857	3.863
$colsample_bytree = 0.6$	28 [1]	0.580	0.533	3.858	4.072
$colsample_bylevel = 0.6$	22 [1]	0.575	0.570	3.808	3.952

†Data set D was split into 6575 training observations and 1064 test observations. The complete test set was used to calculate logloss. Default parameter configuration: Stopwords = NO, Stemming = YES, extended preprocessing = YES, word counts = YES, max. rounds = 40, early stopping if performance does not improve for 1 round, $\eta = 0.5$, max_delta_step = 1, max_depth = 20, $\gamma = 0.6$, $\lambda = 1e^{-4}$, $\alpha = 0$, min_child_weight = 0, subsample = 0.75, colsample_by_tree = 1, colsample_by_level = 1.
††Reason for stopping at r-th iteration: [1] = Logloss minimal or approximately constant; [2] = $\bar{\hat{p}}_{n_f}^{(1)} - acc > 0.01$ in subsequent iteration; [3] = maximal number of iterations reached. Reasons [2] and [3] entail the possibility to reduce logloss further (indicated by $\leq$).

Table A10: Performance measures of XGBoost in data set C. All parameters except one are kept constant at the default configuration.†

Parameter configuration	$iter$††	acc	$\bar{\hat{p}}_{n_f}^{(1)}$	$logloss$	$sharp$
$default$	19 [2]	0.597	0.605	$\leq$3.361	3.321
$max_delta_step = 0.33$	39 [2]	0.590	0.597	$\leq$3.327	3.265
$max_delta_step = 3$	10 [2]	0.582	0.586	$\leq$3.503	3.809
$\eta = 0.25$	36 [2]	0.589	0.594	$\leq$3.371	3.498
$\eta = 1$	7 [2]	0.578	0.564	$\leq$3.539	3.929
$max_depth = 10$	30 [2]	0.586	0.594	$\leq$3.374	3.291
$max_depth = 30$	17 [2]	0.600	0.608	$\leq$3.385	3.377
$\gamma = 0.3$	16 [2]	0.587	0.595	$\leq$3.482	3.533
$\gamma = 0.9$	24 [2]	0.597	0.602	$\leq$3.300	3.290
$\lambda = 0$	15 [1]	0.044	0.031	8.489	8.568
$\lambda = 1$	24 [1]	0.594	0.602	3.278	3.305
$\alpha = 0.2$	24 [2]	0.594	0.602	$\leq$3.312	3.295
$min_child_weight = 0.01$	20 [2]	0.562	0.569	$\leq$3.608	3.731
$subsample = 0.5$	19 [2]	0.597	0.592	$\leq$3.437	3.412
$subsample = 1$	20 [2]	0.592	0.599	$\leq$3.358	3.451
$colsample_bytree = 0.6$	27 [1]	0.598	0.575	3.347	3.549
$colsample_bylevel = 0.6$	17 [2]	0.582	0.588	$\leq$3.408	3.571

†Data set C was split into 47.930 training observations and 1064 test observations. The complete test set was used to calculate logloss. Default parameter configuration: Stopwords = NO, Stemming = YES, extended preprocessing = YES, word counts = YES, max. rounds = 40, early stopping if performance does not improve for 1 round, $\eta = 0.5$, max_delta_step = 1, max_depth = 20, $\gamma = 0.6$, $\lambda = 1e^{-4}$, $\alpha = 0$, min_child_weight = 0, subsample = 0.75, colsample_by_tree = 1, colsample_by_level = 1.
††Reason for stopping at r-th iteration: [1] = Logloss minimal or approximately constant; [2] = $\bar{\hat{p}}_{n_f}^{(1)} - acc > 0.01$ in subsequent iteration; [3] = maximal number of iterations reached. Reasons [2] and [3] entail the possibility to reduce logloss further (indicated by $\leq$).

3.5 Similarity-based reasoning

$dist.type$ determines how similarities between verbal answers and coding index entries are calculated. There exist other parameters that can be varied:

- whether to exclude stopwords or not (only available if $dist.type = "wordwise"$);
- whether to use stemming or not;
- whether to do extended preprocessing;
- whether to remove punctuation marks;
- weights to use when calculating the distance between two strings (only if $dist.type = "fulltext"$ or $dist.type = "wordwise"$);
- a threshold above which strings are considered dissimilar;
- number of draws (n.draws).

We do not exclude stopwords because the coding index should determine which words are useful. Stemming is not used because job titles from the coding index were not stemmed. We use extended preprocessing because this makes verbal answers and entries from the coding index more similar. The removal of punctuation marks is not needed because this is already done during extended preprocessing.

The similarity calculations do not only depend on $dist.type$ but also on the algorithm used and the weights used therein. Both wordwise distance and fulltext distance are calculated using the *optimal string alignment*-distance from the R-package `stringdist` (van der Loo, 2014). There exist separate weights for character deletion, character insertion, character substitution, and adjacent character transposition. For simplicity we set all four weights to one, known as the Damerau-Levenshtein distance. A threshold of two thus refers to the maximal number of character operations to turn one string into another. The threshold should not be too high because the number of misspelled characters is usually rather low.

The number of draws (n.draws) determines how many samples are drawn for Monte Carlo integration. Computations become more exact as one increases the number of draws, but they will also take longer.

Tables A11 and A12 explore how the results change when using different parameter configurations. For all computations in the main paper we set n.draw to 100. The threshold is set to 1 despite evidence that logloss and accuracy increase further if it were set to 2. The reason is that improvements are rather small and due to a few additional observations that can be classified when using a larger threshold. If the threshold were fixed at two, issues arise when distinct job titles differ by just two characters, leading to decreased $\bar{\hat{p}}_{n_f}^{(1)}$ (a sign of poor calibration).

Table A11: Performance measures of similarity-based reasoning in data set D under different configurations.†

$dist.type$	$thresh$	$n.draw$	$n.match$	acc	$\bar{\hat{p}}_{n_f}^{(1)}$	$logloss$	$sharp$
$fulltext$	0	10	356	0.320	0.314	7.104	7.092
$fulltext$	0	50	356	0.320	0.315	7.102	7.091
$fulltext$	0	250	356	0.320	0.315	7.102	7.091
$fulltext$	1	10	396	0.355	0.346	6.747	6.775
$fulltext$	1	50	396	0.355	0.346	6.747	6.775
$fulltext$	1	250	396	0.355	0.347	6.747	6.775
$fulltext$	2	10	425	0.369	0.339	6.586	6.826
$fulltext$	2	50	425	0.370	0.338	6.586	6.826
$fulltext$	2	250	425	0.369	0.338	6.586	6.826
$substring$	-	50	962	0.557	0.444	3.893	4.716
$wordwise$	0	10	801	0.515	0.467	4.875	5.036
$wordwise$	0	50	801	0.508	0.466	4.878	5.035
$wordwise$	0	250	801	0.513	0.466	4.877	5.035
$wordwise$	1	10	901	0.536	0.446	4.552	5.423
$wordwise$	1	50	901	0.536	0.445	4.552	5.424
$wordwise$	1	250	901	0.535	0.444	4.555	5.424
$wordwise$	2	10	996	0.541	0.262	4.569	7.455
$wordwise$	2	50	996	0.543	0.262	4.567	7.456
$wordwise$	2	250	996	0.540	0.262	4.567	7.455

†Data set D was split into 6575 training observations and 1064 test observations. The complete test set was used to calculate logloss. Parameter configuration: Stopwords = NO, Stemming = NO, extended preprocessing = YES, optimal-string-alignment with (d,i,s,t) = (1,1,1,1).

Table A12: Performance measures of similarity-based reasoning in data set C under different configurations.†

$dist.type$	$thresh$	$n.draw$	$n.match$	acc	$\bar{\hat{p}}_{n_f}^{(1)}$	$logloss$	$sharp$
$fulltext$	0	10	596	0.417	0.402	5.800	5.798
$fulltext$	0	50	596	0.416	0.402	5.800	5.798
$fulltext$	0	250	596	0.417	0.402	5.799	5.798
$fulltext$	1	10	644	0.442	0.430	5.472	5.435
$fulltext$	1	50	644	0.442	0.430	5.473	5.435
$fulltext$	1	250	644	0.441	0.430	5.473	5.435
$fulltext$	2	10	672	0.451	0.430	5.321	5.325
$fulltext$	2	50	672	0.451	0.430	5.321	5.325
$fulltext$	2	250	672	0.453	0.430	5.321	5.325
$substring$	-	50	1006	0.555	0.440	3.582	4.394
$wordwise$	0	10	887	0.519	0.446	4.334	4.867
$wordwise$	0	50	887	0.518	0.446	4.334	4.867
$wordwise$	0	250	887	0.518	0.446	4.333	4.867
$wordwise$	1	10	955	0.549	0.465	3.959	4.642
$wordwise$	1	50	955	0.547	0.465	3.959	4.643
$wordwise$	1	250	955	0.547	0.465	3.959	4.643
$wordwise$	2	10	1005	0.558	0.413	3.866	5.386
$wordwise$	2	50	1005	0.558	0.413	3.867	5.386
$wordwise$	2	250	1005	0.558	0.413	3.867	5.386

†Data set C was split into 47.930 training observations and 1064 test observations. The complete test set was used to calculate logloss. Parameter configuration: Stopwords = NO, Stemming = NO, extended preprocessing = YES, optimal-string-alignment with (d,i,s,t) = (1,1,1,1).

4 Part D: Detailed Results

4.1 Agreement Rate Among Top 1 vs. Production Rate (Automated Coding)

Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases)

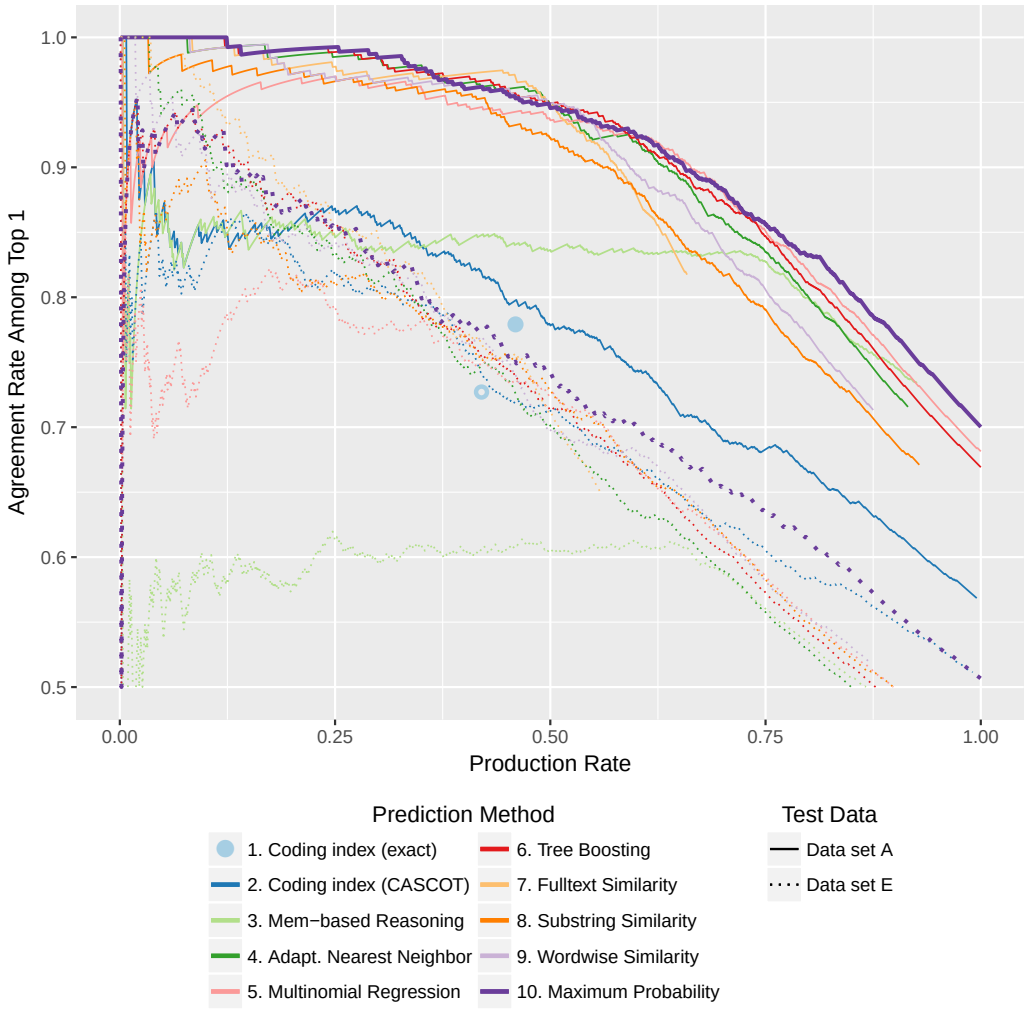


Figure A2: Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases); **Training data: *Data set A*** ($N = 31,867$)

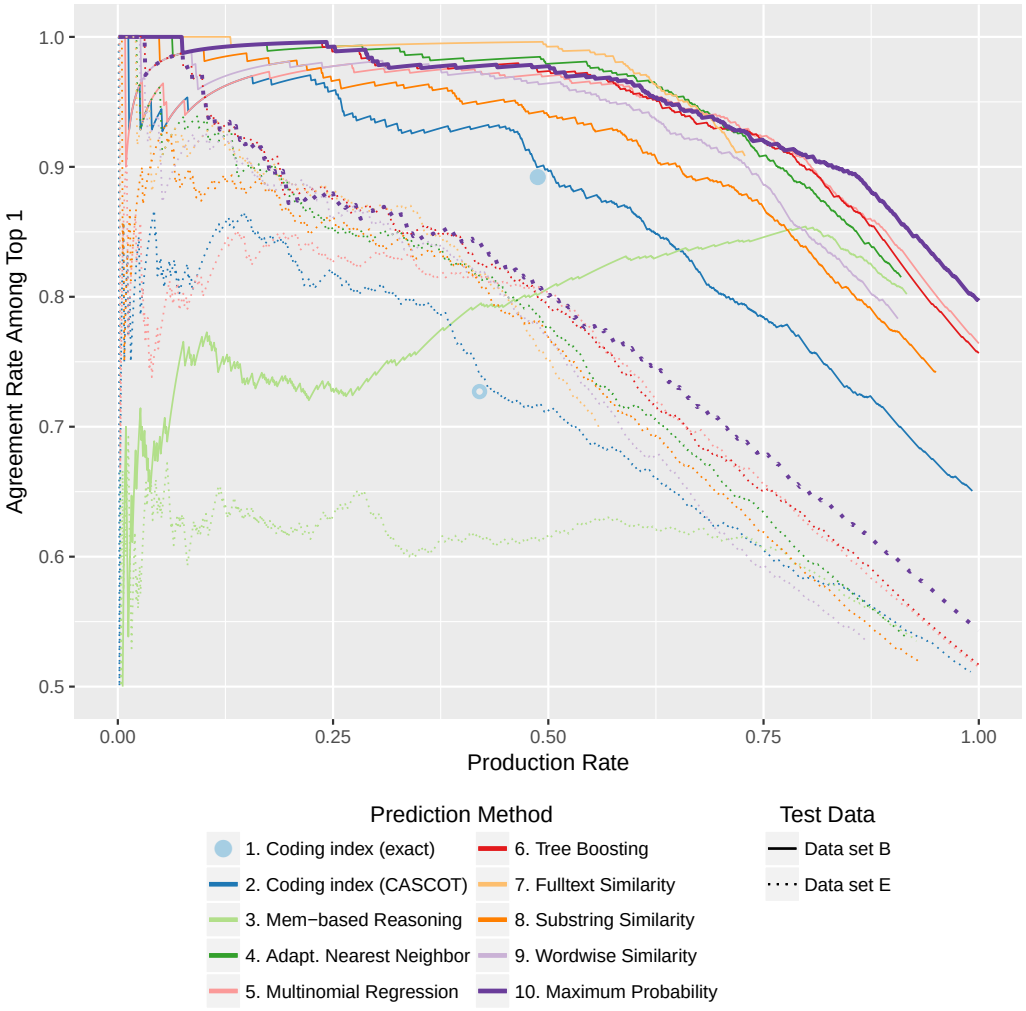


Figure A3: Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases); **Training data: *Data set B*** **($N = 54,880$)**

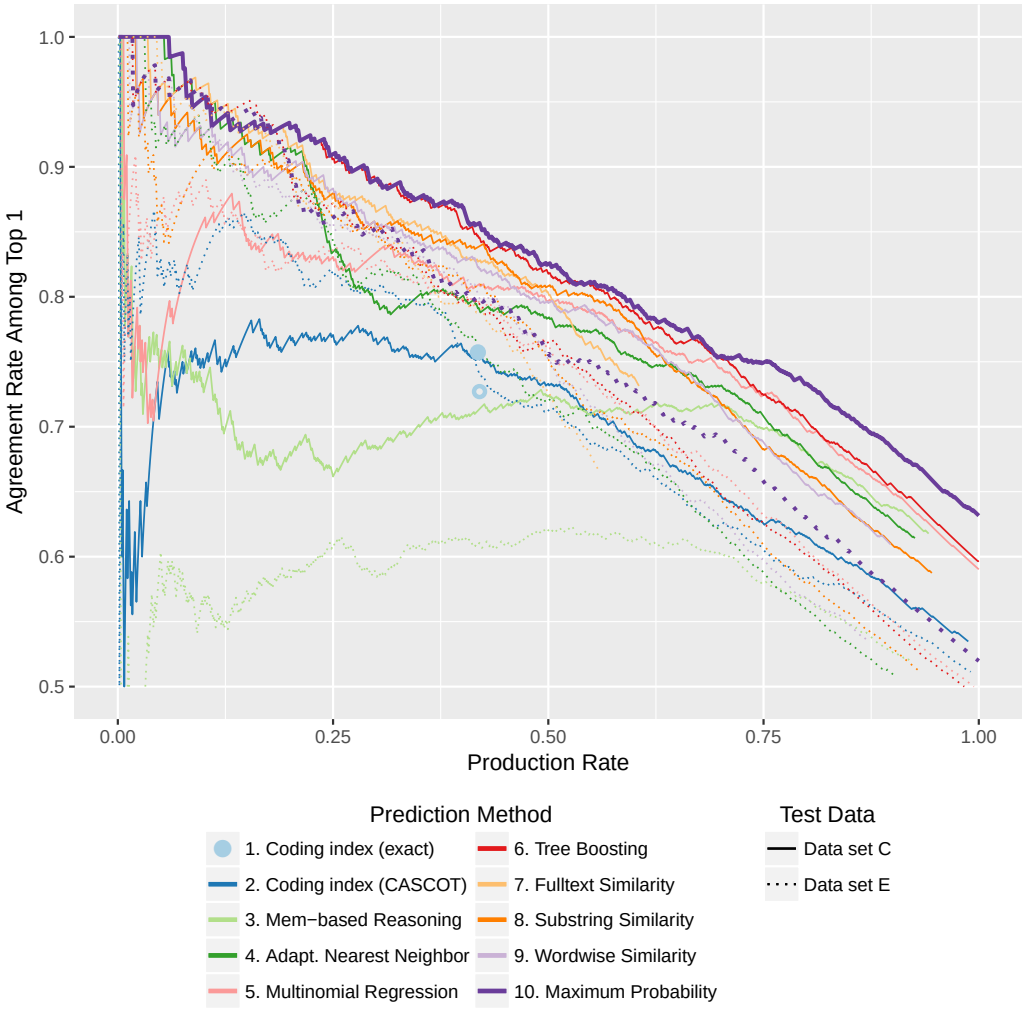


Figure A4: Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases); **Training data: *Data set C*** ($N = 47,930$)

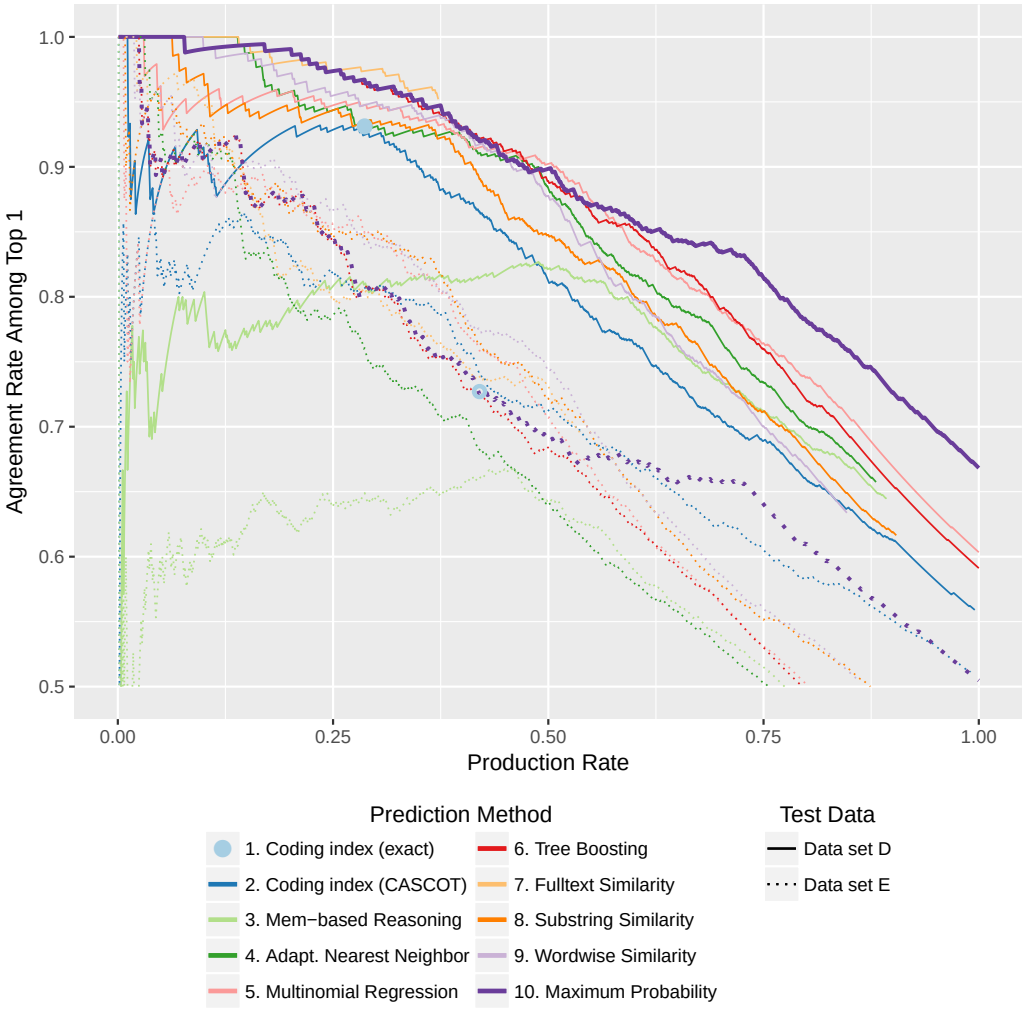


Figure A5: Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases); **Training data: *Data set D*** ($N = 6,575$)

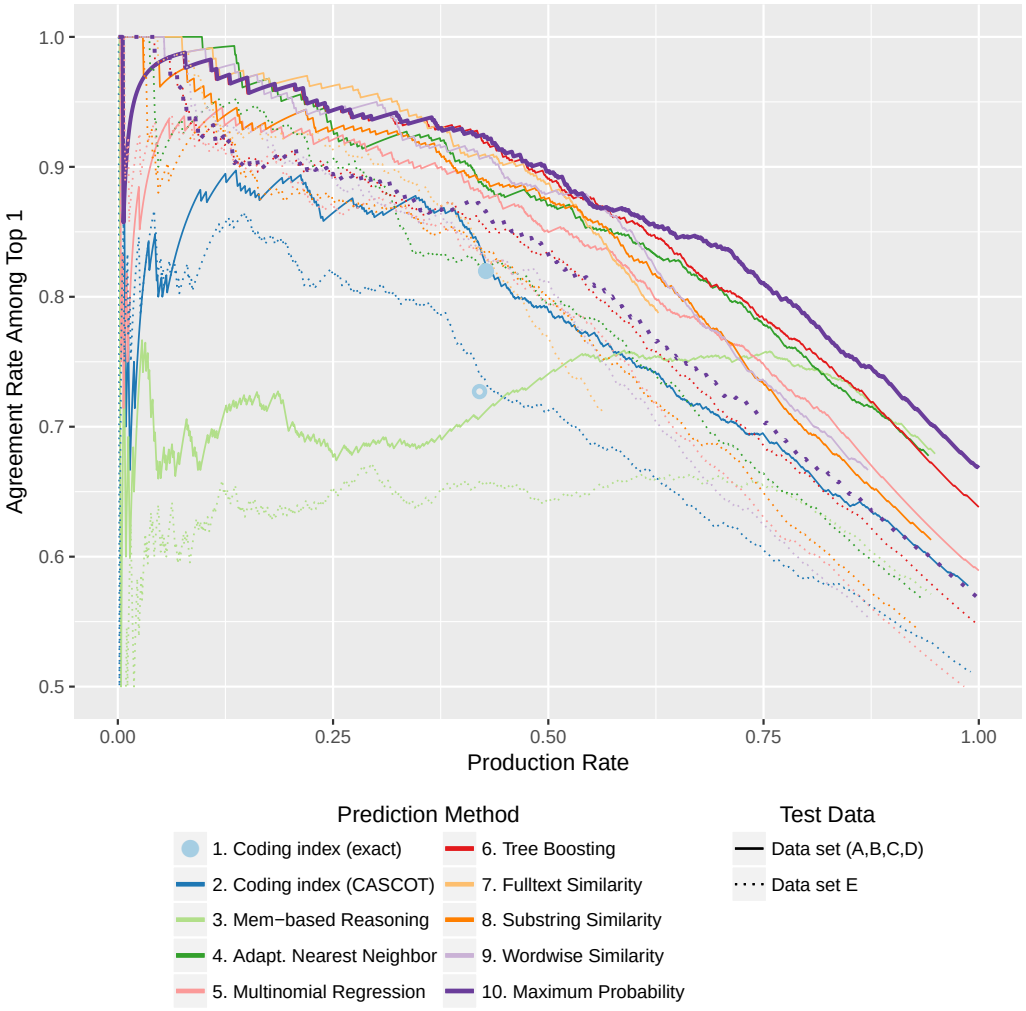


Figure A6: Agreement Rates of the most probable category at various production rates (appropriate for automated coding with allowance for residual cases); **Training data: *all data sets combined*** ($N = 144.444$)

4.2 Reliability Diagrams (k = 1)

Reliability Diagram: Ideal probabilistic predictions should match the observed relative frequencies, the diagonal; Point size is proportional to the number of observations within each bin.

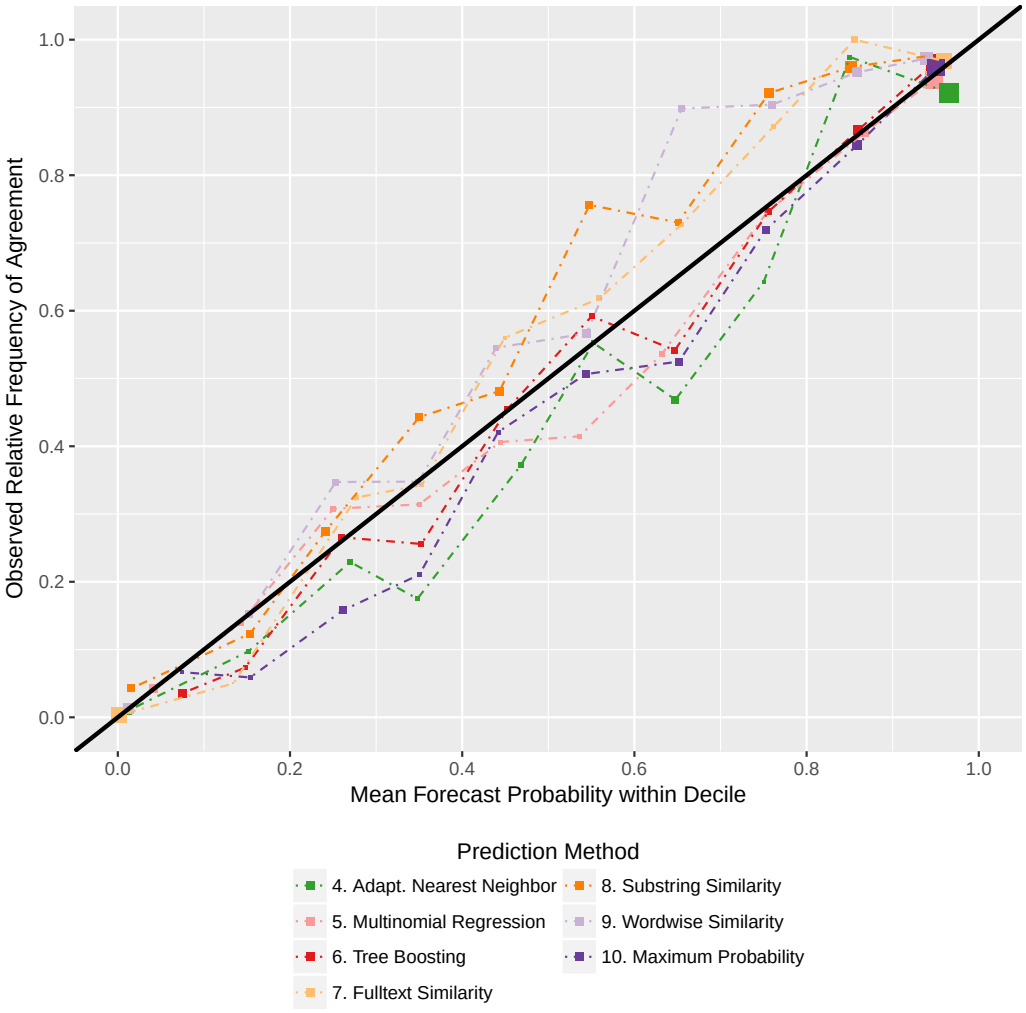


Figure A7: Training data: *Data set A* ($N = 31,867$), Test data: *Data set A* ($N = 1.064$)

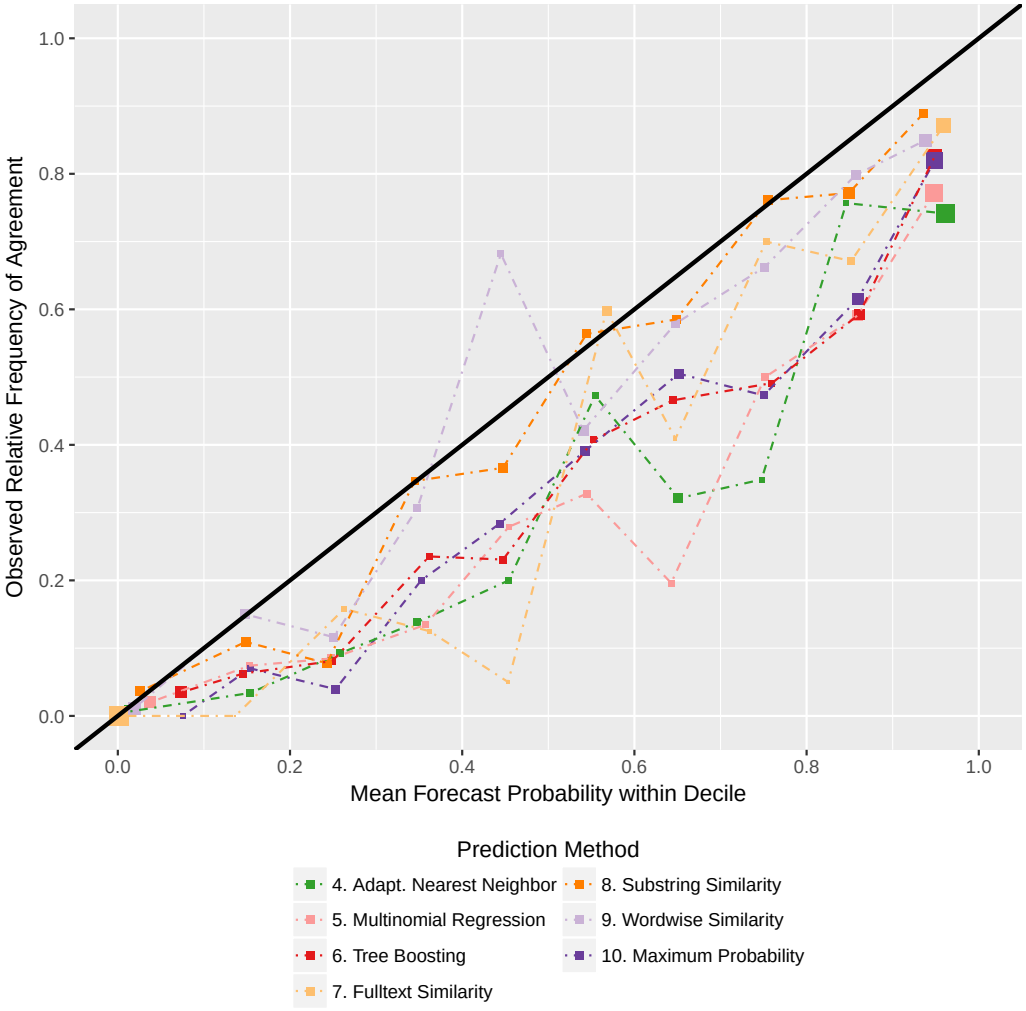


Figure A8: Reliability Diagram of most probable category (k=1); Training data: *Data set A* ($N = 31,867$), Test data: *Data set E* ($N = 1.064$)

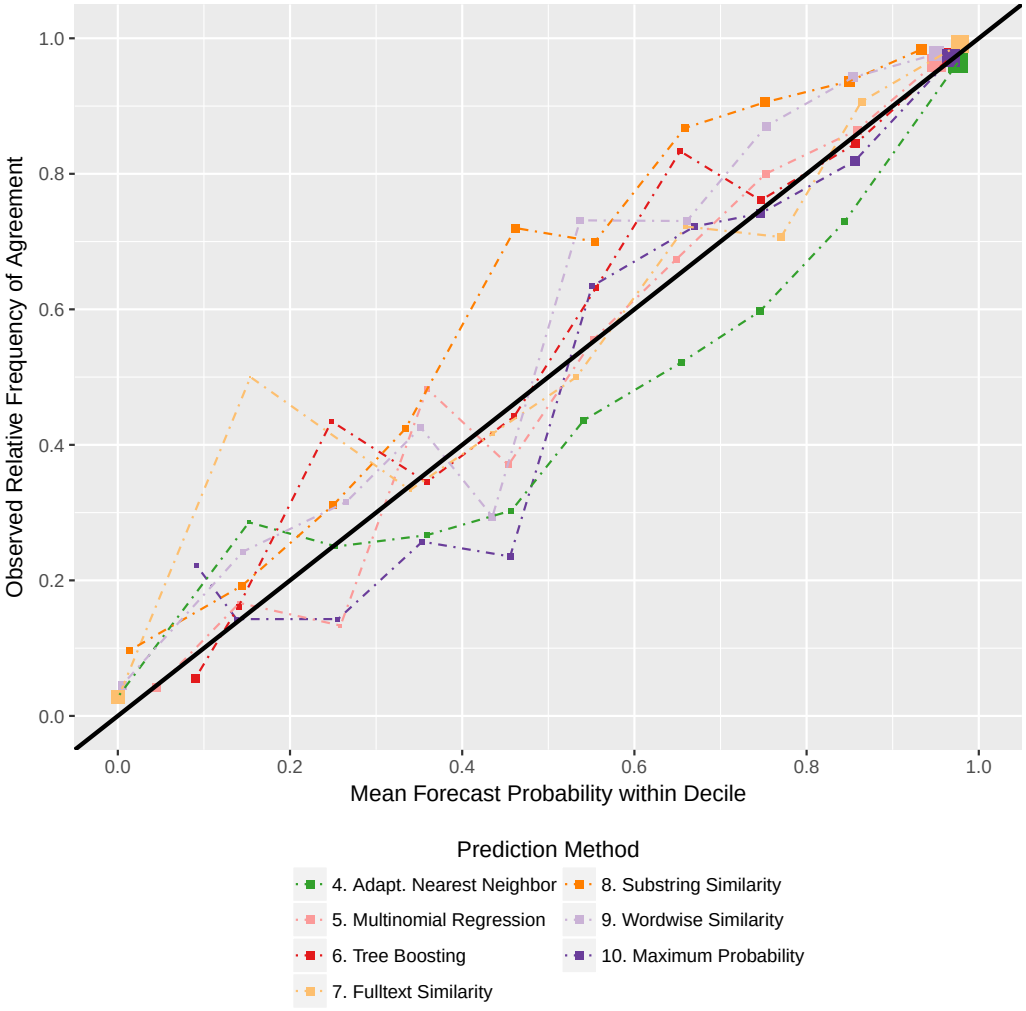


Figure A9: Reliability Diagram of most probable category (k=1); Training data: *Data set B* ($N = 54,880$), Test data: *Data set B* ($N = 1.064$)

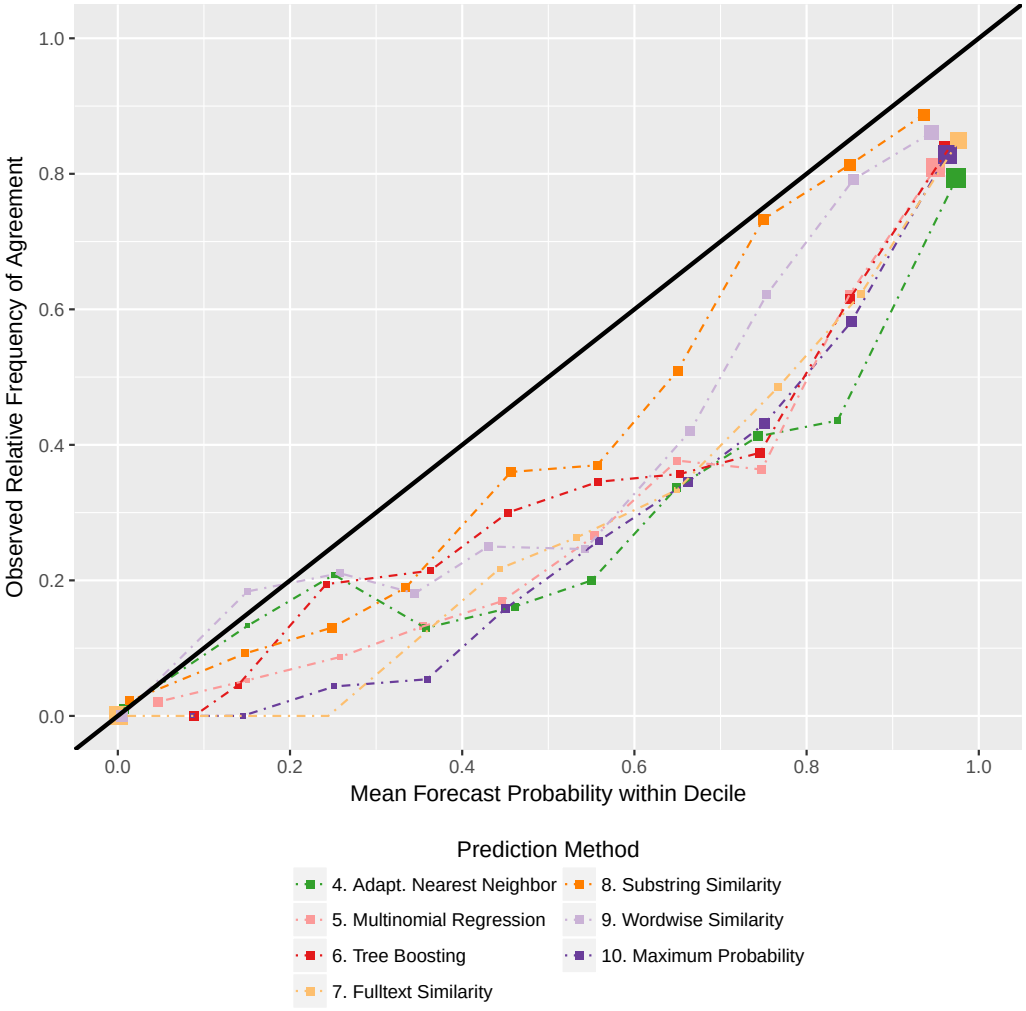


Figure A10: Reliability Diagram of most probable category (k=1); Training data: *Data set B* ($N = 54,880$), Test data: *Data set E* ($N = 1.064$)

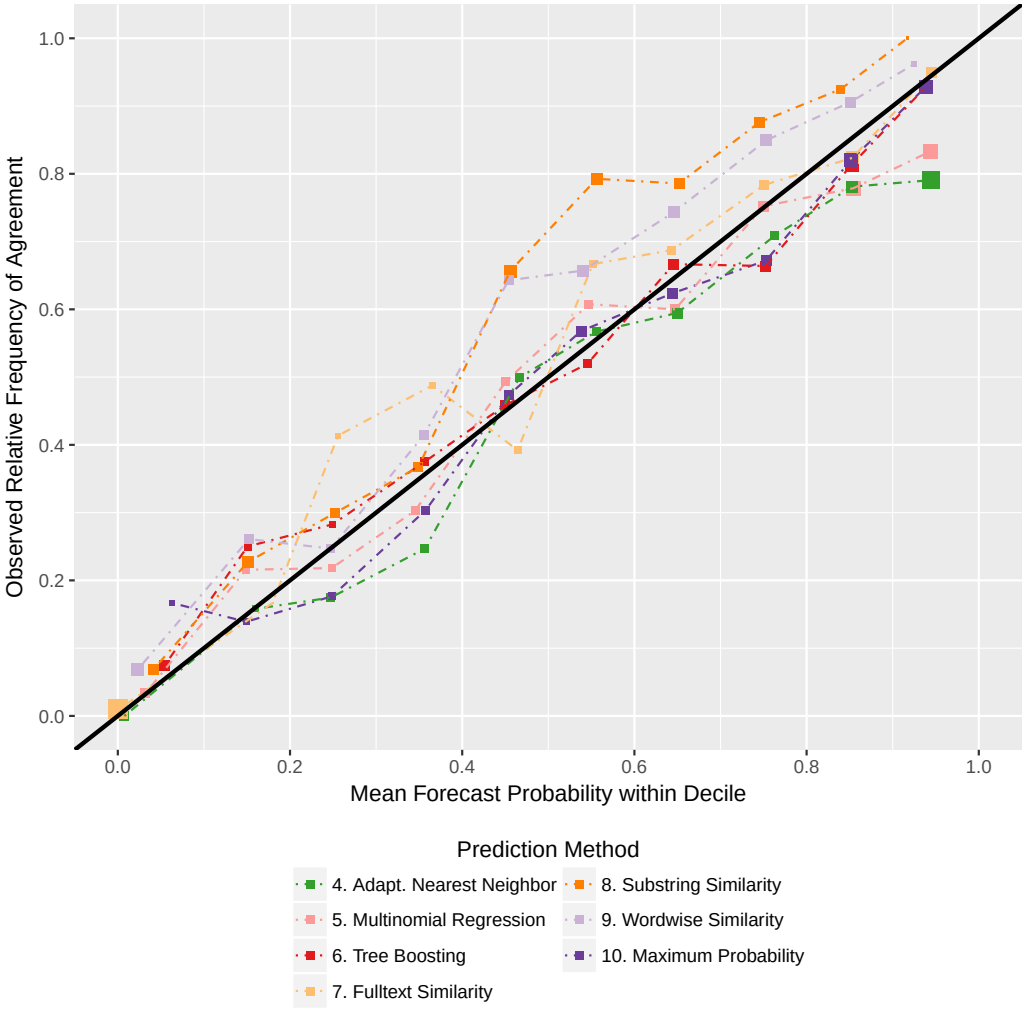


Figure A11: Reliability Diagram of most probable category (k=1); Training data: *Data set C* ($N = 47,930$), Test data: *Data set C* ($N = 1.064$)

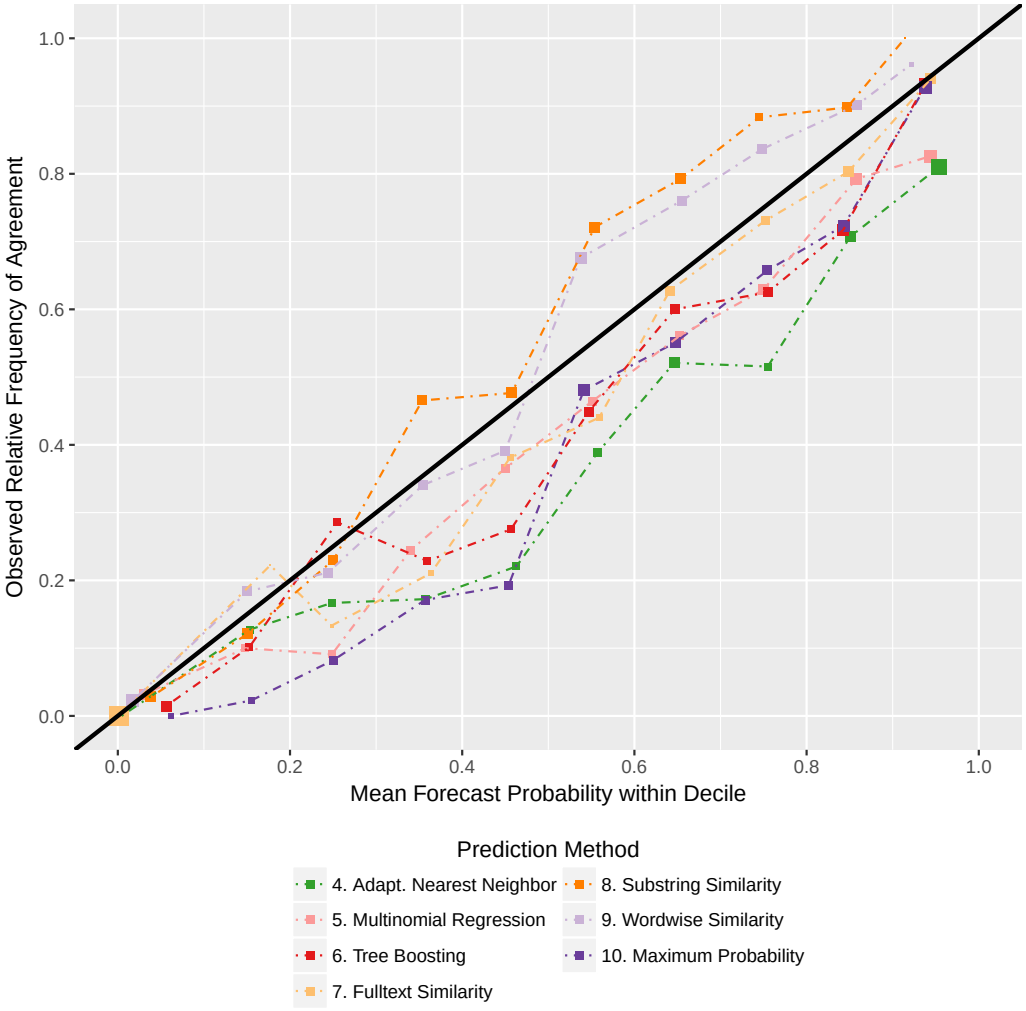


Figure A12: Reliability Diagram of most probable category (k=1); Training data:*Data set C* ($N = 47,930$), Test data: *Data set E* ($N = 1.064$)

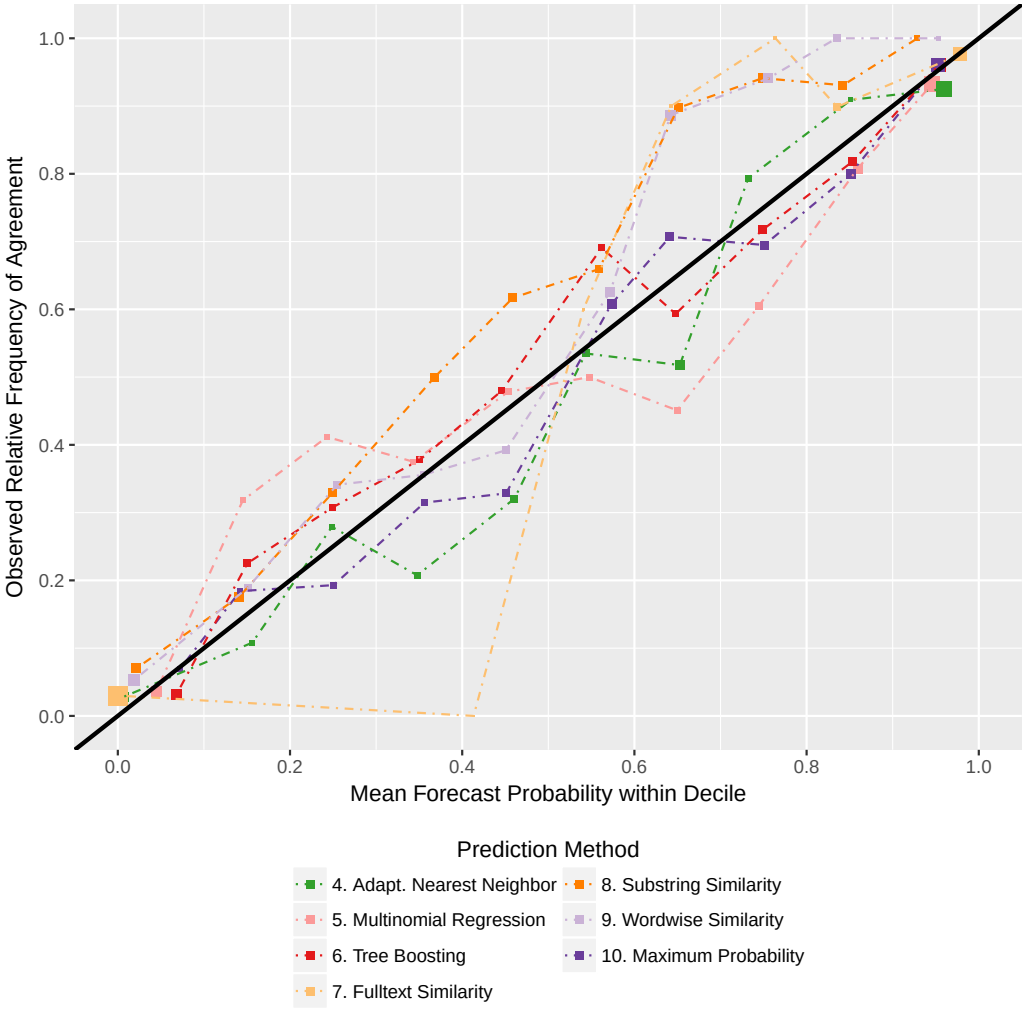


Figure A13: Reliability Diagram of most probable category (k=1); Training data: *Data set D* ($N = 6,575$), Test data: *Data set D* ($N = 1.064$)

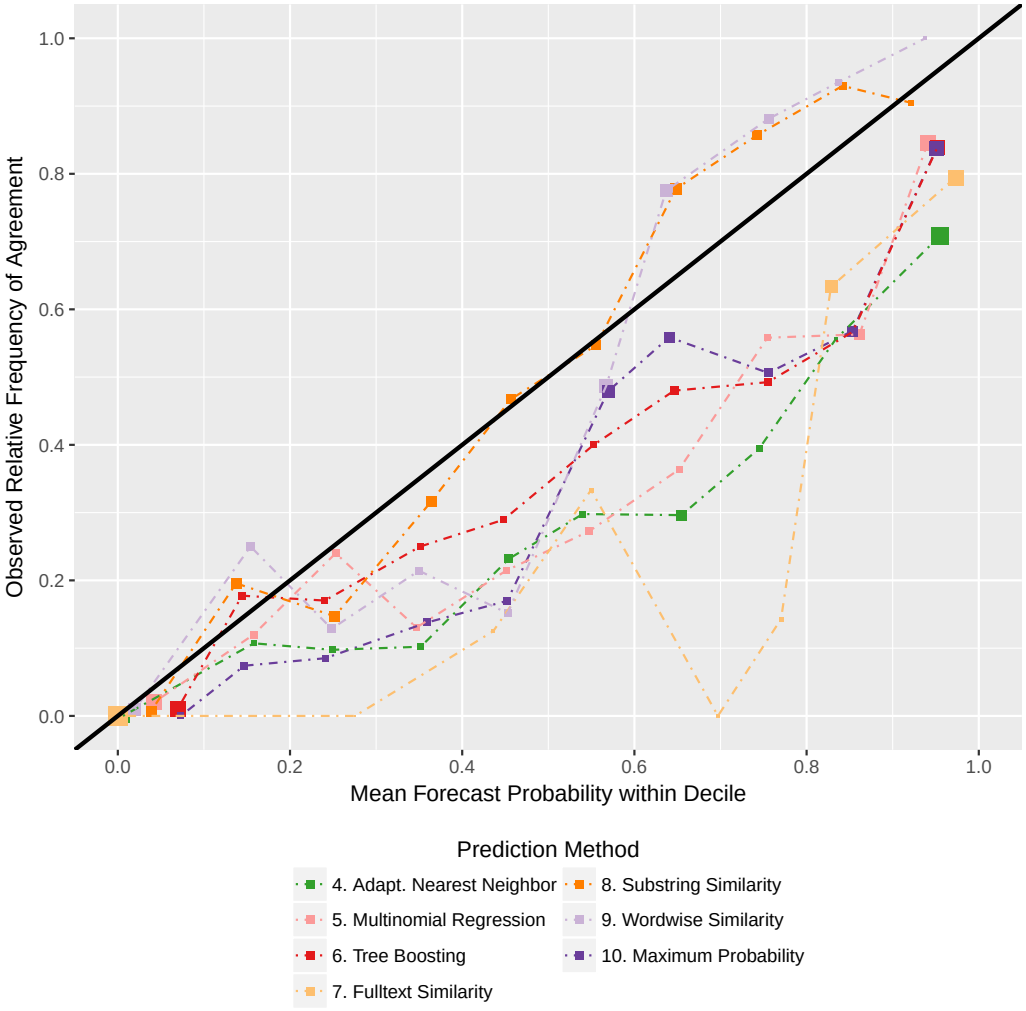


Figure A14: Reliability Diagram of most probable category (k=1); Training data: *Data set D* ($N = 6,575$), Test data: *Data set E* ($N = 1.064$)

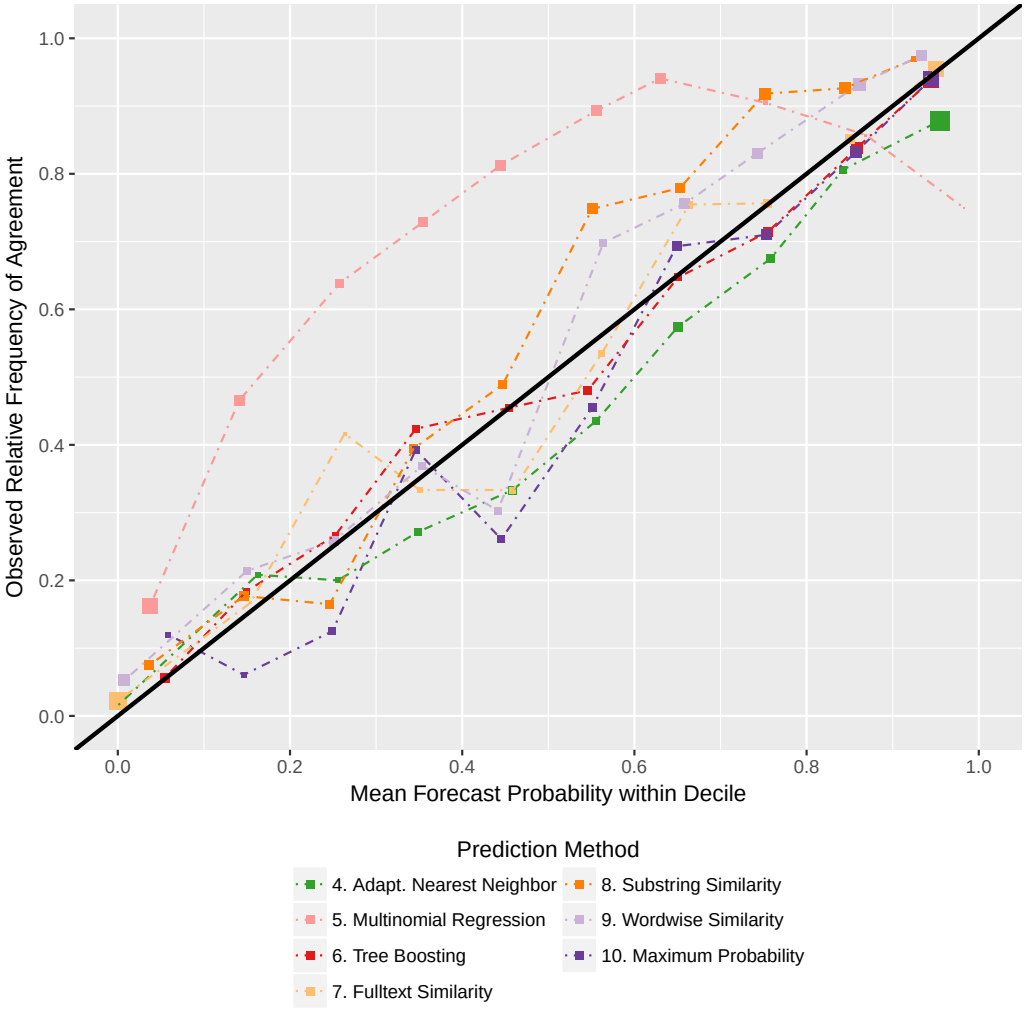


Figure A15: Reliability Diagram of most probable category (k=1); Training data: *all data sets combined* ($N = 144.444$), Test data: *all data sets combined* ($N = 1.064$)

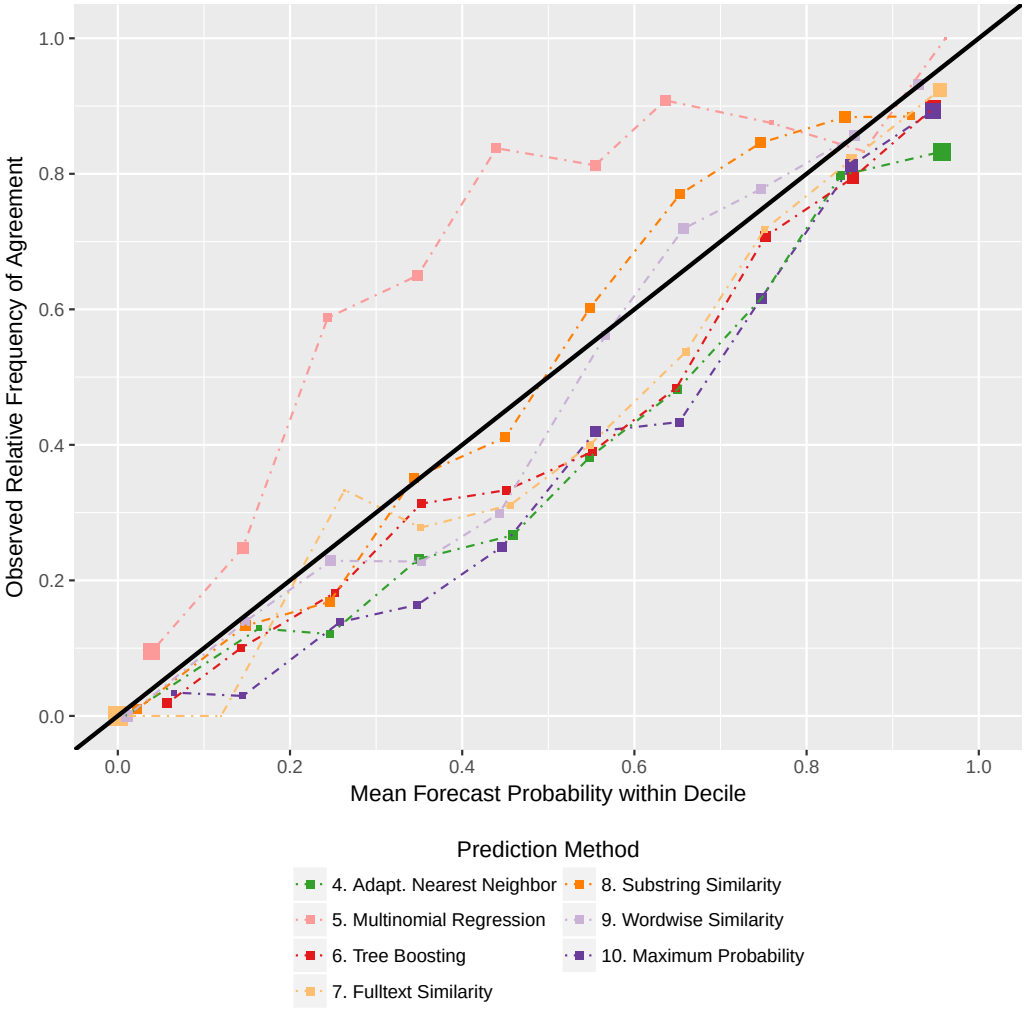


Figure A16: Reliability Diagram of most probable category (k=1); Training data: *all data sets combined* ($N = 144.444$), Test data: *Data set E* ($N = 1.064$)

4.3 Agreement Rate Among Top 5 vs. Production Rate (Computer-Assisted Coding)

Agreement Rates of five most probable category at various production rates (appropriate for interview coding)

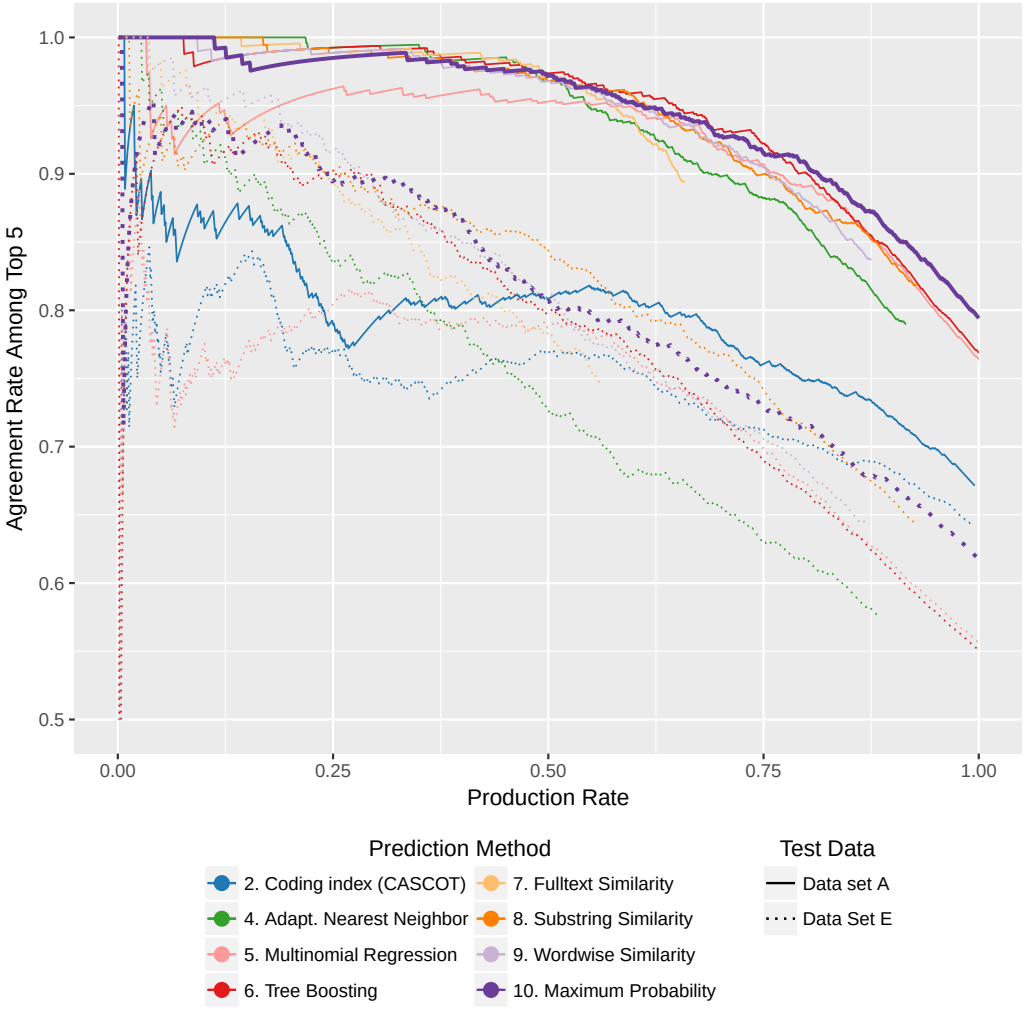


Figure A17: Agreement Rates of five most probable categories at various production rates (appropriate for interview coding); **Training data: *Data set A*** ($N = 31,867$)

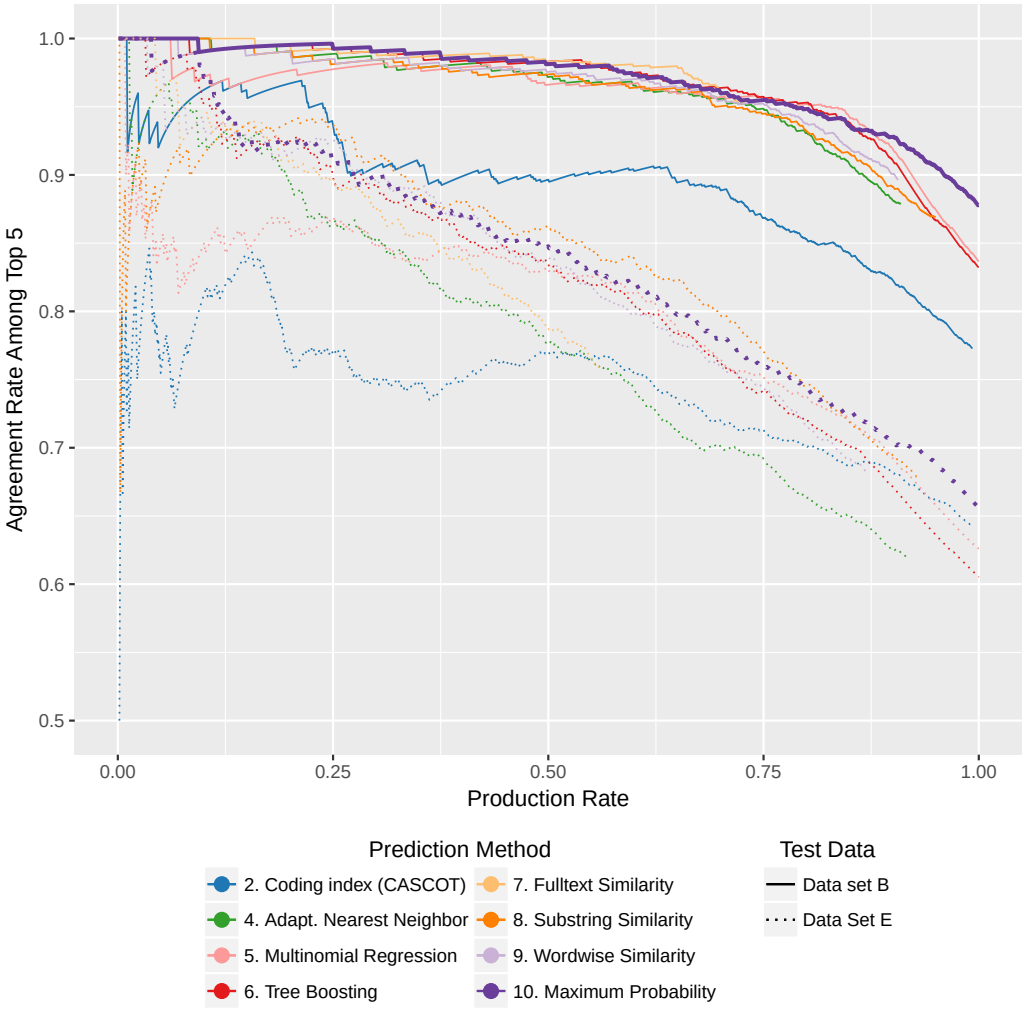


Figure A18: Agreement Rates of five most probable categories at various production rates (appropriate for interview coding); **Training data: *Data set B*** ($N = 54,880$)

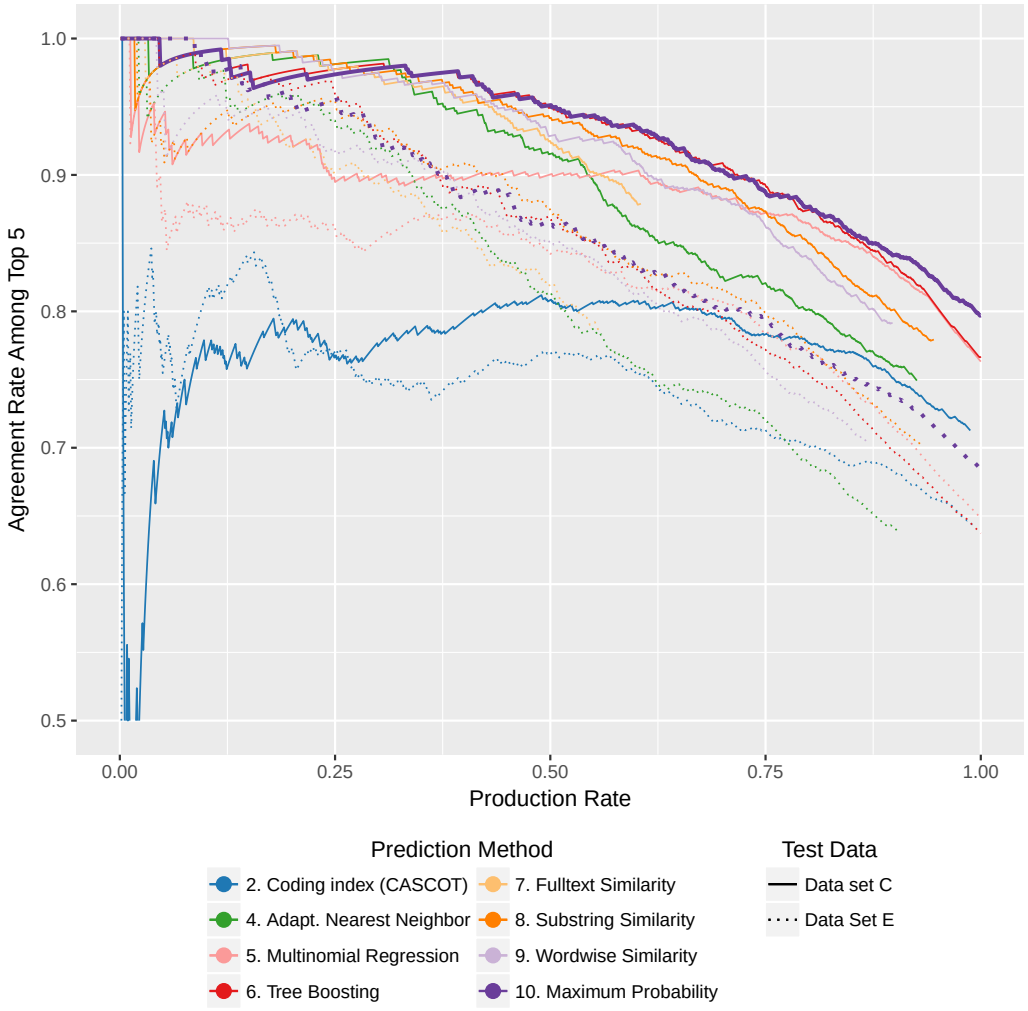


Figure A19: Agreement Rates of five most probable categories at various production rates (appropriate for interview coding); **Training data:** ***Data set C*** **(**$N = 47,930$**)**

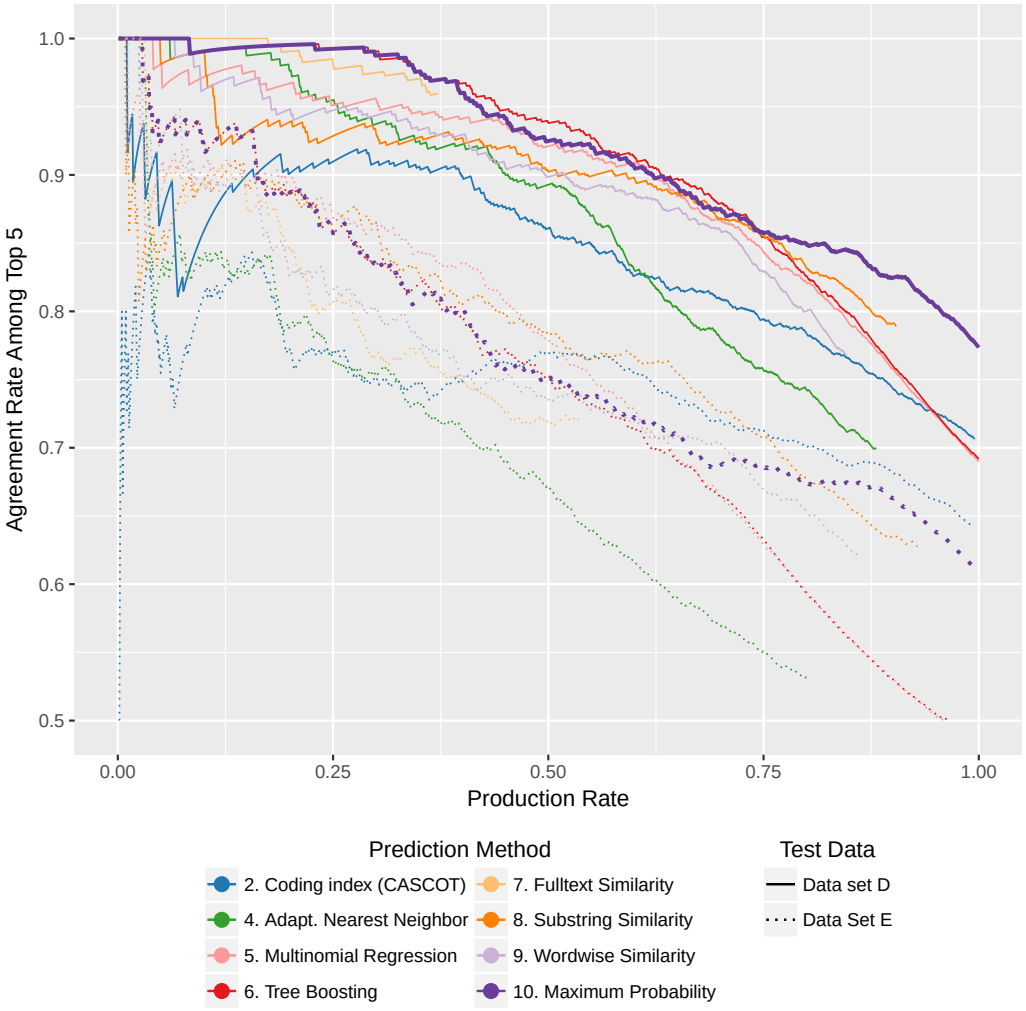


Figure A20: Agreement Rates of five most probable categories at various production rates (appropriate for interview coding); **Training data: *Data set D*** ($N = 6,575$)

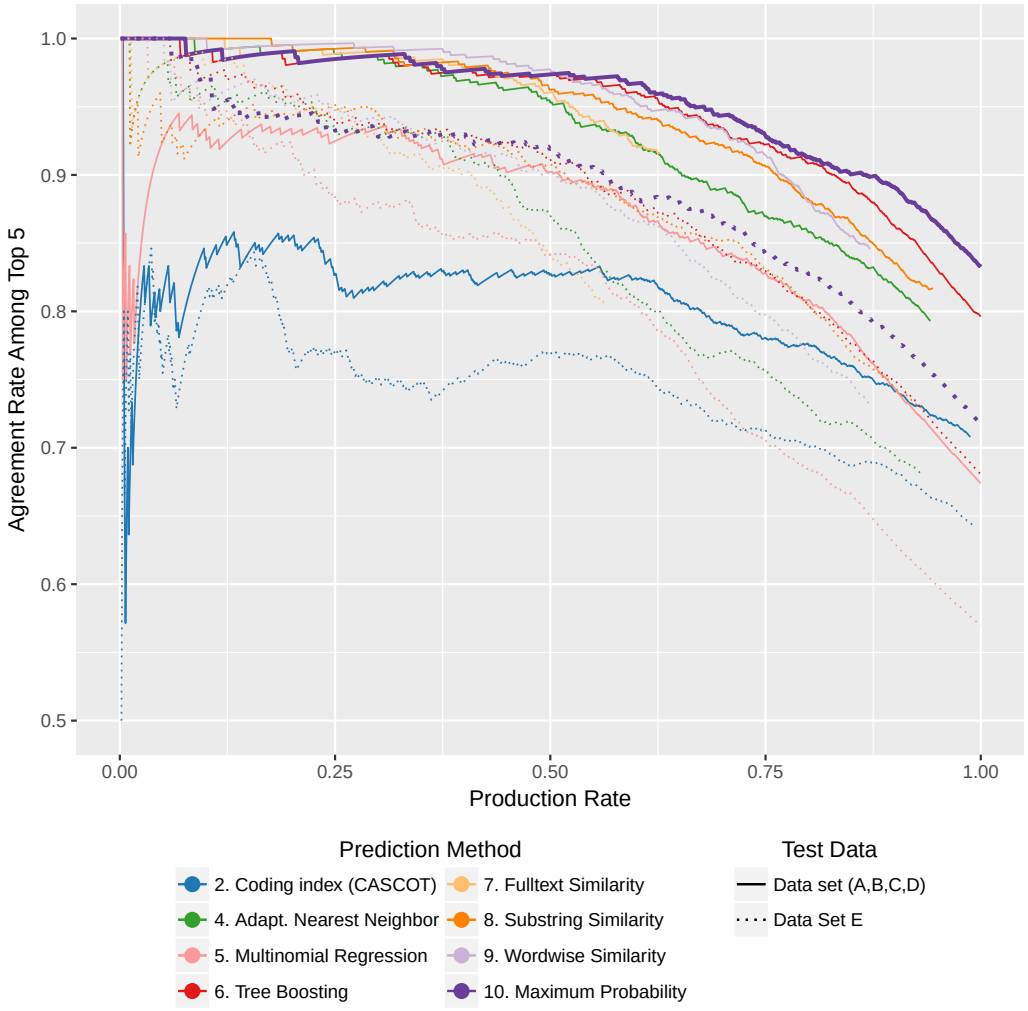


Figure A21: Agreement Rates of five most probable categories at various production rates (appropriate for interview coding); **Training data:** ***all data sets combined*** **(**$N = 144.444$**)**

4.4 True Positives among Top 5 vs. False Positives (Computer-Assisted Coding)

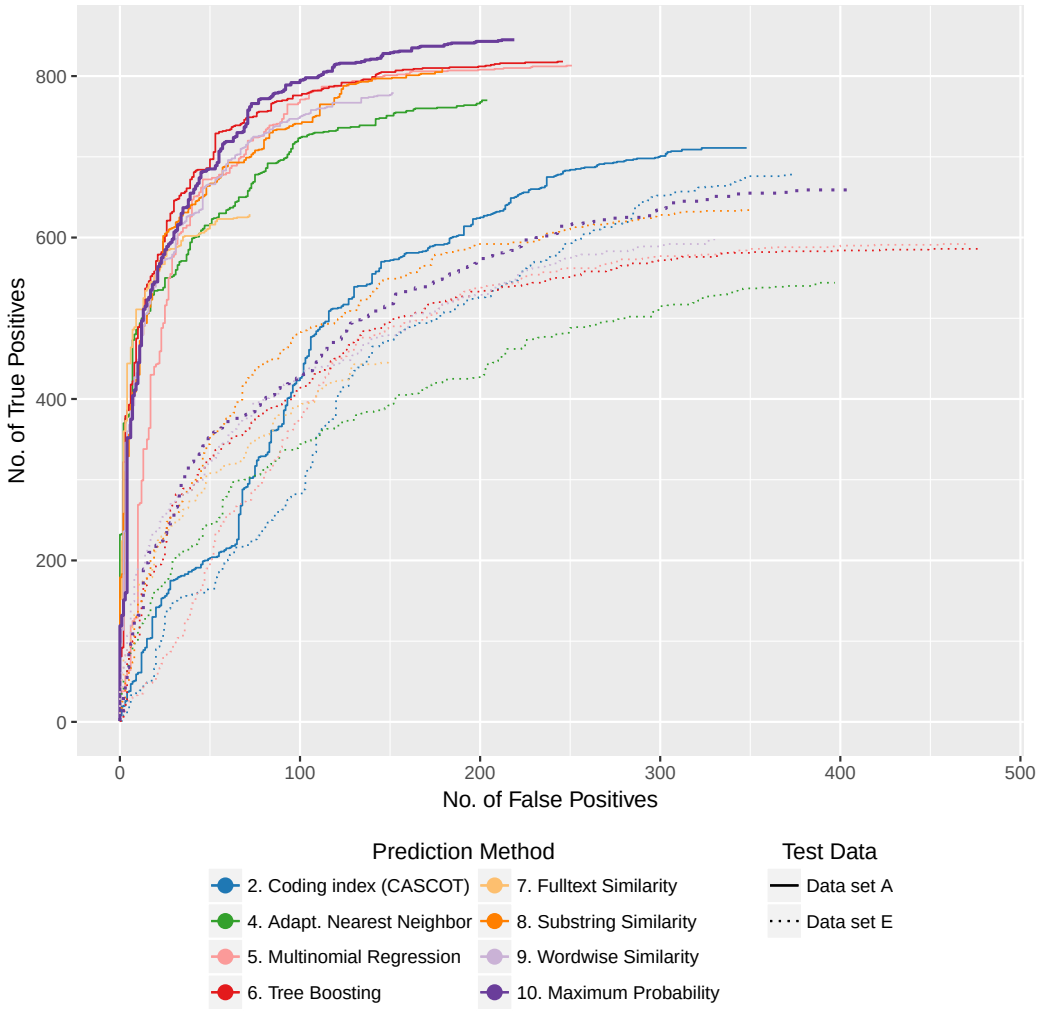


Figure A22: True positives vs. false positives ($k = 5$); **Training data: *Data set A*** **($N = 31,867$)**

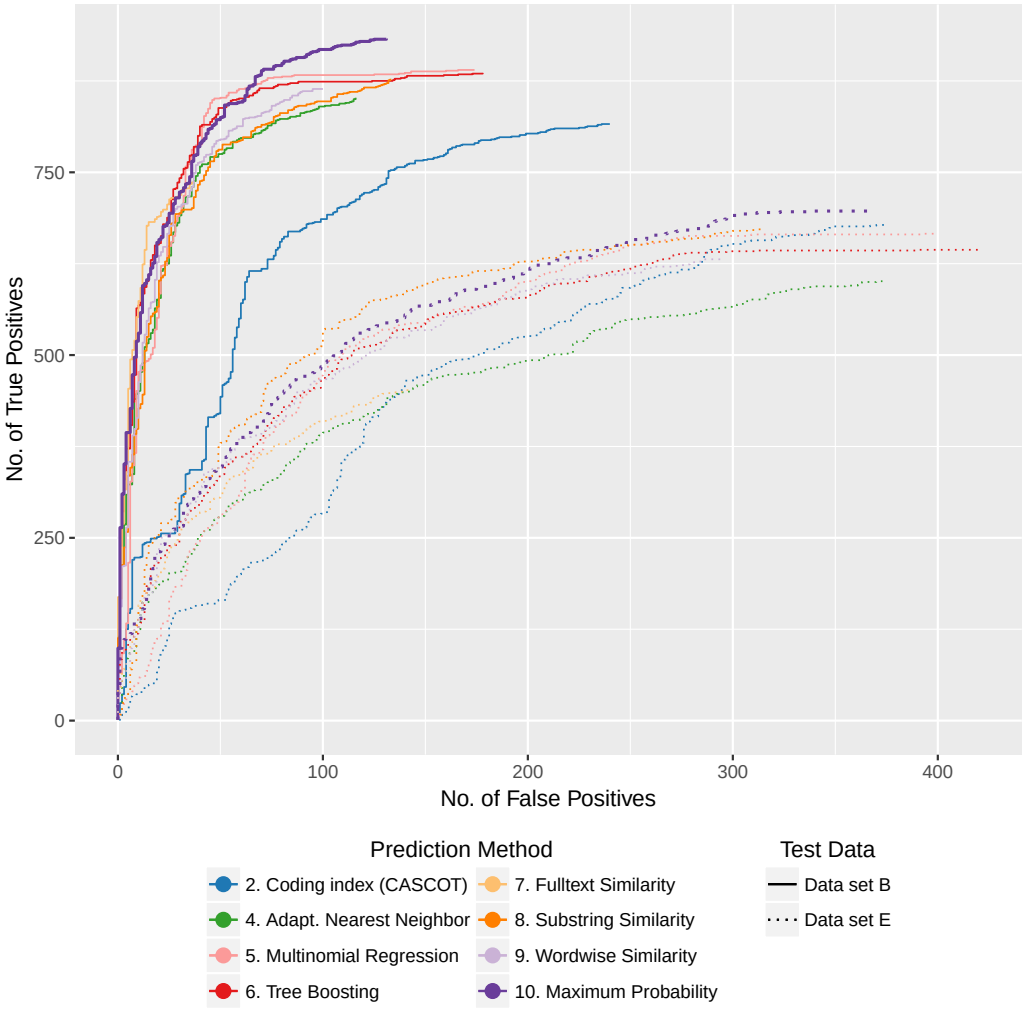


Figure A23: True positives vs. false positives ($k = 5$); **Training data: *Data set B*** **($N = 54,880$)**

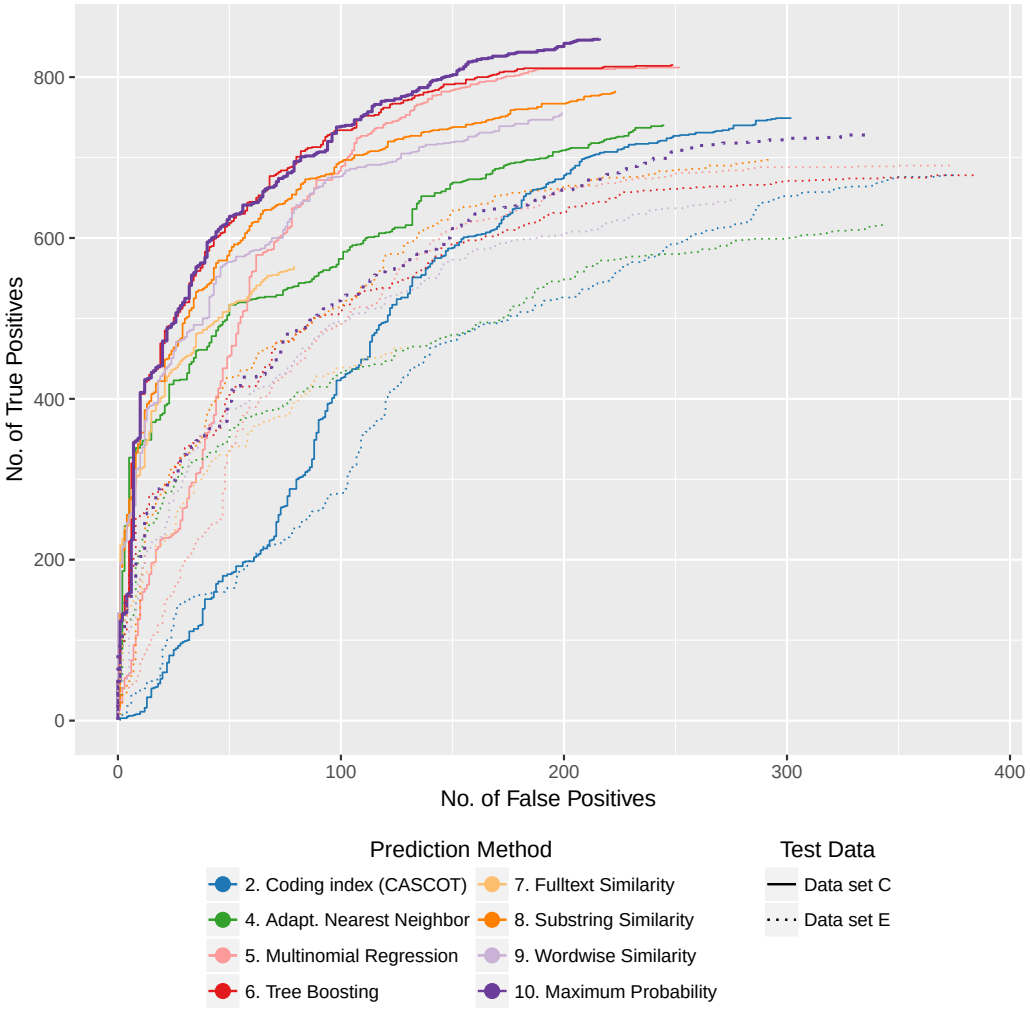


Figure A24: True positives vs. false positives ($k = 5$); **Training data: *Data set C*** **($N = 47,930$)**

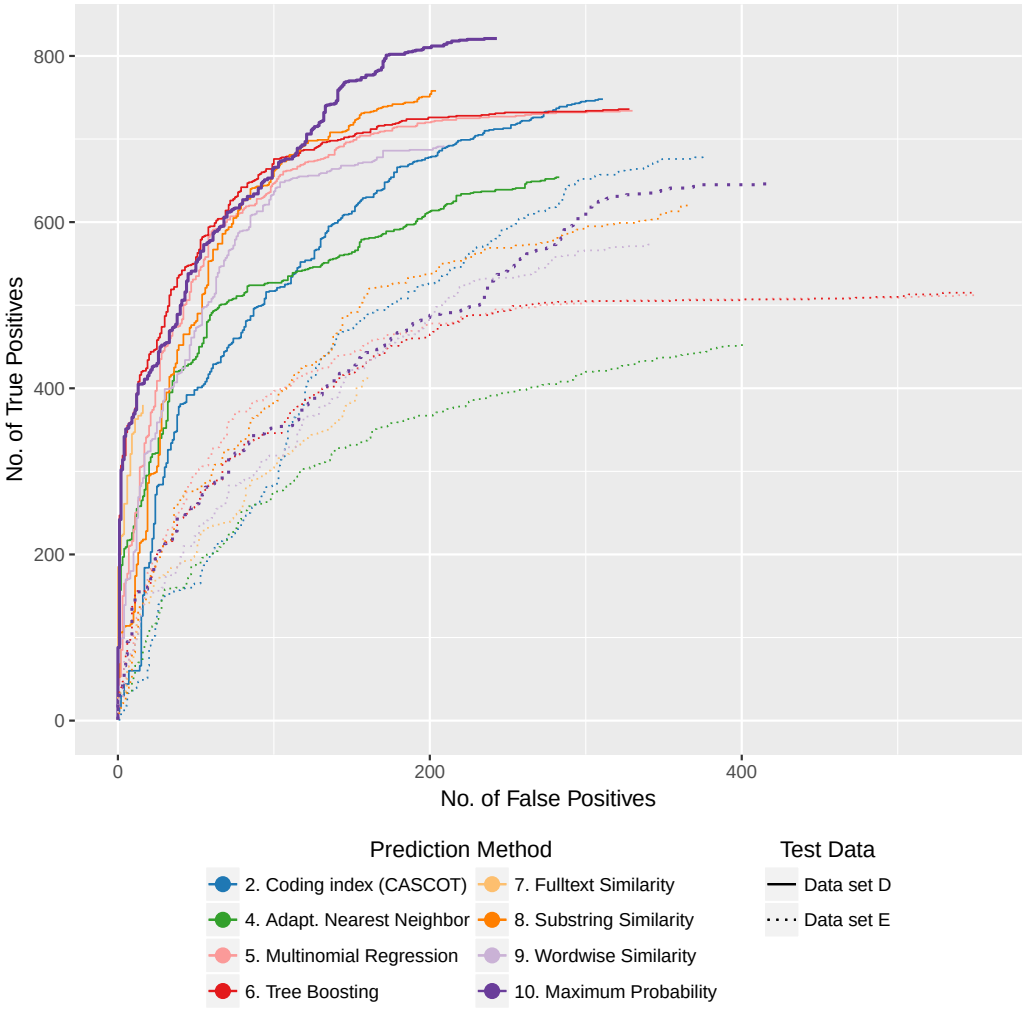


Figure A25: True positives vs. false positives ($k = 5$); **Training data: Training data: *Data set D*** **($N = 6,575$)**

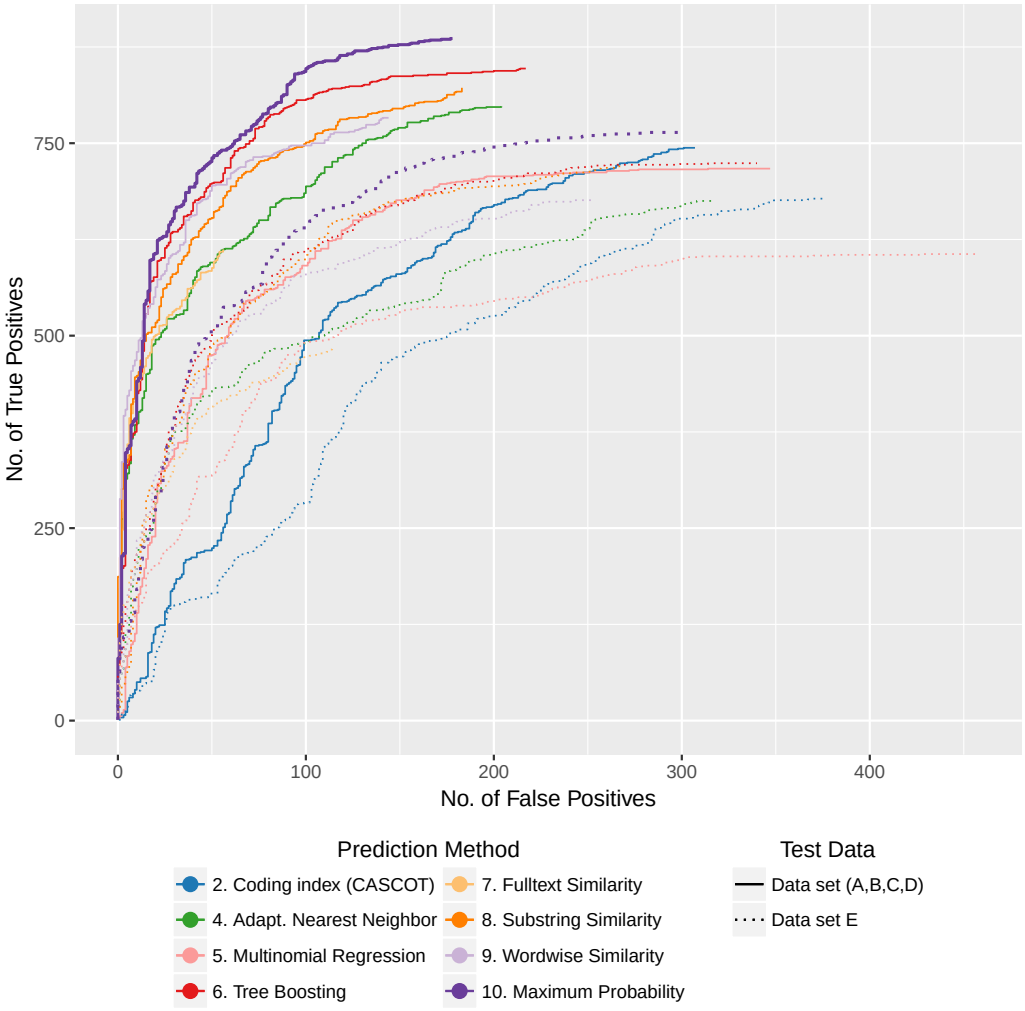


Figure A26: True positives vs. false positives ($k = 5$); **Training data: *all data sets combined*** **($N = 144.444$)**

4.5 Reliability Diagrams (k = 5)

Reliability Diagram: Ideal probabilistic predictions should match the observed relative frequencies, the diagonal; Point size is proportional to the number of observations within each bin. True value is positive if it is among the five most probable categories (appropriate for interview coding)

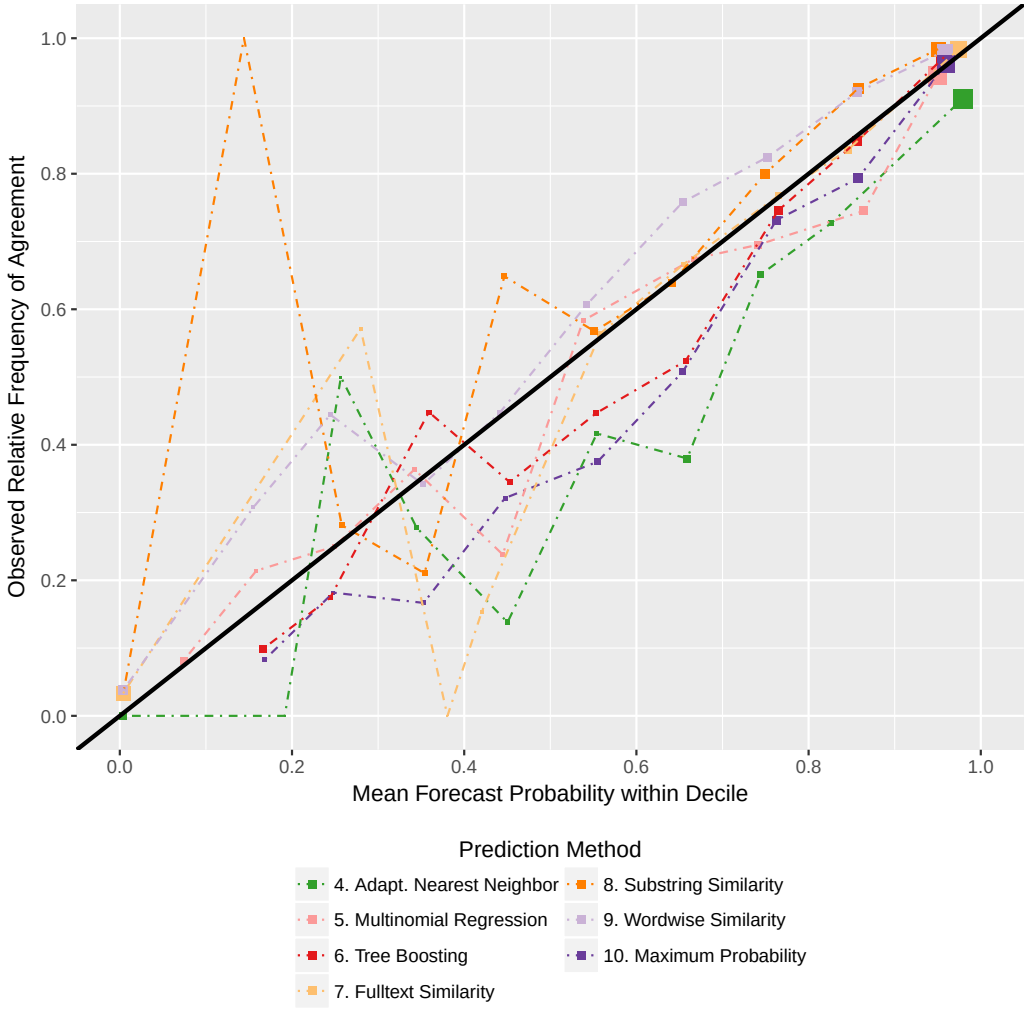


Figure A27: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set A* ($N = 31,867$), Test data: *Data set A* ($N = 1.064$)

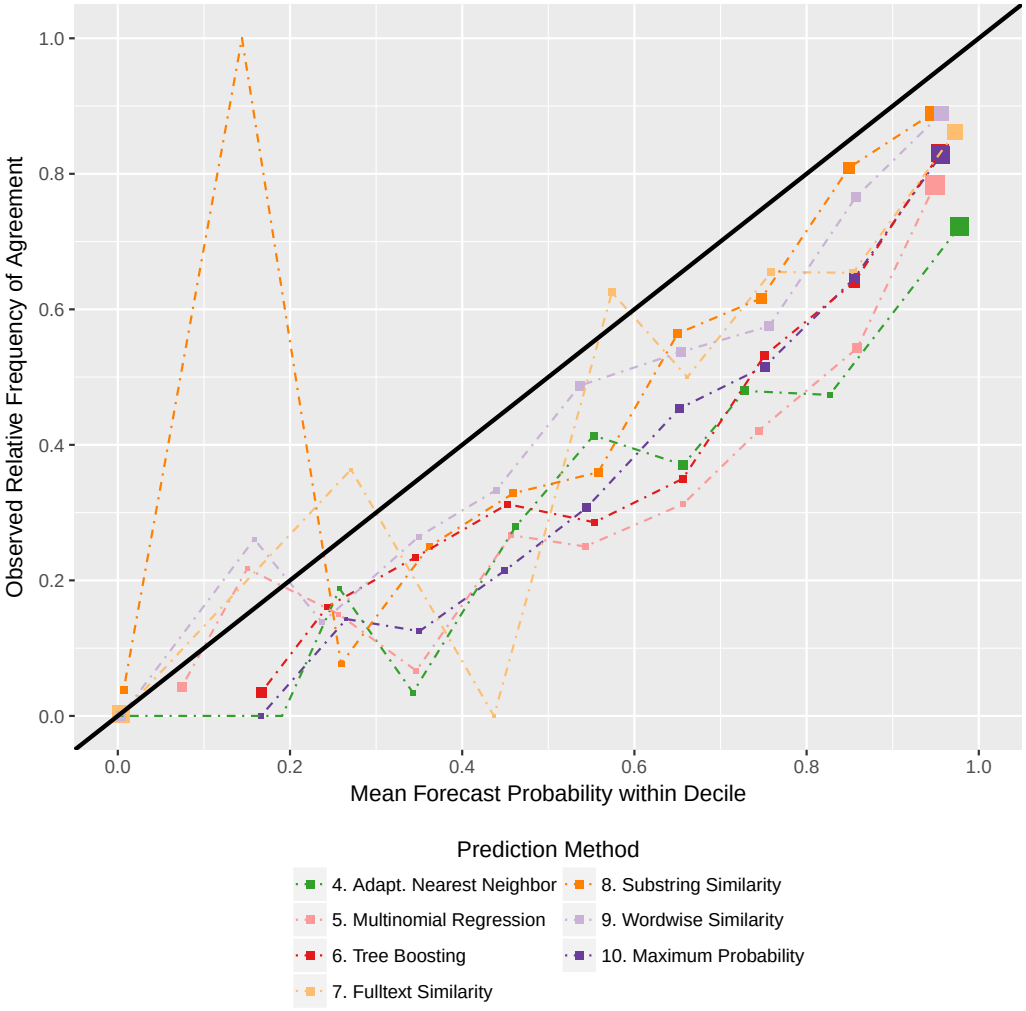


Figure A28: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set A* ($N = 31,867$), Test data: *Data set E* ($N = 1.064$)

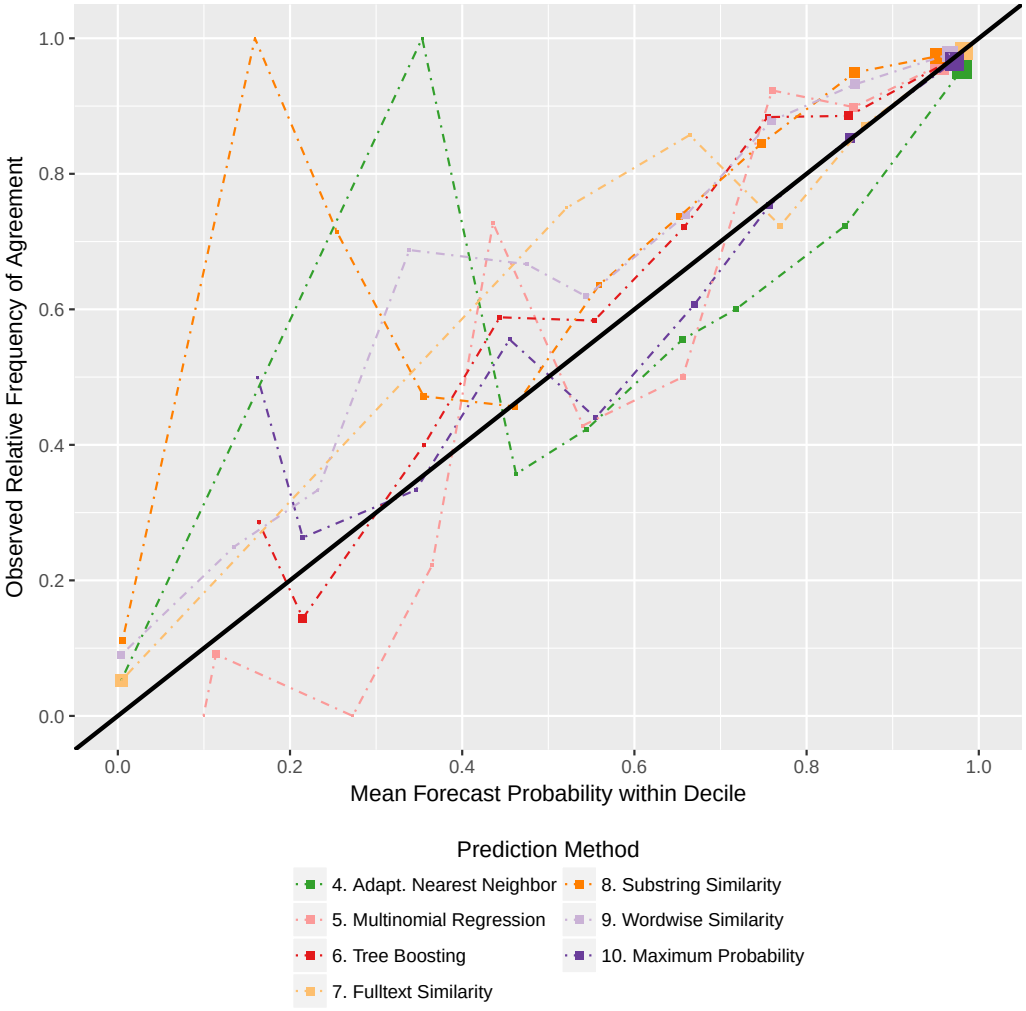


Figure A29: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set B* ($N = 54,880$), Test data: *Data set B* ($N = 1.064$)

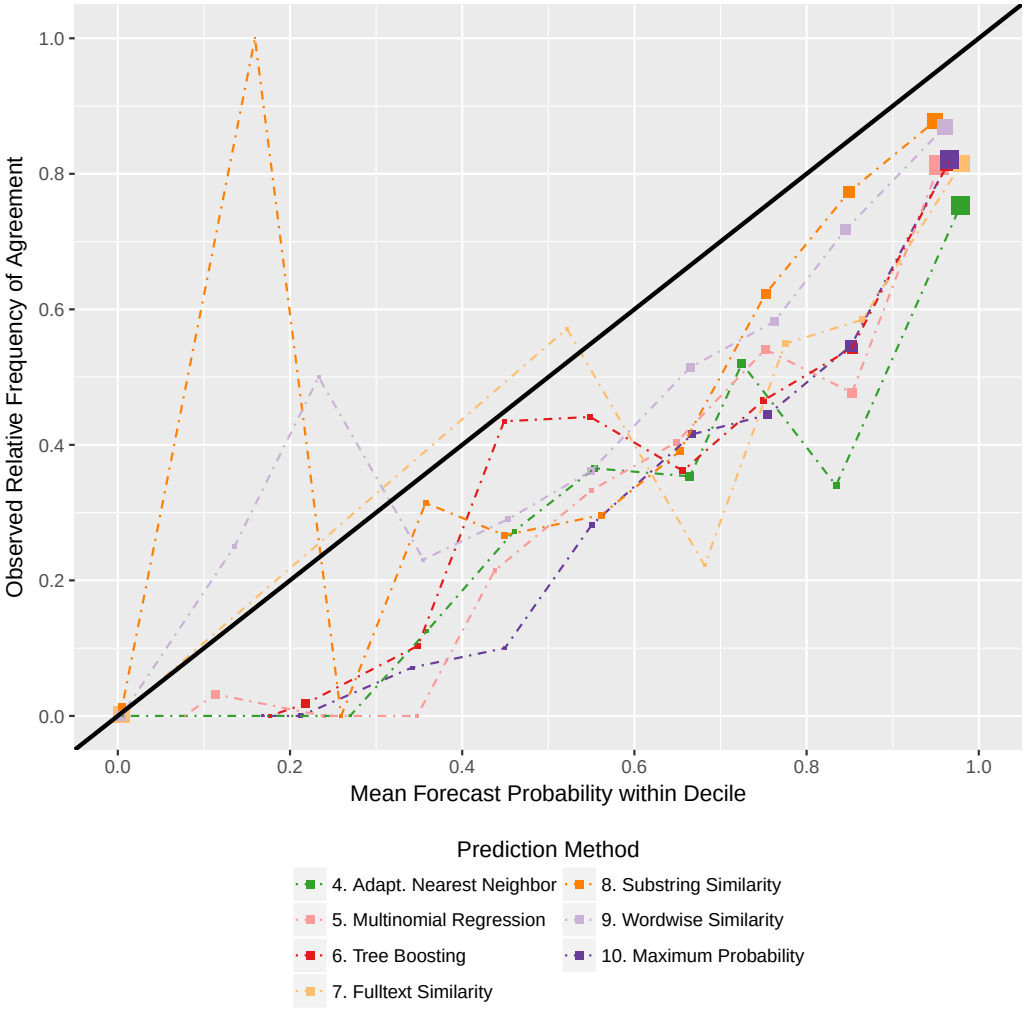


Figure A30: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set B* ($N = 54,880$), Test data: *Data set E* ($N = 1.064$)

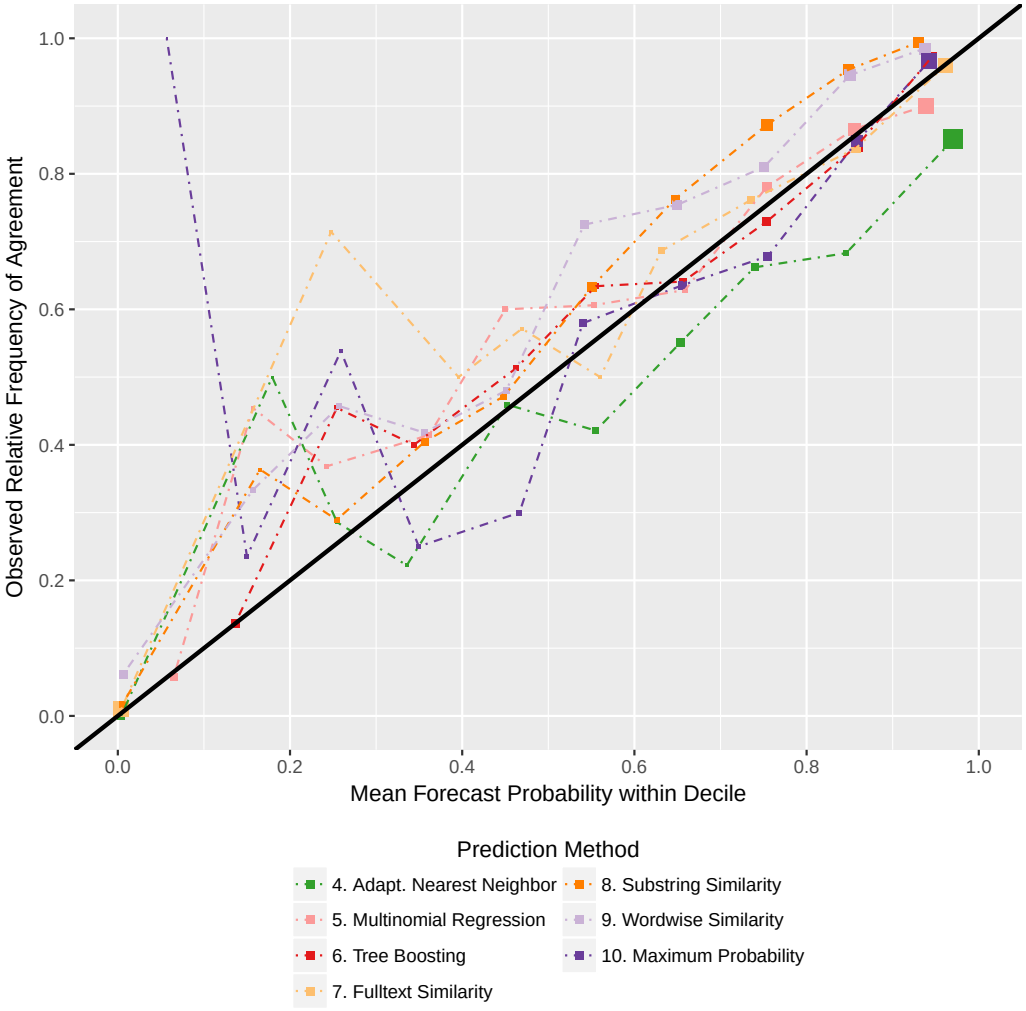


Figure A31: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set C* ($N = 47,930$), Test data: *Data set C* ($N = 1.064$)

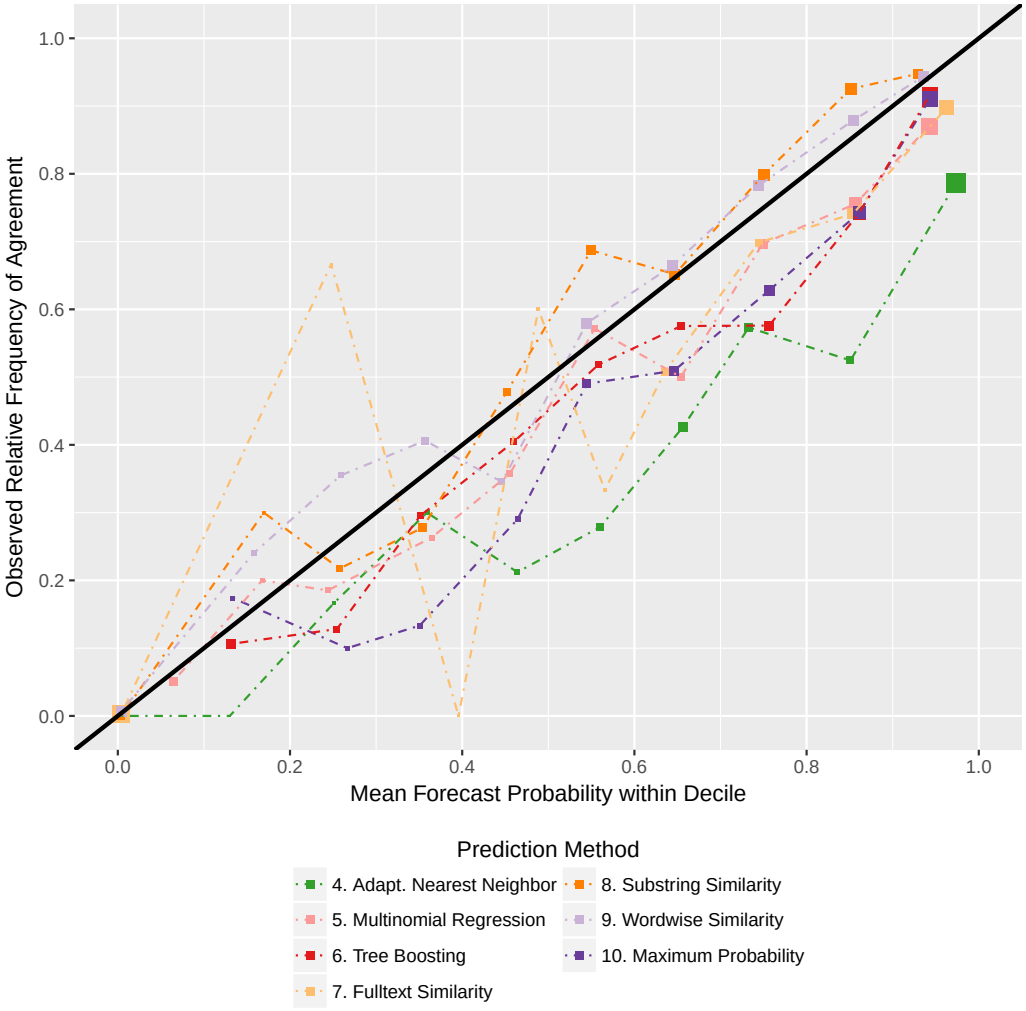


Figure A32: Reliability Diagram of $k = 5$ most probable categories; Training data:*Data set C* ($N = 47,930$), Test data: *Data set E* ($N = 1.064$)

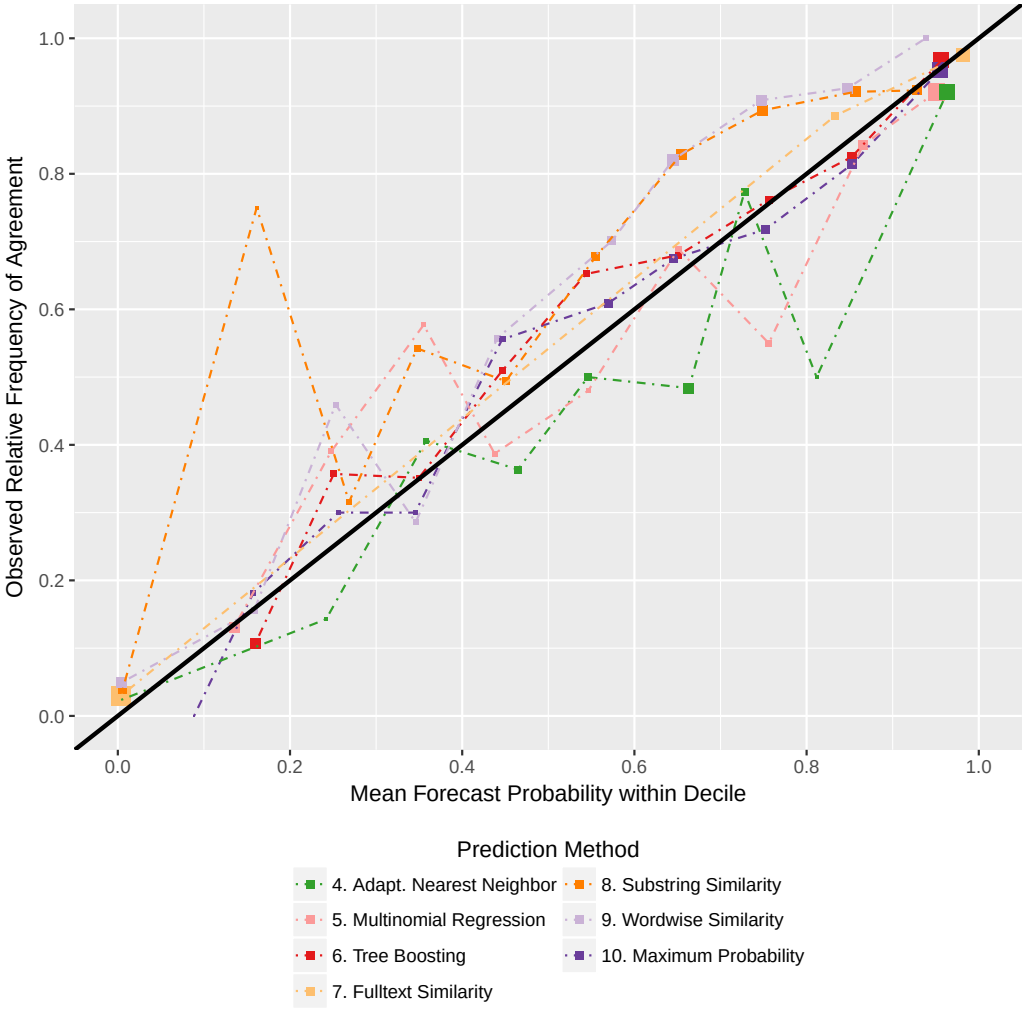


Figure A33: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set D* ($N = 6,575$), Test data: *Data set D* ($N = 1.064$)

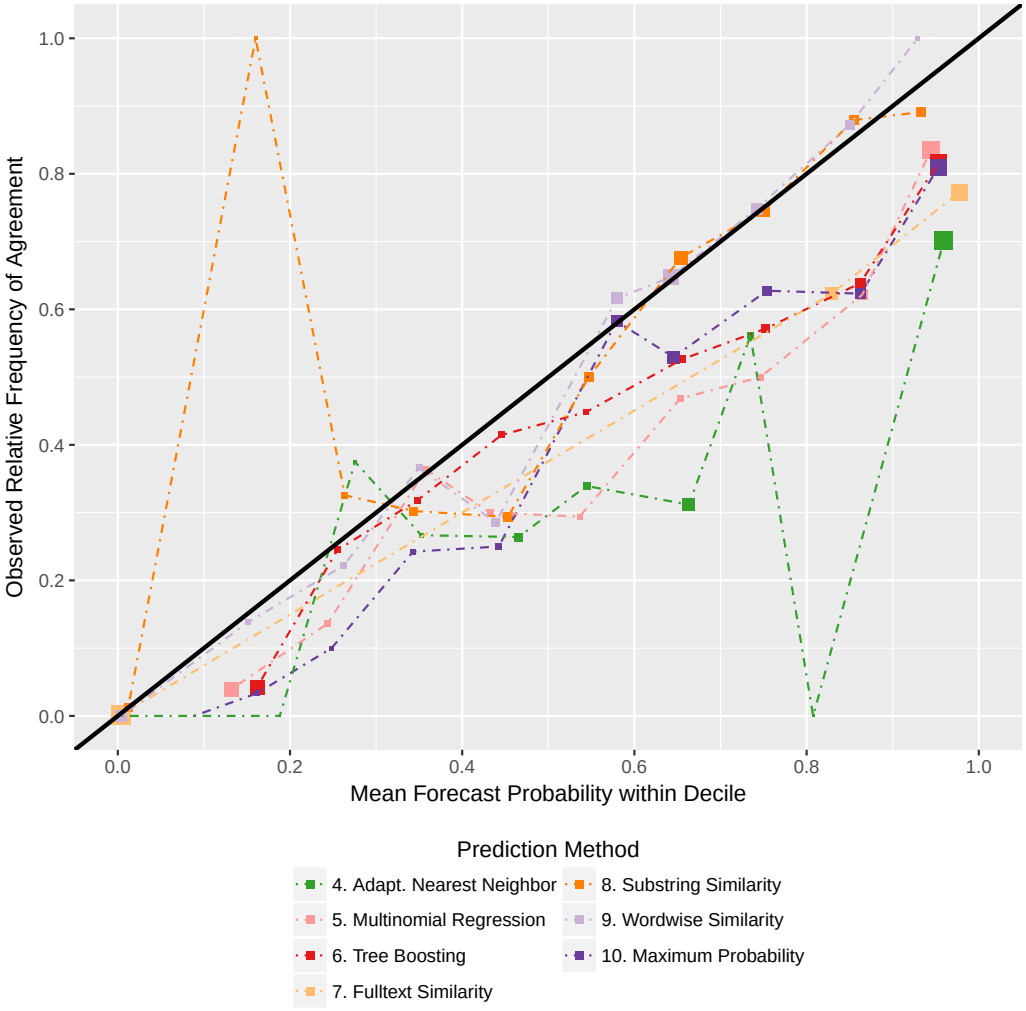


Figure A34: Reliability Diagram of $k = 5$ most probable categories; Training data: *Data set D* ($N = 6,575$), Test data: *Data set E* ($N = 1.064$)

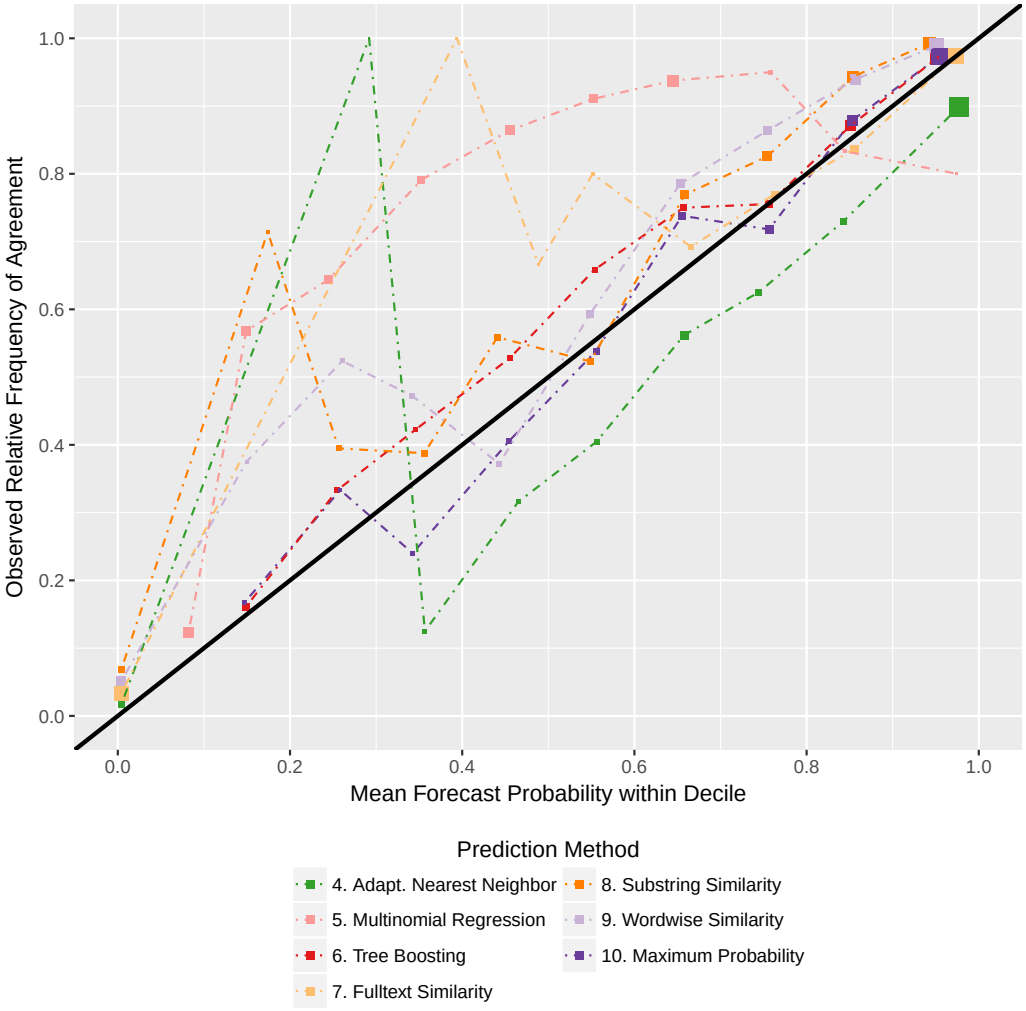


Figure A35: Reliability Diagram of $k = 5$ most probable categories; Training data: *all data sets combined* ($N = 144.444$), Test data: *all data sets combined* ($N = 1.064$)

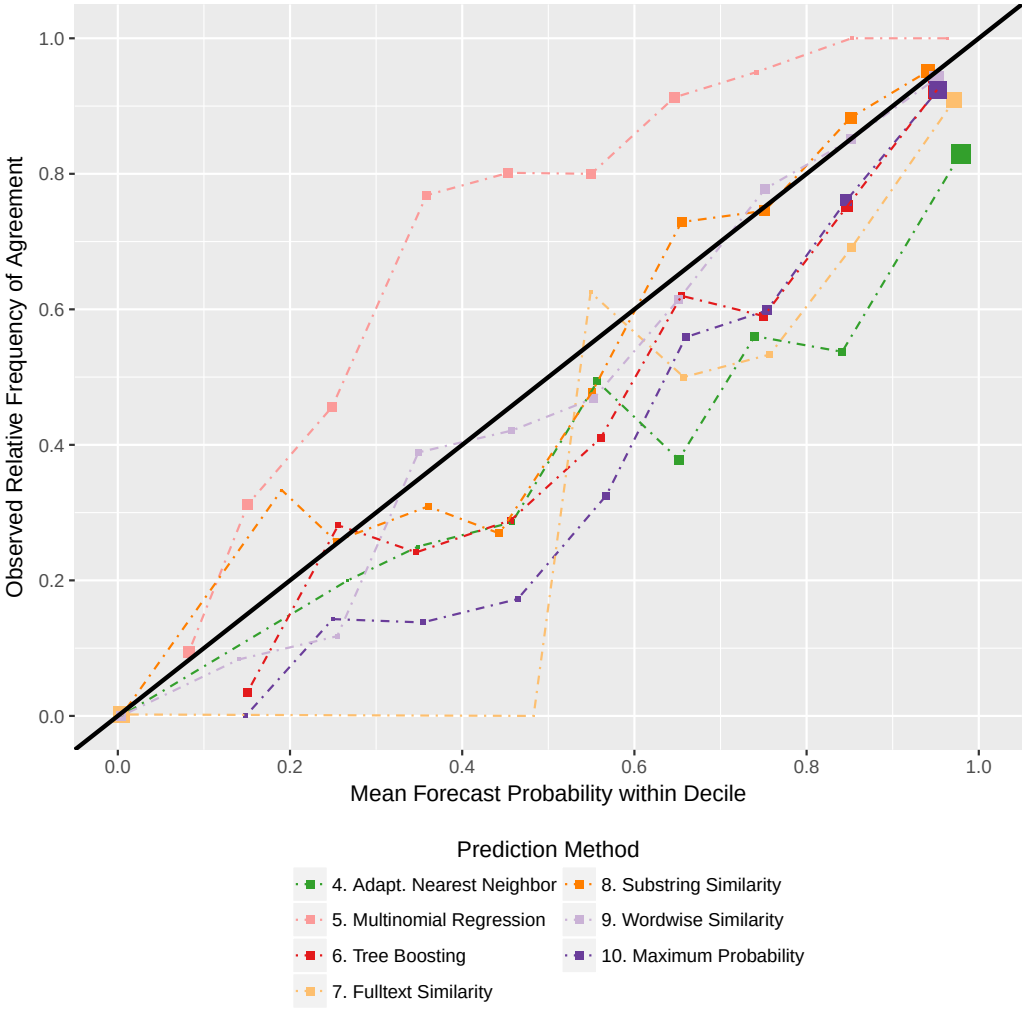


Figure A36: Reliability Diagram of $k = 5$ most probable categories; Training data: *all data sets combined* ($N = 144.444$), Test data: *Data set E* ($N = 1.064$)

4.6 Sharpness

Tree boosting has greater sharpness than the three similarity-based algorithms, but it is outperformed by the Maximum Probability algorithm (see Table A13).

Table A13: Sharpness ± standard errors [avg. bits]†

	Training data				
	(A)	*(B)*	*(C)*	*(D)*	*(A,B,C,D)*
	Test data are from the same data set as the training data.				
4. Adapt. Nearest Neighbor	1.58 ± 0.09	1.35 ± 0.09	1.70 ± 0.08	1.64 ± 0.10	1.43 ± 0.07
5. Multinomial Regression	2.63 ± 0.08	2.23 ± 0.08	3.17 ± 0.08	3.19 ± 0.09	6.64 ± 0.06
6. Tree Boosting (XGBoost)	2.78 ± 0.08	2.25 ± 0.08	3.32 ± 0.08	3.79 ± 0.09	2.98 ± 0.08
7. Fulltext Similarity	4.61 ± 0.13	3.39 ± 0.13	5.43 ± 0.13	6.78 ± 0.14	4.86 ± 0.13
8. Substring Similarity	3.55 ± 0.08	3.06 ± 0.07	4.39 ± 0.07	4.72 ± 0.08	3.81 ± 0.07
9. Wordwise Similarity	3.74 ± 0.10	2.95 ± 0.09	4.64 ± 0.08	5.42 ± 0.08	4.01 ± 0.09
10. Maximum Probability	2.29 ± 0.06	1.66 ± 0.05	2.93 ± 0.06	3.05 ± 0.07	2.51 ± 0.06
	Test data are from data set E.				
4. Adapt. Nearest Neighbor	2.01 ± 0.10	1.43 ± 0.09	1.96 ± 0.09	2.41 ± 0.12	1.67 ± 0.08
5. Multinomial Regression	3.07 ± 0.09	2.53 ± 0.08	3.55 ± 0.08	4.03 ± 0.09	6.78 ± 0.06
6. Tree Boosting (XGBoost)	3.51 ± 0.09	2.64 ± 0.08	3.71 ± 0.08	4.45 ± 0.10	3.16 ± 0.08
7. Fulltext Similarity	5.74 ± 0.13	5.10 ± 0.14	5.85 ± 0.13	5.41 ± 0.14	5.46 ± 0.14
8. Substring Similarity	3.92 ± 0.08	3.42 ± 0.08	4.63 ± 0.07	4.97 ± 0.07	3.96 ± 0.07
9. Wordwise Similarity	4.26 ± 0.09	3.67 ± 0.10	4.99 ± 0.08	5.62 ± 0.07	4.22 ± 0.09
10. Maximum Probability	2.74 ± 0.06	2.01 ± 0.06	3.18 ± 0.06	3.43 ± 0.07	2.65 ± 0.06

†Adapted Nearest Neighbor (algorithm 4) and Multinomial Regression (algorithm 5) frequently predict $\hat{p}_{n_f k} = 0$ for several categories k. $\hat{p}_{n_f k} \log_2 \hat{p}_{n_f k}$ is not defined in this case. In our implementation it is then set to 0, which explains the excellent sharpness of algorithms 4. and 5.

4.7 Logloss

The Maximum Probability algorithm is always best in terms of logloss. If the test data are from the same data set as the training data, Tree Boosting is the second best algorithm. If the test data are from data set E, the Substring Similarity algorithm comes in second (see Table A14).

Table A14: Logloss ± standard errors [avg. bits] †

	Training data				
	(A)	*(B)*	*(C)*	*(D)*	*(A,B,C,D)*
Test data are from the same data set as the training data.					
4. Adapt. Nearest Neighbor	∞	∞	∞	∞	∞
5. Multinomial Regression	∞	∞	∞	∞	4.36 ± 0.12
6. Tree Boosting (XGBoost)	2.92 ± 0.13	2.10 ± 0.11	3.36 ± 0.13	3.83 ± 0.14	2.89 ± 0.12
7. Fulltext Similarity	4.62 ± 0.16	3.42 ± 0.15	5.47 ± 0.15	6.75 ± 0.15	4.92 ± 0.16
8. Substring Similarity	3.15 ± 0.13	2.50 ± 0.11	3.58 ± 0.12	3.89 ± 0.13	3.13 ± 0.11
9. Wordwise Similarity	3.41 ± 0.14	2.48 ± 0.12	3.96 ± 0.13	4.56 ± 0.15	3.46 ± 0.13
10. Maximum Probability	2.73 ± 0.13	1.78 ± 0.11	3.10 ± 0.13	3.23 ± 0.14	2.56 ± 0.12
Test data are from data set E.					
4. Adapt. Nearest Neighbor	∞	∞	∞	∞	∞
5. Multinomial Regression	∞	∞	∞	∞	5.17 ± 0.13
6. Tree Boosting (XGBoost)	5.64 ± 0.17	4.85 ± 0.17	4.65 ± 0.15	6.16 ± 0.16	4.01 ± 0.14
7. Fulltext Similarity	6.72 ± 0.17	6.72 ± 0.18	6.51 ± 0.16	7.03 ± 0.17	6.26 ± 0.16
8. Substring Similarity	5.09 ± 0.16	4.66 ± 0.16	4.35 ± 0.14	5.49 ± 0.16	4.01 ± 0.13
9. Wordwise Similarity	5.66 ± 0.17	5.43 ± 0.17	4.93 ± 0.15	5.78 ± 0.16	4.61 ± 0.15
10. Maximum Probability	5.22 ± 0.18	4.65 ± 0.18	4.32 ± 0.15	5.25 ± 0.17	3.82 ± 0.15

†Adapted Nearest Neighbor (algorithm 4) and Multinomial Regression (algorithm 5) predict sometimes $\hat{p}_{nfk} = 0$ although category k was selected in the validation data ($y_{nfk} = 1$). Then, $\log_2 \text{loss} = \infty$

5 Part E: Similarity-based Reasoning: Connecting Approximate String Matching and a Hierarchical Bayesian Model

From a bird's eye view, similarity-based reasoning proceeds as follows. Start with a list of distinct job titles and training data. The basic idea is to calculate for each job title from the list the relative frequencies how often the job title was coded into various categories. Whenever a prediction is required for a new verbal answer that is identical with one of these job titles, output the relative frequencies.

We elaborate this idea in the following subsections, improving it in three ways. Firstly, we argue that relative frequencies are unsuited and Bayesian hierarchical models can be used to calculate more appropriate probabilities instead. Secondly, lists of job titles are often part of a coding index, which associate every job title with a category. We describe in the second subsection how this information can be used within the Bayesian hierarchical model to improve the results. Thirdly, the requirement that verbal answers from respondents and job titles from the list need to be identical is very strict and makes this technique useless for all non-identical verbal answers. This condition is loosened when we extend the Bayesian hierarchical models to account for various forms of similarity between verbal answers and job titles. We will need several probability density functions along the way, which are derived in a final subsection using standard Bayesian arguments.

A note about language: We will call each job title from the list a "cell". If verbal answers and job titles are identical/similar (i.e., the distance between the verbal answer and the cell c is below some predefined threshold), we say that the verbal answer is covered by cell c or, with the same meaning, the verbal answer belongs to cell c. This abstract language is used because our method is not dependent on how individuals (in our case their answers) and cells are set in relation. Many different ways to assign individuals into cells are conceivable.

5.1 A Bayesian hierarchical model if verbal answers and entries from the coding index are identical

In its simplest form, the algorithm searches the coding index for an entry that is identical to the respondent's verbal answer. If such an entry is found, common sense suggests that one should always, with probability one, assign the corresponding code. However, our experience from analyzing some data sets is that identical answers that match a single entry in the coding index, do not always have the same code, possibly because human coders had access to additional information from respondents, like a second verbal answer. It thus seems inappropriate to always assign the same code, but several different codes may be possible. Given that a verbal answer and an entry from the coding index are identical, how can one determine the probability distribution over possible categories?

The solution proposed here relies on training data, verbal answers that were collected and coded in previous studies. We write $y_{nc}^{(k)} \in \{0, 1\}$ to indicate that the verbal

answer from the n-th individual is identical with the c-th entry in the coding index and was/was not (1/0) coded into the k-th category (see Box 1 for a complete description of our notation). Let $\#\{y_c^{(k)}\} := \sum_{n=1,n\in c}^{N} y_{nc}^{(k)}$ denote the number of individuals whose verbal answers are covered by cell c, i.e., identical with the c-th entry from the coding index, and who were coded in category $y^{(k)}$. Further, $\#\{c\} := \sum_{k=1}^{K} \#\{y_c^{(k)}\}$ denotes the number of individuals whose answers are covered by cell c.

What is the probability distribution over categories? Assume that a future (indexed by f) verbal answer is covered by the c-th cell from the coding index, denoted by $C_f = c$. Using the subset of observations from the training data whose answers are covered by cell c, denoted by $\{y_{nc} : n \in c\}$, one can calculate relative frequencies how often each category realized within this set of identical verbal answers. This is the maximum likelihood estimate. To obtain the distribution over categories on how identical future verbal answers will be coded, a standard strategy would be to interpret the observed relative frequencies from the training data as probabilities, setting $p(y_{fc}^{(k)} = 1|C_f = c, \{y_{nc} : n \in c\}) = \frac{\#\{y_c^{(k)}\}}{\#\{c\}}$. However, this naive strategy only works well if $\#\{c\}$ is large. Yet, respondents are free to choose any verbal answer they like to describe their occupation, which means that many texts in the coding index are mentioned only a few times in the training data. If, for example, the answer "banker" was mentioned only a single time and coded into the k-th category, one would not want to code all future bankers into the c-th category with 100% probability, as the maximum likelihood estimate would suggest, because several other categories might be appropriate as well.

To circumvent this difficulty of the maximum likelihood estimate, we employ a Bayesian approach to estimate the posterior predictive distribution $p(y_{fc}^{(k)} = 1|C_f = c, y)$. The subsequent development of the Bayesian procedure was inspired by Gelman et al. (2014).

We assume that categories $\{y_{nc} : n = 1, ..., N; n \in c\}$ are drawn independently from a categorical distribution that is conditional on cell c and its parameters θ_c,

$$y_{nc}|C = c, \theta_c \overset{iid}{\sim} \text{Categorical}(\theta_{c1}, ..., \theta_{cK}) \tag{3}$$

Expressed in words, this modeling assumption means that only the cell c (the coding index entry) determines which category is being selected. The selection is not deterministic but this label-generating mechanism is random. The categorical distribution implies that each θ_{ck} may be interpreted as the conditional probability for outcome y if it is generated by cell c, $\theta_{ck} = p(y^{(k)} = 1|C = c)$. We believe this model is a good representation of what human coders do when they select a category based on a single-word input.

Box 1: Notation and Model Assumptions

Outcome variables
We have observed codes (or labels) $y = (y_1, ..., y_N)$ from $n = 1, ..., N$ individuals and will observe a future code y_f. The classification consists of K categories, $k = 1, ..., K$. Every observation y_n (and also the future y_f) is thus a vector of length K, $y_n = (0, ..., 1, ..., 0)$ where the k-th element equals 1 (i.e., $y_n^{(k)} = 1$), if the k-th code realizes and 0 otherwise. $y \in \{(y^{(1)}, ..., y^{(K)}) : y^{(k)} \in \{0, 1\}, \sum y^{(k)} = 1\}$ is the categorical outcome variable to be predicted.
To incorporate knowledge about which cell c, $c = 1, ..., C$, generates the outcome variable, we look at the conditional distributions $y_n|C_n = c$. To highlight the dependence on c, we often write y_{nc} or $y_{nc}^{(k)}$ ($\in \{0, 1\}$) as a shortcut. Note that C_n (from already observed individuals) and C_f (from a future individual) are subject-specific random variables.
For convenience we write $\#\{y_c^{(k)}\} := \sum_{n=1, n\in c}^{N} y_{nc}^{(k)}$ to denote the number of respondents whose verbal answers are covered by cell c and who were coded in category $y^{(k)}$. Further, $\#\{c\} := \sum_{k=1}^{K} \#\{y_c^{(k)}\}$ denotes the number of respondents whose answers are covered by cell c.

Parameters
$\theta = (\theta_1, ..., \theta_c, ..., \theta_C)$ is a vector of parameters that determines the distributions $y_n|C_n = c$ for all cells c. All θ_c are vectors of length K such that $\theta = (\theta_1, ..., \theta_c, ..., \theta_C) = ((\theta_{11}, ..., \theta_{1K}), ..., (\theta_{c1}, ..., \theta_{cK}), ..., (\theta_{C1}, ..., \theta_{CK}))$.

Hyperparameters
$\phi = (\phi_1, ..., \phi_K)$ are hyperparameters governing the distribution of each θ_c.

Model Assumptions

$$y_{nc}|C = c, \theta_c \overset{iid}{\sim} \text{Categorical}(\theta_{c1}, ..., \theta_{cK})$$

$$p(\theta_1, ..., \theta_C|\phi) = \prod_{c=1}^{C} p(\theta_c|\phi)$$

$$\theta_{c1}, ..., \theta_{cK}|\phi_1, ..., \phi_K \sim \text{Dirichlet}(\phi_1, ..., \phi_K)$$

$$p(\phi) \propto 1 \text{ or } p(\phi_R, \phi_{\bar{R}}) \propto 1 \text{ (see text)}$$

We now need to choose a prior distribution for θ_c. Without peeking at the data and saving the effort of analyzing and grouping $C > 50.000$ entries from the coding index, it is fair to assume symmetry among parameters $\theta_1, ..., \theta_C$. This is a main simplification in our model because index entries, of course, differ, but relevant differences are ignored in our prior beliefs and need to be learned from the data.

In probabilistic language, this symmetry is represented by an exchangeable distribution. A probability distribution $p(\theta_1, ..., \theta_C)$ is exchangeable if it is invariant to index permutations. Following (Gelman et al., 2014, p. 104ff.), the exchangeable distribution used by us is

$$p(\theta_1, ..., \theta_C|\phi) = \prod_{c=1}^{C} p(\theta_c|\phi) \tag{4}$$

The rationale for this formula is provided by de Finneti's theorem, which proves for $C \to \infty$ that any exchangeable distribution on $(\theta_1, ..., \theta_C)$ can be expressed as a mixture of independent and identical distributions. The distributions of individual components are assumed to follow a Dirichlet distribution,

$$\theta_{c1}, ..., \theta_{cK}|\phi_1, ..., \phi_K \sim \text{Dirichlet}(\phi_1, ..., \phi_K) \tag{5}$$

The Dirichlet prior is conjugate to the categorical data distribution. It was chosen in order to make our formulas analytically tractable and to simplify the computations. To reduce the number of free parameters in ϕ, we force all elements to be identical, $\phi_1 = ... = \phi_K$, signaling that we have a priori no preference for any particular category. An alternative, not yet further explored by us, might be to draw $\phi_1, ..., \phi_K$ from some exchangeable distribution. To reflect our ignorance about the unknown hyperparameter ϕ, a noninformative hyperprior distribution $p(\phi) \propto 1$ is used.

The model assumptions stated so far are standard within the context of Bayesian hierarchical models. It will be shown below how to derive the formulas that will be used for computation. Most importantly, we provide explicit formulas for the conditional posterior predictive distribution $p(y_{fc}^{(k)} = 1|C_f = c, \phi, y)$, Equation (20), and for the marginal posterior distribution $p(\phi|y)$, Equation (22). The distribution of central interest for our application, the posterior predictive distribution, is then obtained by integrating out the hyperparameter ϕ,

$$p(y_{fc}^{(k)} = 1|C_f = c, y) = \int p(y_{fc}^{(k)} = 1|C_f = c, \phi, y)p(\phi|y)d\phi \tag{6}$$

We use Monte Carlo integration to evaluate the integral (see Robert and Casella, 2004, for an overview). This means we need to draw B realizations $\phi^{(1)}, ..., \phi^{(B)}$ from the marginal distribution $p(\phi|y)$. The law of large numbers ensures that the quantity $\frac{1}{B}\sum_{b=1}^{B} h(\phi^{(b)})$ converges for large B towards the desired integral $\int h(\phi)p(\phi|y)d\phi$. Thus, larger B reduce approximation errors but need more time for computations. It can be controlled in our R package with the parameter `n.draw`.

This leads to the question how to draw realizations from the marginal distribution $p(\phi|y)$. Under certain conditions and if the sample size is sufficiently large, posterior

distributions can be approximated by a (multivariate) normal distribution with parameters $N(\hat{\phi}, (I(\hat{\phi}))^{-1})$. $\hat{\phi}$ is the mode of $p(\phi|y)$ and $I(\phi) = -\frac{d^2}{d\phi^2} \log p(\phi|y)$ is the observed Fisher information (Gelman et al., 2014, p. 83ff.). The mode is easily obtained with numerical optimization routines; explicit formulas for $I(\phi)$ are derived below (Equations 25 to 32). Based on graphical checks that the normal distribution approximates $p(\phi|y)$ very well, we decided to draw realizations $\phi^{(1)}, ..., \phi^{(B)}$ from this normal distribution. However, it still is an approximation and there may exist better ways to implement Monte Carlo integration, e.g. using the fact that $\frac{1}{B} \sum_{b=1}^{B} h(\phi^{(b)}) p(\phi^{(b)}|y) / N(\phi^{(b)}, \sigma^2)$ converges to $\int \frac{h(\phi)p(\phi|y)}{N(\phi,\sigma^2)} N(\phi, \sigma^2) d\phi$ as B approaches ∞.

Probabilities $p(y_{fc}^{(k)} = 1|C_f = c, y)$ for all cell-category-combinations are calculated in a model training phase. This makes it fast to predict the probability distribution for future values, because no computations are needed in the prediction phase when time is critical. It is only needed to find the desired cell c in the database and output its K probabilities. If no cell is found (= no index entry is identical with the input string) we output equal probabilities $p(y_{fc}^{(k)} = 1|C_f = c, y) = 1/K$ for all categories.

5.2 Balancing evidence from the coding index and evidence from training data

So far, we described the mathematics of similarity-based reasoning if every verbal answer is covered by a single cell (via identical string matching) and the cells are not associated with a category. In practice, our cells are entries from a coding index and most index entries are associated with a category. How can we use the information from the coding index about possible categories? How should we balance the evidence that a coding index gives about the most appropriate category against the evidence we have from our training data?

In the previous description, we forced all parameters in the parameter vector ϕ to be identical, entailing that no prior preferences are encoded for any particular category. When a coding index associates a cell (=index entry) with a particular category, we have a prior preference. Therefore, we set

$$\phi_k := \begin{cases} \phi_R & \text{if cell } c \text{ refers to category } k, \\ \phi_{\bar{R}} & \text{if cell } c \text{ does not refer to category } k. \end{cases} \tag{7}$$

ϕ_R and $\phi_{\bar{R}}$ are parameters that need to be estimated. The difference is that our model changes. Instead of drawing θ_c from independent and identical distributions, the new model draws θ_c from independent but not identical distributions. ϕ_R encodes for every cell its associated category. All other components in ϕ are still identical, as before.

Due to this differences, we implemented two separate routines in our software. One routine is used if the coding index contains associated categories and the other routine if not. The equations below remain the same with the exception of the observed Fisher information, which needs to be calculated differently.

5.3 Verbal answers and entries from the coding index are similar

So far, we required that respondents' verbal answers should be identical with an entry from the coding index, which avoids that an answer can be covered by more than one cell. However, since respondents are free to answer using any words they like and because spelling errors are always possible, their answers frequently will not match exactly an entry from the coding index. The technique described so far does not allow to make predictions for these cases.

The key innovation here is that we don't use identical string matching, but approximate string matching. We described three measures of string similarities in the main text that we found most useful; many more exist. What follows is not specific to any particular measure of string similarity.

We have an approximate match if the string similarity between a verbal answer and a coding index entry is above a certain threshold (or, equivalently, the distance between both strings is below a certain threshold). The consequence of approximate matching is that a verbal answer can match more than a single entry from the coding index. In the terminology used before, we say that answers can be covered by more than one cell. This raises methodological issues to be discussed next.

Hierarchical models, like the one presented above and summarized in Box 1, are well established in the literature (e.g. Gelman et al., 2014). However, these models usually require that each individual is covered by a single cell only. This assumption is not met anymore when individuals are assigned to all cells from the coding index that are similar enough to the answer given. Since a verbal answer can be covered by multiple cells, our modeling strategy needs some adaptations.

One proposal to adapt hierarchical models for this purpose, developed in the field of educational statistics, are multiple membership models (Hill and Goldstein, 1998, Browne et al., 2001, Goldstein, 2011). These models calculate a weighted mean over all relevant cell parameters θ_c. Observed values are then drawn from a distribution, $y_n|\theta \sim \text{Categorical}(h(\sum_{c=1}^{C} w_{n,c}\theta_c))$ or similar, substituting our original equation (3). This formulation requires the researcher to specify weights $w_{n,c}$ in advance, that may, for example, represent the proportional time an individual n has spent at different schools c or the researcher's probabilistic belief that individual n belongs to cells c. In our situation, the obvious choice would be to give equal weights to all possible cells, $w_{n,c} = \frac{1}{\text{No. of similar cells}}$ if the c-th cell is similar to the n-th verbal answer and $w_{n,c} = 0$ otherwise. Yet, this approach is rather simplistic and neglects the possibility that, given data about the correlational structure between cells c and outcome variables y, one may also infer better probabilities of cell membership.

We employ two separate strategies to deal with answers that are covered by more than one cell. The first strategy is used for training; the second strategy is used for prediction.

The strategy used for training is a naive simplification. The model above requires that only a single cell c will create an outcome y_{nc}. However, due to the multiple membership problem we do not know which cell generates an outcome. For this reason, we transform the data and repeat the outcomes as often as necessary. For example, if cells $3, 5$, and 8 are similar to a respondent's verbal answer and may determine its outcome (=cover the respondent), we will treat the single outcome from this respondent

as if we had three outcomes from three distinct respondents, the first one covered only by cell 3, the second one covered only by cell 5, and the third one covered only by cell 8. This brings us in the same situation as before when we considered identical string matching. We admit that this strategy may not be optimal. During development we tried alternative approaches, but the calculations were too slow to be useful.

The second strategy deals with the prediction phase. The challenge is that we have from Equation 6 a posterior predictive distribution $p(y_{fc}^{(k)} = 1|C_f = c, y)$ that is conditional on cell $C_f = c$. For answers being covered by more than one cell, a range of different cells might generate the outcome and multiple posterior predictive distributions are predicted. How can we choose a single posterior predictive distribution to use?

The posterior predictive distribution $p(y_{fc}^{(k)} = 1|C_f = c, y)$ is appropriate only if we know exactly which cell will generate the future value y_f. However, if the verbal answer is similar to many cells, several cells become candidates. In this case, we take a weighted average over all candidate posterior predictive distributions,

$$p(y_{fc}^{(k)} = 1|y) = \sum_{c=1}^{C} p(y_{fc}^{(k)} = 1|C_f = c, y)p(C_f = c|y) \tag{8}$$

This approach draws inspiration from the Bayesian Model Averaging literature (see Hoeting et al., 1999, for an overview), which proposes averaging over posterior distributions from different plausible models. The central formula from that literature closely resembles Equation (8), with the single difference that we condition on $C_f = c$ whereas the Model Averaging literature would write C without the subscript $_f$. In that context, C refers to different probabilistic models under consideration and all probabilities are implicitly dependent on the models' assumptions.

In our situation the interpretation of C_f is somewhat different. The reason is that Hoeting et al. (1999) describe global models, which are applicable for all individuals in the data. By contrast, our models are local in the sense that we use separate submodels for each cell and C_f refers to a submodel. The submodels are not applicable for all individuals due to the hierarchical supermodel we use. We cannot apply standard Model Averaging techniques as a consequence.

We write C_f with a subscript to highlight its dependence on a future individual f. The verbal answer from this individual will be needed to determine the possible cells (submodels). If the individual's verbal answer is covered by a single cell c, we can safely set $p(C_f = c|y) = 1$, making C_f a deterministic variable as it was with identical matching above. In other situations C_f is not observable and we regard it as a random variable, which determines the random mechanism (cell c) that will generate the future outcome y_f.

It is possible to some extent to model the cell selection mechanism C_f of a future observation based on previous observations. Based on highly pragmatic reasoning, we set

$$p(C_f = c|y) \begin{cases} = 0 & \text{if dissimilar} \\ \propto \sqrt[\#\{c\}]{p(\{y_{nc} : n \in c\}|\hat{\phi})}, & \text{if similar and } \#\{c\} \geq 1 \\ \propto 0.0000001, & \text{if similar and } \#\{c\} < 1 \end{cases} \tag{9}$$

and divide it by the normalizing constant $\sum_{c=1}^{C} p(C_f = c|y) = 1$ to sum to 1.

$\hat{\phi} = \arg\max_\phi p(\phi|y)$ is the posterior mode of the marginal posterior density $p(\phi|y)$ (Equation 22) and $p(\{y_{nc} : n \in c\}|\phi)$ is the conditional prior predictive density (Equation 18).

The formula above describes what actually is calculated within our algorithm. The reasoning about $p(C_f = c|y)$ is as follows. Mathematically, we have

$$p(C_f = c|y) = \frac{p(\{y_{nc} : n \in c\}|C_f = c, \{y_{nc} : n \notin c\})p(C_f = c|\{y_{nc} : n \notin c\})}{\sum_{c'=1}^{C} p(\{y_{nc'} : n \in c'\}|C_f = c', \{y_{nc'} : n \notin c'\})p(C_f = c'|\{y_{nc'} : n \notin c'\})} \tag{10}$$

Consider the first case in Equation (9). If a verbal answer and an entry from the coding index are not similar, the term $p(C_f = c|\{y_{nc} : n \notin c\}) = p(C_f = c)$ is set to zero as nothing would suggest that this cell will generate the future outcome y_f. Using Equation 10, we see that $p(C_f = c|y)$ becomes zero as well.

Now consider the second case in Equation (9). If the verbal answer and the cell are similar, we still have no preference which of the similar cells will generate the future outcome y_f and set $p(C_f = c)$ to $\frac{1}{\text{No. of similar cells}}$. The same term also appears in the denominator and $p(C_f = c|\{y_{nc} : n \notin c\}) = p(C_f = c)$ cancels out (likewise in case three below).

The term $p(\{y_{nc} : n \in c\}|C_f = c, \{y_{nc} : n \notin c\})$ requires more attention. Since the observed values $\{y_{nc} : n \in c\}$ cannot depend on future cells C_f, this is equal to

$$\begin{aligned} p(\{y_{nc} : n \in c\}|\{y_{nc} : n \notin c\}) = p(y_{1c}|\{y_{nc} : n \notin c\})\cdot \\ \cdot p(y_{2c}|y_{1c}, \{y_{nc} : n \notin c\}) \cdot ...\cdot \\ \cdot p(y_{Nc}|y_{1c}, ..., y_{(N-1)c}, \{y_{nc} : n \notin c\}) \end{aligned} \tag{11}$$

which is a product over probabilities of $\#\{c\}$ individuals that are covered by cell c. As a result, cells covering many individuals will have low probabilities and cells covering few individuals will have high probabilities. If we would insert this in Equation (10), cells covering the fewest individuals (often just one) would in general have highest weights $p(C_f = c|y)$. This is undesired as these are exactly the cells that provide the least evidence about its internal random mechanism. Also, our interest with $p(\{y_{nc} : n \in c\}|C_f = c, \{y_{nc} : n \notin c\}$ is not necessarily to obtain the joint probability of all $\{y_{nc} : n \in c\}$, but the probability of one typical realization from $\{y_{nc} : n \in c\}$. Since the joint probability can be written as a product of $\#\{c\}$ terms, the geometric mean $\sqrt[\#\{c\}]{p(\{y_{nc} : n \in c\}|\{y_{nc} : n \notin c\})}$ appears to be an appropriate choice.

The full Bayesian approach to calculate $p(\{y_{nc} : n \in c\}|\{y_{nc} : n \notin c\})$ would use numerical integration and proceed parallel to Equation (6) to determine

$$p(\{y_{nc} : n \in c\}|\{y_{nc} : n \notin c\}) = \int p(\{y_{nc} : n \in c\}|\phi)p(\phi|\{y_{nc} : n \notin c\})d\phi \tag{12}$$

This approach leads to numerical difficulties because $p(\{y_{nc} : n \in c\}|\phi)$ is numerically zero for some cells with large $\#\{c\}$. As a pragmatic resort, we employ an Empirical Bayes approach and insert the posterior mode $\hat{\phi}$ of $p(\phi|y)$ (still better would

be $p(\phi|\{y_{nc} : n \notin c\})$, but differences are negligible) into the prior predictive density, $p(\{y_{nc} : n \in c\}|\hat{\phi})$, to approximate the quantity of interest.

Regarding the third case in Equation (9), if a cell covers no entries from the training set, $\#\{c\} = 0$, a probability for $p(\{y_{nc} : n \in c\}|\phi)$ cannot be calculated. According to the formula it would equal one, but $p(\varnothing) = 0$. To still allow that this cell might generate future outcomes, we set $p(C_f = c|y) \propto 0.0000001$ to a small value.

5.4 Derived formulas

Standard calculus is used to derive the following equations. All model assumptions used below are summarized in Box 1. Our argument to derive the conditional posterior distribution and the marginal posterior distribution draws on (Gelman et al., 2014, p. 108ff.) who utilizes a similar strategy for binomial data in a setting where every individual belongs to a single group/cell.

The **joint posterior density** of all parameters is

$$\begin{aligned} p(\theta_1, ..., \theta_C, \phi|y) &\propto p(\phi) \cdot p(\theta_1, ..., \theta_C|\phi) \cdot p(y_1, ..., y_N|\theta_1, ..., \theta_C, \phi) \\ &\propto p(\phi) \cdot (\prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \prod_{k=1}^{K} \theta_{ck}^{\phi_k - 1}) \cdot (\prod_{c=1}^{C} \prod_{n=1;n\in c}^{N} \prod_{k=1}^{K} \theta_{ck}^{y_{nc}^{(k)}})) \end{aligned} \tag{13}$$

Using the independence assumptions from our model, we see that the **conditional posterior density** has the form

$$\begin{aligned} p(\theta_1, ..., \theta_C|\phi, y) &= \frac{p(\theta_1, ..., \theta_C|\phi) \cdot p(y_{11}, ..., y_{NC}|\theta_1, ..., \theta_C, \phi)}{p(y_{11}, ..., y_{NC}|\phi)} \\ &= \frac{p(\theta_1, ..., \theta_C|\phi) \cdot p(y_{11}, ..., y_{NC}|\theta_1, ..., \theta_C, \phi)}{\int p(\theta_1, ..., \theta_C|\phi) p(y_{11}, ..., y_{NC}|\theta_1, ..., \theta_C) d\theta_1, ..., \theta_C} \\ &= \frac{(\prod_{c=1}^{C} p(\theta_c|\phi))(\prod_{c=1}^{C} \prod_{n=1;n\in c}^{N} p(y_{nc}|\theta_c))}{\int (\prod_{c=1}^{C} p(\theta_c|\phi) \prod_{n=1;n\in c}^{N} p(y_{nc}|\theta_c)) d\theta_1, ..., \theta_C} \\ &= \frac{(\prod_{c=1}^{C} p(\theta_c|\phi) \prod_{n=1;n\in c}^{N} p(y_{nc}|\theta_c))}{\prod_{c=1}^{C} \int p(\theta_c|\phi) \prod_{n=1;n\in c}^{N} p(y_{nc}|\theta_c)) d\theta_c} \\ &= \prod_{c=1}^{C} \frac{p(\theta_c|\phi) p(\{y_{nc} : n \in c\}|\theta_c)}{p(\{y_{nc} : n \in c\}|\phi)} \\ &= \prod_{c=1}^{C} p(\theta_c|\phi, \{y_{nc} : n \in c\}) \end{aligned} \tag{14}$$

showing that the conditional posterior density $p(\theta_1, ..., \theta_C|\phi, y)$ is a product of independent densities $p(\theta_c|\phi, \{y_{nc} : n \in c\})$.

Inserting the model densities into this formula gives

$$\begin{aligned}
p(\theta_1, ..., \theta_C|\phi, y) &\propto \prod_{c=1}^{C} p(\theta_c|\phi) p(\{y_{nc} : n \in c\}|\theta_c, \phi) \\
&\propto \prod_{c=1}^{C} (\frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \prod_{k=1}^{K} \theta_{ck}^{\phi_k - 1}) \cdot (\prod_{n=1; n \in c}^{N} \prod_{k=1}^{K} \theta_{ck}^{y_{nc}^{(k)}}) \\
&\propto \prod_{c=1}^{C} \prod_{k=1}^{K} (\theta_{ck}^{\phi_k - 1} \cdot \theta_{ck}^{\sum_{n=1, n \in c}^{N} y_{nc}^{(k)}}) \\
&\propto \prod_{c=1}^{C} (\prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1}) \qquad (15)
\end{aligned}$$

This shows that $p(\theta_c|\phi, \{y_{nc} : n \in c\}) \propto \prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1}$, which is the kernel of a Dirichlet distribution. All terms in this product are thus independent Dirichlet distributions,

$$\theta_c|\phi, \{y_{nc} : n \in c\} \overset{iid}{\sim} \text{Dirichlet}(\phi_1 + \#\{y_c^{(1)}\}, ..., \phi_K + \#\{y_c^{(K)}\}) \qquad (16)$$

Its density has a closed, analytic form. It is

$$p(\theta_c|\phi, \{y_{nc} : n \in c\}) = \frac{\Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})}{\prod_{k=1}^{K} \Gamma(\phi_k + \#\{y_c^{(k)}\})} \prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1} \qquad (17)$$

Result from (15) and (17) are used to derive the **conditional prior predictive distribution**. Being calculated from model assumptions only, it is the joint prior probability given ϕ, $p(\{y_{nc} : n \in c\}|\phi)$, to observe all observed labels in y that belong to cell c,

$$\begin{aligned}
p(\{y_{nc} : n \in c\}|\phi) &= \frac{p(\theta_c|\phi) p(\{y_{nc} : n \in c\}|\theta_c, \phi)}{p(\theta_c|\{y_{nc} : n \in c\}, \phi)} \\
&= \frac{(\frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \prod_{k=1}^{K} \theta_{ck}^{\phi_k - 1}) \cdot (\prod_{n=1; n \in c}^{N} \prod_{k=1}^{K} \theta_{ck}^{y_{nc}^{(k)}})}{\frac{\Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})}{\prod_{k=1}^{K} \Gamma(\phi_k + \#\{y_c^{(k)}\})} \prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1}} \\
&= \frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \cdot \frac{\prod_{k=1}^{K} \Gamma(\phi_k + \#\{y_c^{(k)}\})}{\Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})} \qquad (18)
\end{aligned}$$

which is well-known to be the Dirichlet-multinomial density for categorical outcomes. For computational purposes we use its logarithm,

$$\begin{aligned}
\ln p(\{y_{nc} : n \in c\}|\phi) &= \ln \Gamma(\sum_{k=1}^{K} \phi_k) - \ln \Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\}) \\
&\quad + \sum_{k=1}^{K} (\ln \Gamma(\phi_k + \#\{y_c^{(k)}\}) - \ln \Gamma(\phi_k)) \qquad (19)
\end{aligned}$$

Our goal is to make predictions about future values y_f. The **conditional posterior predictive distribution** is needed for this. Inserting the conditional posterior density (17) in the nominator and inserting the same formula with increased $\#\{y_c^{(k)}\}$ also in the denominator, we obtain the conditional predictive probability $p(y_{fc}^{(k)} = 1|\phi, y)$ that a future value in cell c will belong to category k

$$
\begin{aligned}
p(y_{fc}^{(k)} = 1|\phi, y) &= \frac{p(\theta_c|\phi, y)p(y_{fc}^{(k)} = 1|\theta_c, \phi, y)}{p(\theta_c|y_{fc}^{(k)} = 1, \phi, y)} \\
&= \frac{(\frac{\Gamma(\sum_{k'=1}^K \phi_{k'} + \#\{y_c^{(k')}\})}{\prod_{k'=1}^K \Gamma(\phi_{k'} + \#\{y_c^{(k')}\})} \prod_{k'=1}^K \theta_{ck'}^{\phi_{k'} + \#\{y_c^{(k')}\} - 1}) \cdot (\theta_{ck})}{\frac{\theta_{ck}^{\phi_k + \#\{y_c^{(k)}\}}}{\Gamma(1+\phi_k + \#\{y_c^{(k)}\})} \frac{\Gamma(1+\sum_{k'=1}^K \phi_{k'} + \#\{y_c^{(k')}\})}{\prod_{k'=1, k' \neq k}^K \Gamma(\phi_{k'} + \#\{y_c^{(k')}\})} \prod_{k'=1, k' \neq k}^K \theta_{ck'}^{\phi_{k'} + \#\{y_c^{(k')}\} - 1}} \\
&= \frac{\Gamma(\sum_{k'=1}^K \phi_{k'} + \#\{y_c^{(k')}\})}{\Gamma(1 + \sum_{k'=1}^K \phi_{k'} + \#\{y_c^{(k')}\})} \frac{\Gamma(1 + \phi_k + \#\{y_c^{(k)}\})}{\Gamma(\phi_k + \#\{y_c^{(k)}\})} \\
&= \frac{\phi_k + \#\{y_c^{(k)}\}}{\sum_{k'=1}^K \phi_{k'} + \#\{y_c^{(k')}\}} \qquad (20)
\end{aligned}
$$

where we use for the last equality the property $\Gamma(1 + n) = n\Gamma(n)$ of the gamma function.

As one would expect, this is identical with the conditional posterior expectation, $p(y_{fc}^{(k)} = 1|\phi, y) = \mathbb{E}(\theta_c|\phi, \{y_{nc} : n \in c\})$. It is well known that the posterior expectation is a weighted average between observed data and prior information,

$$
\begin{aligned}
p(y_{fc}^{(k)} = 1|\phi, y) &= \omega \frac{\#\{y_c^{(k)}\}}{\#\{c\}} + (1 - \omega) \frac{\phi_k}{\sum_{k'=1}^K \phi_{k'}} \\
&= \omega \hat{P} + (1 - \omega)\mathbb{E}(\theta_c|\phi) \qquad (21)
\end{aligned}
$$

with weights $\omega = \frac{\#\{c\}}{\#\{c\} + \sum_{k'=1}^K \phi_{k'}}$ depending on the number of respondents whose verbatim answers belong to cell c. For large $\#\{c\}$, $p(y_{fc}^{(k)} = 1|\phi, y)$ is well approximated by the relative frequency $\hat{P} = \frac{\#\{y_c^{(k)}\}}{\#\{c\}}$, whereas small $\#\{c\}$ (=little experience with this cell) can seriously weight down the relative frequencies towards the prior expectation $\mathbb{E}(\theta_c|\phi)$. Usually, $\frac{\phi_k}{\sum_{k'=1}^K \phi_{k'}}$ will be very small due to the large number of categories K and might well be negligible.

The conditional posterior distributions (16) and (20) alone are of limited interest because they depend on ϕ, itself an unknown parameter. Its marginal distribution is needed for full Bayesian inference. Using the joint posterior density (13), some simplifications used in (15), and the conditional posterior density (17), we can determine

the **marginal posterior distribution**. It is

$$\begin{aligned} p(\phi|y) &= \frac{p(\theta_1, ..., \theta_C, \phi|y)}{p(\theta_1, ..., \theta_C|\phi, y)} \\ &\propto \frac{p(\phi) \cdot \prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1}}{\prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})}{\prod_{k=1}^{K} \Gamma(\phi_k + \#\{y_c^{(k)}\})} \prod_{k=1}^{K} \theta_{ck}^{\phi_k + \#\{y_c^{(k)}\} - 1}} \\ &\propto p(\phi) \prod_{c=1}^{C} \frac{\Gamma(\sum_{k=1}^{K} \phi_k)}{\prod_{k=1}^{K} \Gamma(\phi_k)} \frac{\prod_{k=1}^{K} \Gamma(\phi_k + \#\{y_c^{(k)}\})}{\Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})} \end{aligned} \tag{22}$$

For computational purposes it is helpful to do calculations on a logarithmic basis, and with some unknown constant c we have,

$$\begin{aligned} \ln p(\phi|y) = c + \ln(p(\phi)) + \sum_{c=1}^{C} (\sum_{k=1}^{K} (\ln \Gamma(\phi_k + \#\{y_c^{(k)}\}) - \ln \Gamma(\phi_k)) + \\ + \ln \Gamma(\phi_1 + ... + \phi_K) - \ln \Gamma(\sum_{k=1}^{K} \phi_k + \#\{y_c^{(k)}\})) \end{aligned} \tag{23}$$

We determine the **observed Fisher information** next. Two cases will be distinguished, depending on how many components ϕ has.

In the following, $\psi^{(i)}(x) = \frac{\partial^{i+1} \ln \Gamma(x)}{\partial x^{i+1}}$ denotes the polygamma function of order i (digamma ($i = 0$) and trigamma ($i = 1$) functions).

If $\phi_1 = ... = \phi_K =: \phi_{(1)}$ (ϕ depends on a single value), the derivative of $\ln p(\phi|y)$ is

$$\begin{aligned} \frac{\partial}{\partial \phi_{(1)}} \ln p(\phi|y) = \sum_{c=1}^{C} (\sum_{k=1}^{K} (\psi^{(0)}(\phi_{(1)} + \#\{y_c^{(k)}\}) - \psi^{(0)}\Gamma(\phi_{(1)})) + \\ + K \cdot \psi^{(0)}(K \cdot \phi_{(1)}) - K \cdot \psi^{(0)}(\sum_{k=1}^{K} \phi_{(1)} + \#\{y_c^{(k)}\})) \end{aligned} \tag{24}$$

and the second order derivative is

$$\begin{aligned} \frac{\partial}{\partial \phi_{(1)} \phi_{(1)}} \ln p(\phi|y) = \sum_{c=1}^{C} (\sum_{k=1}^{K} (\psi^{(1)}(\phi_{(1)} + \#\{y_c^{(k)}\}) - \psi^{(1)}\Gamma(\phi_{(1)})) + \\ + K^2 \cdot (\psi^{(1)}(K \cdot \phi_{(1)}) - \psi^{(1)}(\sum_{k=1}^{K} \phi_{(1)} + \#\{y_c^{(k)}\}))) \end{aligned} \tag{25}$$

yielding the negative observed Fisher information.

Now consider the case that ϕ depends on a two values (see Equation 7),

$$\phi_k := \begin{cases} \phi_R & \text{if cell } c \text{ refers to category } k, \\ \phi_{\bar{R}} & \text{if cell } c \text{ does not refer to category } k. \end{cases} \tag{26}$$

Let D_{ck} be an indicator that equals 1 if cell c refers to category k, and 0 otherwise. The partial derivatives are

$$\frac{\partial}{\partial \phi_R} \ln p(\phi|y) = \sum_{c=1}^{C}(\sum_{k=1}^{K} D_{ck} \cdot (\psi^{(0)}(\phi_R + \#\{y_c^{(k)}\}) - \psi^{(0)}(\phi_R)) + \\ + \psi^{(0)}(\phi_R + (K-1)\phi_{\bar{R}}) - \psi^{(0)}(\phi_R + (K-1)\phi_{\bar{R}} + \sum_{k=1}^{K} \#\{y_c^{(k)}\})) \tag{27}$$

and

$$\frac{\partial}{\partial \phi_{\bar{R}}} \ln p(\phi|y) = \sum_{c=1}^{C}(\sum_{k=1}^{K}(1 - D_{ck}) \cdot (\psi^{(0)}(\phi_{\bar{R}} + \#\{y_c^{(k)}\}) - \psi^{(0)}(\phi_{\bar{R}})) + \\ + (K-1) \cdot (\psi^{(0)}(\phi_R + (K-1)\phi_{\bar{R}}) - \psi^{(0)}(\phi_R + (K-1)\phi_{\bar{R}} + \sum_{k=1}^{K} \#\{y_c^{(k)}\}))) \tag{28}$$

This leads to the second order derivatives

$$\frac{\partial^2}{\partial \phi_R \phi_R} \ln p(\phi|y) = \sum_{c=1}^{C}(\sum_{k=1}^{K} D_{ck} \cdot (\psi^{(1)}(\phi_R + \#\{y_c^{(k)}\}) - \psi^{(1)}(\phi_R)) + \\ + \psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}}) - \psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}} + \sum_{k=1}^{K} \#\{y_c^{(k)}\})) \tag{29}$$

$$\frac{\partial^2}{\partial \phi_{\bar{R}} \phi_{\bar{R}}} \ln p(\phi|y) = \sum_{c=1}^{C}(\sum_{k=1}^{K}(1 - D_{ck}) \cdot (\psi^{(1)}(\phi_{\bar{R}} + \#\{y_c^{(k)}\}) - \psi^{(1)}(\phi_{\bar{R}})) + \\ + (K-1)^2 \cdot (\psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}}) - \psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}} + \sum_{k=1}^{K} \#\{y_c^{(k)}\}))) \tag{30}$$

$$\frac{\partial^2}{\partial \phi_R \phi_{\bar{R}}} \ln p(\phi|y) = \sum_{c=1}^{C}(K-1) \cdot (\psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}}) - \psi^{(1)}(\phi_R + (K-1)\phi_{\bar{R}} + \sum_{k=1}^{K} \#\{y_c^{(k)}\})) \tag{31}$$

The observed Fisher information is

$$I(\phi_R, \phi_{\bar{R}}) = -\begin{pmatrix} \frac{\partial^2}{\partial \phi_R \phi_R} \ln p(\phi|y) & \frac{\partial^2}{\partial \phi_R \phi_{\bar{R}}} \ln p(\phi|y) \\ \frac{\partial^2}{\partial \phi_R \phi_{\bar{R}}} \ln p(\phi|y) & \frac{\partial^2}{\partial \phi_{\bar{R}} \phi_{\bar{R}}} \ln p(\phi|y) \end{pmatrix} \tag{32}$$

6 References

References

Albrecht, S., Schmich, P. and Varga, M. (2017). Occupation coding in the german health update (geda-study 2014/15). 7th Conference of the European Survey Research Association.
URL: *https://www.europeansurveyresearch.org/conference/programme2017?sess=4\#630*

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M. and Trahms, A. (2010). Arbeiten und Lernen im Wandel * Teil 1: Überblick über die Studie, *FDZ-Methodenreport 05/2010*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Berg, M., Cramer, R., Dickmann, C., Gilberg, R., Jesske, B., Kleudgen, M., Beste, J., Dummert, S., Frodermann, C., Fuchs, B., Schwarz, S., Trappmann, M. and Trenkle, S. (2017). Codebuch und Dokumentation des Panel 'Arbeitsmarkt und soziale Sicherung' (PASS) * Band I: Datenreport Welle 10, *FDZ-Datenreport 07/2017*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Bröcker, J. and Smith, L. A. (2007). Increasing the reliability of reliability diagrams, *Weather and Forecasting* **22**(3): 651–661.
URL: *https://doi.org/10.1175/WAF993.1*

Browne, W. J., Goldstein, H. and Rasbash, J. (2001). Multiple membership multiple classification (mmmc) models, *Statistical Modeling* **1**(2): 103–124.

Chen, B.-C., Creecy, R. H. and Appel, M. V. (1993). Error control of automated industry and occupation coding, *Journal of Official Statistics* **9**(4): 729–745.

Creecy, R. H., Masand, B. M., Smith, S. J. and Waltz, D. L. (1992). Trading mips and memory for knowledge engineering, *Commun. ACM* **35**(8): 48–64.
URL: *http://doi.acm.org/10.1145/135226.135228*

Drasch, K., Matthes, B., Munz, M., Paulus, W. and Valentin, M.-A. (2012). Arbeiten und Lernen im Wandel * Teil V: Die Codierung der offenen Angaben zur beruflichen Tätigkeit, Ausbildung und Branche, *FDZ-Methodenreport 04/2012*, Forschungsdatenzentrum der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2014). *Bayesian Data Analysis, Third Edition*, Chapman and Hall/CRC Press, Boca Raton.

Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness, *Journal of the Royal Statistical Society: Series B (Statistical*

Methodology) **69**(2): 243–268.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2007.00587.x*

Goldstein, H. (2011). *Multilevel Statistical Models, Fourth Edition*, Wiley, Chichester.

Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M. and Steiner, S. (2017). Three Methods for Occupation Coding Based on Statistical Learning, *Journal of Official Statistics* **33**(1): 101–122.

Hall, A., Siefer, A. and Tiemann, M. (2015). BIBB/BAuA Employment Survey of the Working Population on Qualification and Working Conditions in Germany 2012. vt_1.0, sv_2.0, *Research Data Center at BIBB (ed., data access)*, Federal Institute for Vocational Education and Training, Bonn.
URL: *https://doi.org/10.7803/501.12.1.4.10*

Hartmann, J., Tschersich, N. and Schütz, G. (2012). Die Vercodung der offenen Angaben zur beruflichen Tätigkeit nach der Klassifikation der Berufe 2010 (KldB 2010) und nach der International Standard Classification of Occupations 2008 (ISCO08), *Technical report*, TNS Infratest Sozialforschung, Munich.
URL: *https://metadaten.bibb.de/download/684*

Hill, P. W. and Goldstein, H. (1998). Multilevel modeling of educational data with cross-classification and missing identification for units, *Journal of Educational and Behavioral Statistics* **23**(2): 117–128.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial, *Statistical Science* **14**(4): 382–417.

Hoffmann, R., Lange, M., Butschalowsky, H., Houben, R., Schmich, P., Allen, J., Kuhnert, R., Schaffrath Rosario, A. and Gößwald, A. (2018). KiGGS Wave 2 cross-sectional study – participant acquisition, response rates and representativeness, *Journal of Health Monitoring* **3**(1): 78–91.

Hsu, W.-r. and Murphy, A. H. (1986). The attributes diagram: A geometrical framework for assessing the quality of probability forecasts, *International Journal of Forecasting* **2**(3): 285 – 293.

Jollife, I. T. and Stephenson, D. B. (2012). *Forecast Verification: A Practitioner's Guide to Atmospheric Science, 2nd Edtion*, Wiley, Chichester.
URL: *http://dx.doi.org/10.1002/9781119960003*

Lange, C., Finger, J., Allen, J., Born, S., Hoebel, J., Kuhnert, R., Müters, S., Thelen, J., Schmich, P., Varga, M., von der Lippe, E., Wetzstein, M. and Ziese, T. (2017). Implementation of the European health interview survey (EHIS) into the German health update (GEDA), *Archives of Public Health* **75**(1): 40.
URL: *https://doi.org/10.1186/s13690-017-0208-6*

Lange, M., Hoffmann, R., Mauz, E., Houben, R., Gößwald, A., Schaffrath Rosario, A. and Kurth, B.-M. (2018). KiGGS Wave 2 longitudinal component – data collection design and developments in the numbers of participants in the KiGGS cohort, *Journal of Health Monitoring* **3**(1): 92–107.

Mauz, E., Gößwald, A., Kamtsiuris, P., Hoffmann, R., Lange, M., Schenck, U. v., Allen, J., Butschalowsky, H., Frank, L., Hölling, H., Houben, R., Krause, L., Kuhnert, R., Lange, C., Stephan, M., Neuhauser, H., Christina, P.-M., Richter, A., Schaffrath Rosario, A., Schaarschmidt, J., Schlack, R., Schlaud, M., Schmich, P., Gina, S., Wetzstein, M., Ziese, T. and Kurth, B.-M. (2017). New data for action. Data collection for KiGGS Wave 2 has been completed, *Journal of Health Monitoring* **2**(S3): 2–27.

Murphy, A. H. and Epstein, E. S. (1967). Verification of probabilistic predictions: A brief review, *Journal of Applied Meteorology* **6**(5): 748–755.
URL: *https://doi.org/10.1175/1520-0450(1967)006¡0748:VOPPAB¿2.0.CO;2*

O'Hagan, A. and Forster, J. (2004). *Kendall' Advanced Theory of Statistics, Volume 2B: Bayesian Inference, 2nd Edition*, Arnold, London.

Potts, J. M. (2012). Basic concepts, *in* I. T. Jollife and D. B. Stephenson (eds), *Forecast Verification: A Practitioner's Guide to Atmospheric Science, 2nd Edtion*, Wiley, Chichester, pp. 11–29.
URL: *http://dx.doi.org/10.1002/9781119960003.ch2*

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer, New Xork.

Rohrbach-Schmidt, D. and Hall, A. (2013). BIBB/BAuA Employment Survey 2012, *BIBB-FDZ Data and Methodological Reports Nr. 1/2013. Version 4.1*, Federal Institute for Vocational Education and Training, Bonn.

Roulston, M. S. and Smith, L. A. (2002). Evaluating probabilistic forecasts using information theory, *Monthly Weather Review* **130**(6): 1653–1660.

Sakshaug, J. W., Schmucker, A., Kreuter, F., Couper, M. P. and Singer, E. (2016). Evaluating active (opt-in) and passive (opt-out) consent bias in the transfer of federal contact data to a third-party survey agency, *Journal of Survey Statistics and Methodology* **4**(3): 382–416.
URL: *http://jssam.oxfordjournals.org/content/4/3/382.abstract*

Schierholz, M., Gensicke, M., Tschersich, N. and Kreuter, F. (2018). Occupation coding during the interview, *Journal of the Royal Statistical Society: Series A* **181**(2): 379–407.

Statistisches Bundesamt (2016). *Demographische Standards*, Statistisches Bundesamt, Wiesbaden.

Trappmann, M., Beste, J., Bethmann, A. and Müller, G. (2013). The pass panel survey after six waves, *Journal for Labour Market Research* **46**(4): 275–281.
URL: *https://doi.org/10.1007/s12651-013-0150-1*

van der Loo, M. (2014). The stringdist package for approximate string matching, *The R Journal* **6**: 111–122.
URL: *https://CRAN.R-project.org/package=stringdist*

Eidesstattliche Versicherung

Eidesstattliche Versicherung gemäß § 9 Absatz 1 Buchstabe e) der Promotionsordnung der Universität Mannheim zur Erlangung des Doktorgrades der Sozialwissenschaften:

1. Bei der eingereichten Dissertation mit dem Titel *New Methods for Job and Occupation Classification* handelt es sich um mein eigenständig erstelltes eigenes Werk.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtliche Zitate aus anderen Werken als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bisher nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärung bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.