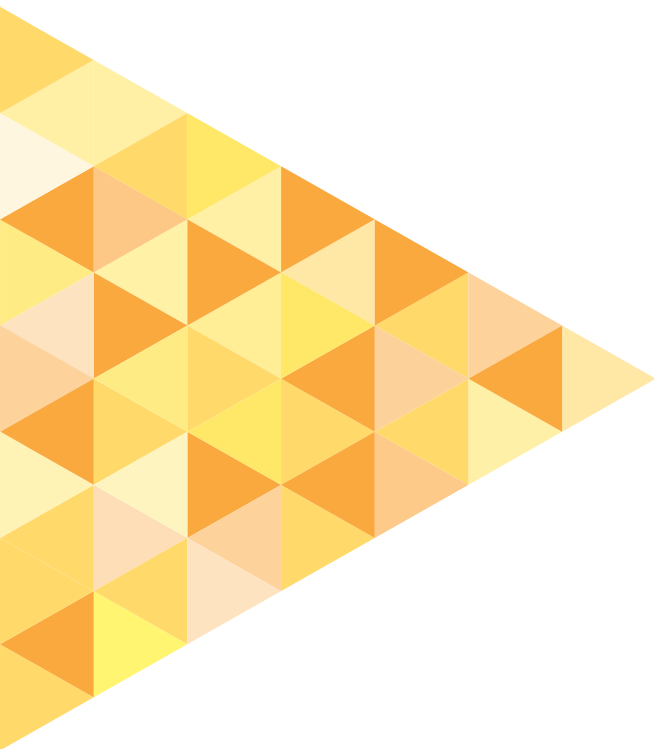


Timo Schnepf | Kai Krüger | Johanna Binnewitt | Stefan Winnige | Anja Firll

Emphasizing the laborious side of machine learning: annotation guidelines and how to develop them



VET | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
REPOSITORY
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |

Citation:
Schnepf, Timo; Krüger, Kai; Binnewitt, Johanna; Winnige, Stefan; Firll, Anja: Emphasizing the laborious side of machine learning: annotation guidelines and how to develop them. Version 1.0 Bonn, 2024. Online: https://res.bibb.de/vet-repository_782577.



© Federal Institute for Vocational Education and Training (BIBB), 2024

Version 1.0
July 2024

Publisher

Federal Institute for Vocational Education and Training
Friedrich-Ebert-Allee 114-116
D-53113 Bonn
Internet: www.vet-repository.info
E-Mail: repository@bibb.de

Despite careful checking of the content, the Federal Institute for Vocational Education and Training accepts no liability for the content of external links. The operators of the linked pages are solely responsible for their content.

CC Lizenz

The content of this work is subject to a Creative Commons Licence (licence type: Attribution-ShareAlike 4.0 International). For further information, please visit our Creative Commons information site online at <https://creativecommons.org/licenses/by-sa/4.0/>.

Bibliographic information from the German National Library

The German National Library catalogues this publication in the German National Bibliography:
urn:nbn:de:0035-vetrepository-782577-1

Emphasizing the laborious side of machine learning: annotation guidelines and how to develop them

Timo Schnepf, Kai Krüger, Johanna Binnewitt, Stefan Winnige, Anja Firl

Abstract (english)

This paper presents an overview on the development and application of Annotation Guidelines (AGs) in the context of the BIBB job advertisements project. It highlights the importance of AGs in ensuring data annotation consistency, fostering reproducibility, and research transparency. It describes the process of creating AGs for the extraction of information from online job advertisements, emphasizing the challenges faced and the iterative approaches adopted to refine these guidelines. The study also examines the transition from classic to 'relational extraction' methods to address complex, compositional concepts in unstructured texts. Highlighting the importance of AGs in enhancing reproducibility, transparency, and the interoperability of computational social science, the paper advocates for the open access of AGs as critical resources for advancing the field.

Abstract (german)

Dieser Beitrag gibt einen Überblick über die Entwicklung und Anwendung von Annotationsrichtlinien oder „Annotationsguidelines“ (AGs) im Rahmen des Stellenanzeigen Projekts des BIBB. Er unterstreicht die Wichtigkeit von AGs in Bezug auf die Qualität von Datenannotationen, die Reproduzierbarkeit sowie die Transparenz im Bereich Computational Social Science. Der Beitrag beschreibt den Prozess der Erstellung von AGs für die Extraktion von Informationen aus Online-Stellenanzeigen. Er hebt die Herausforderungen hervor und beschreibt Lösungsansätze, die zur Verbesserung von AGs genutzt werden können. An praktischen Beispielen untersucht der Beitrag die Frage, wie sogenannte "relationale Extraktionsmethoden" helfen können, komplexe, kompositorische Konzepte in unstrukturierten Texten zu erfassen. Das Papier hebt die Bedeutung von AGs für die Verbesserung der Reproduzierbarkeit, Transparenz und Interoperabilität der Computational Social Science hervor und plädiert für einen offenen Zugang zu AGs als wichtige Ressourcen für die Weiterentwicklung des Fachgebiets.

Table of content

Why is the development of a detailed annotation guideline important?	3
Tools and workflows.....	6
Initiation phase: Building the AG draft.....	8
The first iteration.....	10
Further iterations	11
Insights from annotating online job ads.....	12
Outlook: Classic vs Relational extraction	14
Conclusion	15
Appendix.....	16
Literature	16
Annotation Guideline	17

Why is the development of a detailed annotation guideline important?

The amount of information stored on personal computers and server farms continues to grow. Much of it is stored in, or can be converted to text (e.g. by speech-to-text algorithms). Natural language processing (NLP) makes it possible to extract specific information from such unstructured text. It makes information available for research. A key factor in this process is to link theoretical phenomena of interest to observable text passages. Poor definitions can jeopardise both reproducibility and interoperability. Transparency therefore plays an important role in these research efforts. The following article highlights the importance of annotation guidelines (AGs) in this context. It emphasises the relevance of guidelines and gives an example of the workflow and tools to iterate through this process. We use the development of the annotation guidelines for the BIBB job advertisements project as an example. We describe the challenges using the example of online job advertisements.

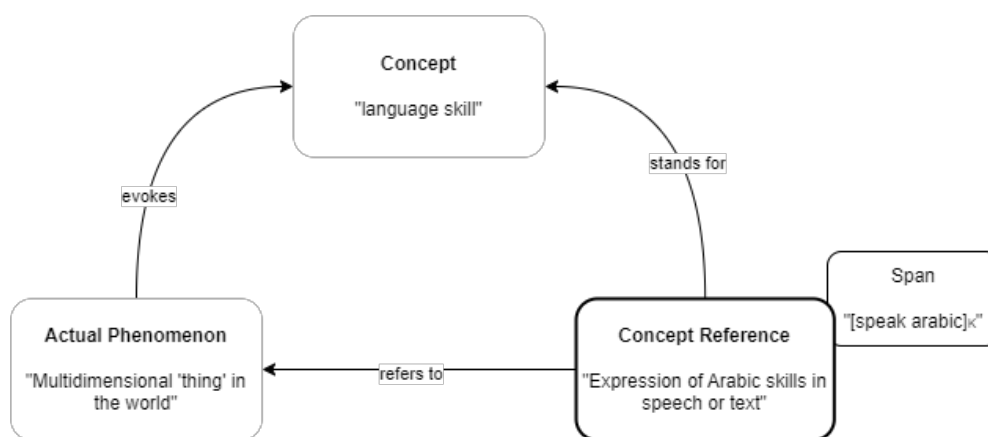


Figure 1: Relation between the phenomenon, the concept and the concept reference, own depiction inspired by Cherry 1966.

Annotation guidelines are the foundation of any NLP task that seeks to systematically extract countable information from text. They serve several purposes:

1. AGs define *which* concepts we want to extract

Annotated text serves as a "gold standard" for machine learning models: which words/sentences represent the concept of interest? In most cases, concepts are

fuzzy, latent variables that we cannot observe directly. Examples include concepts such as 'sentiment', 'fake news', 'skills', 'environmental sustainability'. We use approximations of these phenomena, such as texts or other expressions that we can observe. AGs help us define these concepts in two ways: deductively and inductively (cf. Scharnow 2012, Grimmer et al. 2022). Inductively means that we approach our definitions from theoretical considerations, from the requirements of the research question "What defines our concept from the perspective of an ideal research goal?". Each research question has different requirements. A broad literature review and interviews with experts/stakeholders help to understand concepts from different perspectives. Working groups help to discuss what can and cannot be extracted from a particular source. They therefore help to manage expectations. For example, 'skills' from a qualification and examination perspective are defined as measurable 'competence to act' and may not be the same as 'competencies' as we find them in job advertisements, which are primarily defined as - more abstract - skill requirements. However, unlike bespoke data, online job adverts are ready-made. That is, a researcher 'sees something that already exists in the world and then creatively reuses it' (Salganik 2018: 7). Thus, it is not enough to approach concept definitions only inductively, but also deductively: we take 'something that already exists' and define from this observation what the concept is. For example, by giving and explaining examples. AGs also document where we categorise ambiguous terms (is 'management' a skill or a task?).

2. AGs define *how* we want to extract concepts

Inspired by the "semiotic triangle" (Cherry 1966), Figure 1 shows the relationship between concepts, concept references (CR) and the actual phenomenological "thing" in the world. People use language to describe phenomena as "concepts". These concepts are mental models that emphasise some aspects and ignore others (cf. Entman 1993). A good definition is therefore crucial. However, we can only observe "concept references", expressions of an actual phenomenon. AGs help to define the boundaries of a concept reference, the actual tagged entity, the span. The span is defined by two considerations:

- i. What do we want to study in terms of content? What additional information is part of the concept? Which are not?
- ii. How do we annotate the text technically? What exactly to annotate is not trivial, and the enormous variety of language requires documented annotation rules. When we find "you are fluent in Arabic and Spanish", is this two spans or one? Does "fluent" refer only to "Arabic" or also to "Spanish"? Do we separate them in the annotation or in a follow-up task? How do we deal with even more complex sentences?

The AGs are designed to clarify these questions.

3. AGs serve as documentation and reference for human annotators

As mentioned above, the variety of ways in which (fuzzy) concepts are expressed in language is a challenge for information extraction tasks. Even with sufficient training and practice, it is not possible to memorise every rule and definition. AGs serve as a reliable reference for documentation.

4. AGs serve refinement of AGs

Because natural language is so diverse, AGs will never be "complete". When annotating, AGs help to identify missing parts or contradictions. Their clear structure helps to discuss paragraphs that need refining or restructuring.

5. AGs facilitate evaluation, transparency, replicability and interoperability of research

AGs facilitate research evaluation, transparency, replicability and interoperability. Only if humans know which concepts they want to extract and how, they can program algorithms to automate this task. The results of a data analysis task depend fundamentally on the definitions and extraction rules defined in the AGs. Annotated text based on a detailed AG helps to evaluate algorithms, i.e. by defining silver or gold standards (Alexopolous 2020). They help researchers pick up the baton by making clear which aspects of a phenomenon are highlighted and how they are observed. This allows not only comparability between studies, but also replicability. Given this important aspect of scientific research, we emphasise the importance of transparency in what is, to be honest, a relatively laborious task.

Open access to AGs is not currently the norm. In survey and questionnaire-based research, it is common practice to provide "method reports" and "codebooks" with the data. We see AGs in a similar relationship to data extracted from unstructured text. We relate this situation to the relatively young and interdisciplinary nature of the field. We also observe a lack of a common language among researchers in the field. Each discipline and each author has a different way of naming things. Concepts are called "categories", "abstract entities", "annotation schemes" (cf. Reiter et al. 2020, Alexopoulos 2020, Finlayson & Erjavec 2017). Conceptual references are called "entity references", "specifications" or sometimes just "terms" (cf. Ketschik et al. 2020, Alexopoulos 2020, Finlayson & Erjavec 2017).

We hope that the five arguments above make it clear that AGs need to be well constructed and then made available to the scientific community. How we approach the development of AGs is described in the next section.

Tools and workflows

Before doing so, we will briefly introduce our experience on the rather technical decisions for specific tools and workflows. The first step is to decide which tools are best suited to the annotation task. How fuzzy and convoluted can concept references be? There are now many (online) annotation tools with different functionalities that may be better or worse suited to the annotation task at hand. Finlayson and Erjavec give an overview of different aspects that should be considered when choosing an annotation tool (Finlayson & Erjavec 2017).

Furthermore, Neves and Ševa present an overview of 15 different tools for annotating text in the biomedical domain (Neves & Ševa 2021).

The first thing to consider is the textual level at which the data should be annotated. In our case, we have modelled the annotation as a span annotation task, which means that several

passages in a job advertisement should be annotated, rather than the whole job advertisement or individual sentences.

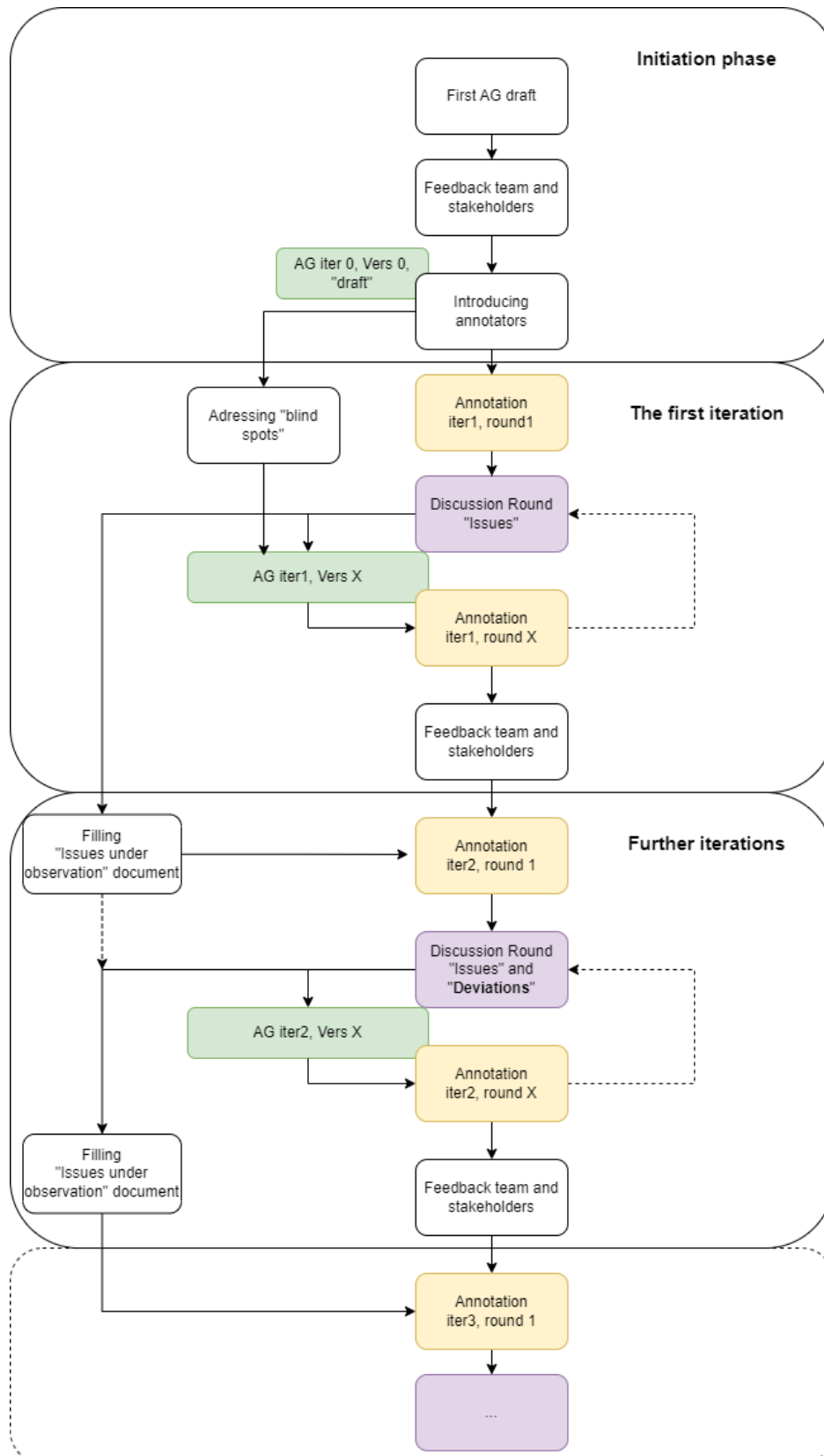


Figure 2: Iteration Process Chart

When suitable annotation software is found, the AG developer, her colleagues and stakeholders develop the AG. How this is done is shown in Figure 2 and the following subsections.

Initiation phase: Building the AG draft

In the initiation phase, the AG developer reviews the literature dealing with the phenomena of interest. Along with the theoretical work, she takes a first look at the unstructured textual data. She writes down concept references and develops a first idea of the concepts to be extracted. They do not have to be worked out, but serve as a common starting point. We structured our AG (see appendix) with paragraphs in the following way:

1. Introduction

Describes the purpose of the research project and the AG.

2. § Paragraph for general annotation rules

Describes general extraction rules, independent of concepts. This can be quite technical, such as how to deal with convoluted phrases or conflicting concept references, and how annotators should deal with new issues not covered by the current AG version.

3. § Paragraph for each concept

It starts with a theoretical introduction to the concept. It then provides a list of previous extractions before documenting annotation rules specific to the concept.

The paragraphs have the following structure:

§ Paragrah G 1 (§ General 1)

§ Subparagraph (§ G 1.1)

Explanation of the Subparagraph

Example 1: “here is the text passage containing the [span]_c marked with the concept”

When not obvious, a sentence concerning the example can be provided.

Counterexample 1: “here is a counterexample without a concept reference”

Sometimes counterexamples can be helpful to distinguish different cases.

§ Paragrah G 2

§ Paragrah G k...

...

§ Paragraph A 1 (§ Concept A 1)

...

The structure allows each annotation rule to be explained in detail. The numbering of the paragraphs allows for easy reference during the discussion (e.g. “§G 4.2 contradicts §K 2.2”). Once the AG developer has received confirmation from the team and stakeholders, she starts to annotate the text and writes down uncertainties. Initially, depending on the complexity, there will be many uncertainties and it is *difficult to assess the importance of each issue*. If there is no obvious solution, the issue is documented in a “discussion document”. The document allows issues to be structured and decisions to be tracked along the AG development process. We used an Excel spreadsheet with the following columns:

1. the text ID,
2. the issue, i. e. the problematic text section (including the span),
3. a short explanation of the problem,
(in the first iteration, this will just be collected, in further iterations problem categories can be used),
4. for each annotator an own column for remarks and
5. a “lesson learned” column
(this is at the beginning empty and documents the discussion results in the following meetings).

We organised each concept in a separate spreadsheet. At the end of the initiation phase, two documents are produced: the AG iter 0, version 0, or "draft", and a "discussion document" containing issues and open questions.

The first iteration

The next step is to introduce other annotators to the annotation policy. In our experience, the easiest way to introduce others is to observe the AG developer annotating. She can give examples of why she chooses a particular concept reference and span length. When colleagues feel confident enough, they start annotating themselves and the AG developer accompanies their first annotations. Many questions will arise. They can be raised directly with the AG developer or collected in the "discussion document". These initial questions are helpful hints for the AG developer. They allow her to specify the AG with regard to her "blind spots". Their inclusion will reduce the introduction and training time for further annotators.

Annotators and the AG developer annotate the first set of text documents. They collect their problems in the "discussion document". Before the discussion meeting, the AG developer cleans up the document, merges similar problems and writes obvious solutions in the "lesson learned" field. Then all the annotators look at the document and make comments. This helps to identify the most important issues, increases the structure of the meeting and reduces meeting time. At the end of the meeting, each issue is either resolved ("lesson learned") or cannot be resolved given the current state of knowledge. We have had good experience of moving the larger issues or those of more general interest to a third document, the "issues under observation" document. These issues are carried forward to the next iteration round, as the level of uncertainty in the annotation decreases with each iteration. We do not recommend initialising inter-annotator agreement measurements in the first iteration.

After the discussion meeting, the AG developer incorporates the "lessons learned" into the AG. At the end of iteration 1, we get the AG iteration 1, version 1.

Further iterations

After the team and stakeholders have reviewed the AG iteration 1, version 1, the fine tuning of the AG begins. As in Iteration 1, further iterations include discussion sessions to discuss new issues and 'issues under observation'. In order to fine-tune the common understanding of spans and to identify misunderstandings, iterations may now include discussion of metric measures of inter-annotator agreement (i.e. Krippendorff's alpha) and deviations in spans.

Deviations can occur when

- 1) annotator A marks a span which annotator B simply does not,
- 2) annotator A and B mark the same span but refer to different concepts,
- 3) annotator A and B refer to the same concept but vary in the span length.

With sufficient agreement between annotators, annotated extracts (spans) can be used as examples of concept references in the AG. This helps new annotators to understand the concepts better and to check if and how a concept reference/span is part of a concept. Further annotation rounds end when the project team decides that a certain level of annotation agreement has been reached and the AG is detailed enough for new annotators to work with it independently. Depending on the complexity of the textual data and the amount of expected content extraction, AGs will never cover every use case and every variety of natural language. This is not necessary. We argue that making AGs available in the first place is already a step towards a transparent research process in CSS.

Insights from annotating online job ads

In the following section we present five practical issues related to the extraction of skills from online job advertisements.

Extensive experience in research is mandatory , interest in teaching would be a plus .



Figure 3: Additional information

Figure 3: Additional information shows three possible span lengths below the text containing two concept references for the skills 'research' and 'teaching'. By structuring the topics with the 'discussion document', we found that concept references are often accompanied by additional information, which in turn can be categorised for each concept. For skills, we found the categories of 'level of experience' and 'urgency' to be common accompaniments. Whether and how to include such information needs to be addressed in the working group: What kind of information is needed? How to annotate it?

Bachelors in a relevant engineering discipline or minimum five years of working experience .



Figure 4: "Or"-connector

Figure 4 shows a problem with "or" statements in the text. If the "or" is ignored, we find two requirements for access to a job, a bachelor's degree and work experience. Option 2 does not allow us to conclude that at least a bachelor's degree is required to enter the profession. This only tells half the story. Important information would also be ignored with an "and" link. Option 2 would also lead to incomplete statements. If, as we have done, we use option 1, we have to include further disentangling processes for all "or"-cases. Whether or not this information is useful to the analyst depends on his research question. The big advantage, in our view, is that AGs allow transparency in this process.

At this job , you 'll **carry out surgical operations** .

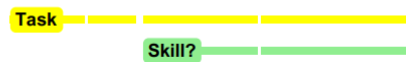


Figure 5: Task vs Skill

Figure 5 shows the inherent link between two concepts, here skills and tasks. Depending on the research question, a distinction may be useful (see Autor 2013). However, when companies formulate their job description, they may implicitly communicate that the candidate must have the skills to 'perform surgical operations'. The same applies to the reverse formulation. If a company requires "programming skills", it is almost certain that the job will include such tasks. When extracting information and categorising it in terms of concepts, it is important to be transparent about the unit on which the research is based.

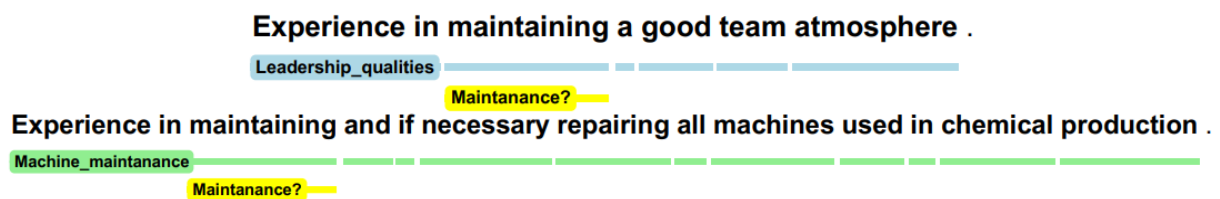


Figure 6: Skills objects

Figure 6 illustrates a problem concerning the ambiguity of verbs in relation to tasks. We show that the object relation of the skill 'maintaining' is crucial for its information value. Simply annotating 'maintaining' would not allow the inherent meaning to be reconstructed. Maintaining a good atmosphere is different from maintaining machines. The length of the span increases further with the subclause "and, if necessary, repair". Therefore, the next section shows an alternative annotation scheme to select the concept-object relation in a simpler way.

Very organized and independent



Figure 7: Syntactic ambiguity

Finally, Figure 7 shows an example of syntactic ambiguity, i.e. where the information provided can be interpreted in several ways. The provided text does not allow to assess whether a candidate has to be only very organised and independent or also "very independent", i.e. whether the adverb binds only to one or to both adjectives. Analysing option 2 implicitly removes additional information about the level of experience of working independently. Our solution was to choose option two for the AG. As there is no "right" answer, the transparency documented in the AGs is important.

Outlook: Classic vs Relational extraction

As shown in the last section, one factor that can make the development of annotation guidelines for sequence labelling challenging is the trade-off between having or wanting to retain conceptual information that appears compositionally at different positions in the text, and the possibility of drastically increasing span lengths when these positions are far apart. It also requires an additional processing step to decompose the meaning of these spans once they have been extracted. Zhang et. Al. (2022) also show for OJA skills extraction that longer spans lead to poorer model performance. A well-known approach that might help here is relation extraction (RE) (see e.g. Qin et. Al. 2022). Here, labelled sequences (extracted entities) can be related to each other by assigning them meaningful labels, which we call relations. Relations are usually at the semantic or syntactic level. While RE has been applied to a variety of NLP tasks, we believe that it has unexplored potential in many cases where sequence labelling tasks target fuzzy concepts such as tasks. As shown in the previous section, tasks and skills can be compositional in nature, depending on how they are modelled. Simply put, a task is something someone does with a particular thing or person. Someone is maintaining a machine. So in this case the task consists of two components: the process of repairing and the object machines that are being maintained. While both components are semantically inseparable in terms of the dimension of analysis, they are independent clauses that do not necessarily have to be syntactically close to each other. The idea, then, is to consider the concept we are interested in analysing (in this example: tasks) as a high-level composition that is not directly labelled in the text. Rather, the individual

components are the sequences for the labelling task, and using RE allows them to be linked together to form the relevant concept. For example, maintenance can be labelled as a process and machines as work content. Then there is also a relation (object being transformed) between the two entities, which allows the task of maintaining machines to be composed independently of the number of additional words in the sentence.

Conclusion

This paper underscores the importance of annotation guidelines in the realm of Natural Language Processing for the social sciences. It emphasizes that AGs not only serve as a foundation for ensuring reproducibility, transparency, and interoperability of research but also play a pivotal role in defining and extracting theoretical – and often not directly observable - concepts accurately. Through the lens of the BIBB job advertisements project, we have navigated the challenges and insights gained from annotating online job ads, highlighting the iterative process of AG development. Additionally, we touched upon the potential of Relational Extraction over ‘classic extraction’ to handle complex, compositional concepts more effectively. This exploration reinforces the argument for open access to AGs, advocating for their recognition as indispensable tools for advancing NLP research in the field of computational social science.

Acknowledgement: This research is part of the project Occupations and Skills Radar funded by the BMBF. We thank all team members that are included in the annotation process.

Appendix

Literature

- &effect (2024), Guide to german online Job ad data, Last access: 10.04.2024, www.oja-guide.de
<https://www.oja-guide.de/steps/taxonomies-and-ontologies>
- Alexopoulos, Panos: Semantic Modeling for Data (2020)
- Cherry, Colin: On human communication a review, a survey, and a criticism (1966)
- Entmann, Robert M.: Framing: Toward Clarification of a Fractured Paradigm (1993)
- Finlayson, Mark Alan; Erjavec Tomaž: Overview of Annotation Creation: Processes and Tools (2017)
- Grimmer, Margaret E. Roberts, Brandon M. Stewart. Princeton: Princeton University Press (2022)
- Ketschik Nora, Overbeck Maximilian, Murr Sandra, Pichler Axel & Blessing André: Interdisziplinäre Annotation von Entitätenreferenzen (2020)
- Neves, Mariana; Ševa, Jurica: An extensive review of tools for manual annotation of documents. In: Brief Bioinform 22 (1), S. 146–163. DOI: 10.1093/bib/bbz130. (2021)
- Reiter Nils, Pichler Axel & Kuhn Jonas (eds.), Reflektierte Algorithmische Textanalyse: Interdisziplinäre Arbeiten in der Creta-Werkstatt. De Gruyter. pp. 203-236 (2020)
- Salganik, Matthew J.: Bit by Bit: Social Research in the Digital Age Gebundene Ausgabe. (2017)
- Scharkow, Michael: Automatische Inhaltsanalyse und maschinelles Lernen (2012)
- Wang, Hailin; Qin Ke; Zakari, Rufai Yusuf; Lu, Guoming; Yin, Jin: Deep neural network-based relation extraction: an overview (2022)
- Qin, Han; Tian, Yuanhe; Song, Yan: Enhancing Relation Extraction via Adversarial Multi-task Learning (2022)
- Zhang, Mike; Nørgaard Jensen, Kristian; Sonniks, Sif Dam; Plank, Barbara: SKILLSPAN: Hard and Soft Skill Extraction from English Job Postings (A2022)

Annotationsguide Stellenanzeigen

Iteration 2, Fassung 10.01.2023

BIBB Stellenanzeigenanalysen

Stefan Winnige
Johanna Binnewitt
Kai Krüger
Timo Schnepf

Inhaltsverzeichnis

Glossar	4
Einleitung	5
1 Allgemeine Regeln (§ A)	6
§ A 1 Grundlegendes	6
§ A 1.1 Umso expliziter und mit weniger Vorannahmen annotiert wird, je besser.	6
§ A 1.2 Ist unklar, ob ein Term Teil eines Konzepts ist, Begriff im Internet recherchieren.	6
§ A 1.3 Unklarheiten sind in „DATUM_unklarheiten_protokoll_iterX.xls“ zu erfassen.	6
§ A 1.4 Terme mit Tippfehler werden mit annotiert.	6
§ A 1.5 Mehrfach genannte Terme bzw. sich wiederholende Terme werden alle annotiert.	6
§ A 1.6 KRen, die nicht auf Wortebene vorliegen sondern als Substring, werden in einer eigenen Kategorie annotiert.	6
§ A 1.7 Widersprechen sich KRen, werden alle annotiert.	6
§ A 2 Die KR muss einen direkten Bezug zur auszuübenden Stelle aufweisen.	7
§ A 3 Span der KR so kurz wie möglich halten, solange maximale Informationen erhalten bleiben.	7
§ A 3.1 Alle relevanten (Zusatz-)Informationen werden annotiert.	7
§ A 3.2 Subjekte (z.B. „Sie“, „du“) nicht mit annotieren, sondern ausschließlich Verb-Objekt-Beziehung.	7
§ A 3.3 Wenig aussagekräftige Verben sowie Modalverben werden i.d.R. nicht annotiert.	7
§ A 3.4 Artikel werden nicht annotiert.	8
§ A 3.5 Datums-, Zeit- und Ortsangaben werden nicht annotiert.	8
§ A 3.6 KRen werden getrennt erfasst, wenn diese eindeutig als selbstständige Einheiten aufgefasst werden können.	8
§ A 3.7 Nähere Beschreibungen von KRen (Detailauflistungen) werden separat als eigenständige KRen annotiert.	8
§ A 3.8 KRen werden getrennt erfasst, wenn diese mit einer „Oder“-Verbindung verknüpft sind.	9
§ A 3.9 KRen werden getrennt erfasst, wenn eine Verschachtelung mehrdeutig ist.	9
§ A 4 Eindeutig verschachtelte KRen werden in einer Span annotiert.	9
§ A 5 Eine Span kann sich auf mehrere Konzepte gleichzeitig beziehen (Doppelreferenz).	10
§ A 5.1 Die KRen können dabei vollständig deckungsgleich sein.	10
§ A 5.2 Die KRen können dabei ineinander verschachtelt sein.	10
2 Kompetenzen (§ K)	11
§ K 1 Grundlegendes	13
§ K 1.1 Zertifikate werden als Kompetenz annotiert.	13
§ K 1.2 Interesse/Offenheit wird als Kompetenz annotiert.	13
§ K 2 Zusatzinformationen: Erfahrungsniveaus/Dringlichkeit/Verstärker werden als Teil der Kompetenz annotiert.	13
§ K 2.1 Begriffe, die das Erfahrungsniveau einer Kompetenz beschreiben, werden immer gemeinsam mit der Kompetenz annotiert.	13
§ K 2.2 Begriffe, die die Dringlichkeit einer Kompetenz beschreiben, werden immer gemeinsam mit der Kompetenz annotiert.	14
§ K 3 Adjektivierte Kompetenzen werden annotiert.	14
3 Tätigkeiten (§ T)	15
§ T 1 Grundlegendes	16
§ T 1.1 „Allgemeinfloskeln“ ohne eigentliche Information nicht annotieren	16
§ T 1.2 „Arbeiten im Team“ ist eine Arbeitsbedingung und keine Tätigkeitsbeschreibung.	16
§ T 1.3 Ansprechpartner sein beschreibt eine Tätigkeit und ist daher als solche zu annotieren	16
§ T 2 Zusatzinformation: Autonomie einer Tätigkeit wird mit annotiert.	16
§ T 3 Zusatzinformation: Umfang einer Tätigkeit wird mit annotiert.	16
4 Arbeitsmittel (§ AM)	17
§ AM 1 Grundlegendes	18
§ AM 1.1 Firmenfahrzeuge sind Arbeitsmittel.	18
§ AM 1.2 Arbeitskleidung ist ein Arbeitsmittel.	18

§ AM 2	Output, der durch die Tätigkeit produziert wird, ist kein Arbeitsmittel	18
§ AM 3	Inputmaterialien wie Zutaten und andere Rohstoffe sind keine Arbeitsmittel.	18
§ AM 4	Schriftliche und potentiell schriftliche Hilfsmittel wie Konzepte, Anleitungen oder Gesetz- zestexte sind Arbeitsmittel. Zertifikate nicht.	18
5	Arbeitsbedingungen (§ AB)	20
§ AB 1	Grundlegendes	20
§ AB 1.1	Weiterbildungen, Fortbildungen usw. sind Arbeitsbedingungen.	20
§ AB 1.2	Angaben zur Arbeitszeit/Schichtarbeit sind Arbeitsbedingungen.	20
§ AB 1.3	Befristung ist eine Arbeitsbedingung.	20
§ AB 1.4	Teamarbeit ist eine Arbeitsbedingung.	21
§ AB 1.5	Kundenkontakt ist eine Arbeitsbedingung.	21
§ AB 2	Formulierungen zur allgemeinen Anforderung der Tätigkeit werden als Arbeitsbedingung annotiert.	21
§ AB 3	Mobilität ist eine Arbeitsbedingung.	21
6	Berufsbezeichnung (§ BB)	22
§ BB 1	Grundlegendes	22
§ BB 1.1	Berufsbezeichnungen umfassen die Anzahl der gesuchten Stellen, d.h. es wird „ein/e“ mitannotiert.	22
§ BB 1.2	Berufsbezeichnungen umfassen die vollständige Gender-Formulierung.	22
§ BB 2	Fachrichtung oder Spezialisierung werden als Teil der Berufsbezeichnung annotiert, Bran- chen und Sektoren alleine nicht.	22
§ BB 3	Berufsbezeichnungen, die innerhalb einer Zertifikatsnennung auftauchen, werden als Be- rufsbezeichnung annotiert.	23
7	Berufserfahrung (§ BE)	24
§ BE 1	Berufserfahrung wird nur annotiert, wenn sie nicht an eine Kompetenz gekoppelt ist. . . .	24

Glossar

Konzept

Theoretisches Konstrukt, abstraktes Phänomen, welches extrahiert werden soll. Wir unterscheiden zwischen den Konzepten Kompetenzen, Arbeitsmitteln, Tätigkeiten, ...(alle aufzählen) Jedes Konzept hat eine Definition. Konzepte können sich sprachlich auf vielfältige Art äußern in Form von Konzept-Referenzen. So wären Bürosoftwarekenntnisse, Lötkenntnisse, Deutschkenntnisse Beispiele für das Konzept Kompetenz.

Unterkategorie

Konzepte besitzen Unterkategorien. D.h. sie haben eine hierarchische Struktur, welche, zukünftig noch zu erstellen ist (Ontologie, Kategorieschema, usw.). Beispielsweise sind „Zertifikate“ oder „Sozialkompetenzen“ Unterkategorien des Konzepts „Kompetenz“.

Konzept-Referenz (KR)

Textliche Repräsentationen des Konzeptes, welches wir extrahieren wollen. Also „Span“, welche die zu extrahierende(n) Information(en) enthält.

Span

Zu annotierender, bzw. markierender Text.

„verschachtelt“

Bezieht sich auf KRen (desselben Konzepts), die sich Teile eines Satzes teilen.

Doppelreferenz

Bezieht sich auf Spans, welche sich auf mehr als nur ein Konzept beziehen.

Zusatzinformation

Zusatzinformationen sind Informationen, welche die KR genauer beschreiben. Für Kompetenzen sind das

1. Erfahrungsniveaus. Bspw. „*Einschlägige* Englischkenntnisse“.
2. Dringlichkeiten. Bspw. „*Programmierkenntnisse zwingend erforderlich.*“
3. Verstärker. Bspw. „*sehr gute* Kenntnisse in Löten.“

Für Tätigkeiten sind das

1. Autonomie. Bspw. „*Eigenverantwortliche* Aquirse von Neukunden“.
2. Umfang. Bspw. „*teilweise* Budgetverwaltung“.

Iteration und Diskussion

Eine Iteration bzw. Iterationsrunde beginnt damit, dass Stellenanzeigen vom Team annotiert werden. Sie endet dann, wenn die Guideline aktualisiert wurde und alle Unklarheiten entweder geklärt oder vertagt wurden. Innerhalb einer Iterationsrunde gibt es mehrere Diskussionen. Die erste Diskussion einer Iteration behandelt die „Annotationsabweichungen“ der Annotatoren. Jede Abweichung mündet entweder in einer „lesson learned“, also einer Entscheidung, wie mit bestimmten Fällen zu verfahren ist oder mit einer Ausführung und Vertagung des Problems ins Unklarheiten-Dokument. Pro Iteration gibt es mehrere Diskussionsrunden, welche von dem/der Guideline-Verantwortlichen organisiert und geleitet werden.

Unklarheiten

Unklarheiten sind offene Diskussionsenden, welche nicht mit dem jetzigen Wissen und der aktuellen Diskussion geklärt werden können. Sie werden im „Unklarheiten-Dokument“ gesammelt

Einleitung

Das Stellenanzeigen Projekt des BIBB hat das Ziel, unstrukturierte Informationen aus online Stellenanzeigen in strukturierte Daten zu überführen. Die vorliegenden Annotations-Guidelines erfüllen dabei mehrere Funktionen:

1. Sie definieren sowohl deduktiv als auch induktiv *welche* Informationen extrahiert werden sollen.
2. Sie beschreiben, *wie* diese Informationen konkret extrahiert werden sollen.
3. Sie dienen als *Nachschlagewerk* für Annotatoren.
4. Sie dienen der Diskussion und Präzisierung von 1. und 2. und sind damit notwendige menschliche Vorarbeit für die folgende computergestützte Informationsextraktion. Erst eine genaue Definition und Eingrenzung der zu extrahierenden Konzepte erlaubt die Güte maschineller Algorithmen zu evaluieren.
5. Sie dienen der *wissenschaftlichen Transparenz* und Interoperabilität der Daten.

Für Annotatoren ist vor allem die Guideline als Nachschlagewerk wichtig. Es ist nicht möglich, alle Vorgaben von Beginn an vollständig zu erfüllen. Beim ersten Lesen der Guidelines wird vermutlich der Kontext fehlen, also die zu annotierende Stellenanzeige. Daher wird empfohlen, dass die Guidelines gemeinsam mit einem/einer erfahrenen Annotator*in durchgesprochen werden. Danach werden gemeinsam einige Stellenanzeigen annotiert. Sobald sich der/die neu eingewiesene*r sicher fühlt, kann die Person mithilfe der Guidelines alleine weiter annotieren.

Abschnitt 1 „Allgemeine Regeln“ beschreibt Annotationsregeln, welche Konzeptunabhängig sind. Paragraph § A 1 bis § A 3 sind eher technischer Natur, verweisen bspw. auf den Umgang mit Mehrfachnennungen (§ A 1.4) oder auf widersprüchliche Konzeptreferenzen (§ A 1.6). In § A 3 bis § A 5 geht es vermehrt um die Spanlänge, also was Teil der Informationsextraktion sein soll, und was nicht. Vor allem Bezüge zur „Verschachtelung“ von KRen sowie von Konzepten erscheinen womöglich etwas unübersichtlich. Dies liegt an der Natur der Sprache und dem Problem, dass sich zusammenhängende Informationen in „verschachtelter“ Form in Stellenanzeigen-Sätzen wieder finden. Diese § sortieren mögliche Fälle und geben Anweisung, wie mit ihnen umzugehen ist. § A 6 beschreibt letztlich noch den Umgang mit Detailauflistungen.

Nachfolgend werden die sechs Konzepte näher beschrieben und welche Herausforderungen bei deren Annotation auftreten können. Neben einer kurzen theoretischen Definition der Konzepte findet sich hier auch beispielhaft ein Überblick über bereits zugewiesene KRen zum Konzept. Diese Aufzählungen sollen bei unklaren Begriffen helfen, ob diese oder ähnliche Begriffe bereits in der Vergangenheit valide zugewiesen wurden.

Die Paragraphen besitzen immer diesselbe Gliederungsstruktur:

§ Paragraph

§ Unterparagraph

- An dieser Stelle steht ein Einleitungs- und Erklärungssatz zum Paragraphen, wenn dieser für sich nicht selbsterklärend ist.
 - **Beispiel 1:** „Hier steht ein Beispieltext aus einer Stellenanzeige, welcher den Paragraphen verdeutlicht. Es kann mehrere Beispiele pro Paragraph geben.“
 - * Spezifische Beispiele können in diesem Abschnitt noch etwas detaillierter beschrieben werden.
 - **Gegenbeispiel 1:** „Diese Beispiele beschreiben Problemfälle, und wo bspw. nicht annotiert werden sollte.“
 - * Auch Gegenbeispiele können hier näher beschrieben sein.

1 Allgemeine Regeln (§ A)

§ A 1 Grundlegendes

§ A 1.1 Umso expliziter und mit weniger Vorannahmen annotiert wird, je besser.

- **Beispiel 1:** „Sie [arbeiten gemeinsam im Team]_{AB}“
 - * Implizit könnte man annehmen, dass hier auch die Kompetenz „Teamarbeit“ referenziert wird. Der sprachliche Kontext bezieht sich jedoch auf die Arbeitsbedingung, nicht auf eine implizite Anforderung, dass der Bewerber „teamfähig“ zu sein hat. Daher keine Kompetenz-Annotation.
- **Beispiel 2:** „Sie nutzen [Adobe Photoshop]_{AM}“
 - * Explizit wird nur ein Arbeitsmittel genannt, es wird keine Kompetenzanforderung formuliert und ist somit auch nicht als solche zu annotieren.

§ A 1.2 Ist unklar, ob ein Term Teil eines Konzepts ist, Begriff im Internet recherchieren.

- Es können immer wieder Begriffe in Stellenanzeigen fallen, die unbekannt sind, bspw. unbekannte Software (Arbeitsmittel). Falls Konzeptzuweisung unklar, bitte Begriff online recherchieren.

§ A 1.3 Unklarheiten sind in „DATUM_unklarheiten_protokoll_iterX.xls“ zu erfassen.

- In „DATUM_unklarheiten_protokoll_iterX.xls“ werden die Unklarheiten, die auch nach einer Internetrecherche nicht aufgelöst werden konnten, getrennt nach Konzepten erfasst. Zur Identifizierung der betroffenen Stellenanzeige wird die Stea ID genutzt. Diese findet sich in Inception im Dokumentnamen (oben links).

§ A 1.4 Terme mit Tippfehler werden mit annotiert.

§ A 1.5 Mehrfach genannte Terme bzw. sich wiederholende Terme werden alle annotiert.

§ A 1.6 KRen, die nicht auf Wortebene vorliegen sondern als Substring, werden in einer eigenen Kategorie annotiert.

- KRen, welche Teil von Wörtern sind (Substrings) werden mit der Kategorie-Ausprägung: „KONZEPT_part“ auf Wortebene annotiert. Meist handelt es sich um Arbeitsmittel, Berufsbezeichnungen oder Arbeitsbedingungen. Grund dafür ist, dass Inception Annotationen nur entweder auf Token- oder auf Zeichenebene für alle Kategorien zulässt.
 - **Beispiel 3:** „Sie besitzen [[MS-Officekenntnisse]_{AM_P}]_K] und sind [[MS-Wordfachmann]_{AM_P}]_K“
 - **Beispiel 4:** „...zum sofortigen Eintritt mehrere [[CNC-Maschinenbediener]_{AM_P} (m/w)]_{BB}“
 - **Beispiel 5:** „[Exceltabellen]_{AM_P}“
 - **Beispiel 6:** „[Bäckerlehre]_{BB_P}“

§ A 1.7 Widersprechen sich KRen, werden alle annotiert.

- Es kann vorkommen, dass sich die Aussagen in einer Stellenanzeige widersprechen. In diesem Fall werden alle Aussagen annotiert.
 - **Beispiel 7:** „Sie verfügen über [mindestens fünf Jahre Berufserfahrung]_{BE}. [...] [Berufseinsteiger]_{BE} sind bei uns willkommen!“
 - * fünf Jahre Berufserfahrung und Berufseinsteiger widersprechen sich.

§ A 2 Die KR muss einen direkten Bezug zur auszuübenden Stelle aufweisen.

- Bezieht sich die KR (hier Arbeitsmittel) auf die beschriebene Stelle, oder auf die Firma, die Abteilung, Zulieferer etc.? Wenn kein direkter Stellenbezug ersichtlich ist, oder unklar, ob KR auf ausgeschriebene Stelle referenziert, nicht annotieren.
 - **Beispiel 8:** „Unsere Firma XY spezialisiert sich auf Weblösungen mit D3 Bibliotheken“
 - * Daraus kann nicht geschlossen werden, dass in der *eigentlichen Tätigkeit* tatsächlich mit D3 programmiert wird. Potenzielle KRen, die sich in Unternehmensbeschreibungen u.ä. befinden, werden nicht annotiert, wenn sie keinen direkten Tätigkeitsbezug haben.
 - **Beispiel 9:** „spannende Tätigkeit in einem innovativen Unternehmen“
 - * Es handelt sich um eine Unternehmensbeschreibung, von der keine Rückschlüsse auf allgemeine „innovative“ Arbeitsbedingungen der ausgeschriebenen Stelle gezogen werden können.
 - **Beispiel 10:** „Für unseren Kunden suchen wir Mitarbeiter für die Industrieverpackung.“
 - * „Industrieverpackung“ bezieht sich auf die Branche. Annotieren erfordert genaues lesen, sonst könnte fälschlich davon ausgegangen werden, es handelt sich hier um eine Tätigkeit.

§ A 3 Span der KR so kurz wie möglich halten, solange maximale Informationen erhalten bleiben.

§ A 3.1 Alle relevanten (Zusatz-)Informationen werden annotiert.

- **Beispiel 11:** „Sie [sprechen fließend Arabisch]_K“ versus
 - * Würde man im dagegen wie folgt annotieren: Sie sprechen fließend [Arabisch]_K ginge die Zusatzinformation zum Sprachniveau (”fließend”) verloren. Es wird nicht die maximale Information der KR erfasst. Was im konkreten Fall ”maximale Information” bedeutet, und wann die Spanlänge so kurz wie nötig zu halten ist, ist nicht immer eindeutig und muss nach eigenem Ermessen entschieden werden. Unsicherheiten bitte immer dokumentieren.
- **Beispiel 12:** „[An Probleme gehst Du strategisch heran, um sie zu lösen]_K“
 - * In Beispiel 12 ist die Kompetenz ”Probleme lösen” in einer langen Span beschrieben. Spans können länger sein, falls notwendig.
- **Beispiel 13:** „Dein Aufgabengebiet umfasst ..., das [Prüfen und die Behebung von Störungen [vor Ort]_{AB} oder per Support]_T.“
- **Beispiel 14:** „Aufgabengebiet: [Reinigen sämtlicher Bereiche im Unternehmen]_T.“
 - * Auch genauere Angaben zu Tätigkeiten sind mit zu annotieren. „vor Ort“ aus Beispiel 13 beschreibt darüber hinaus die Arbeitsbedingung ”Mobilität”.

§ A 3.2 Subjekte (z.B. „Sie“, „du“) nicht mit annotieren, sondern ausschließlich Verb-Objekt-Beziehung.

- **Beispiel 15:** „Sie [ermitteln Benutzeranforderungen]_T, [planen und erstellen Anwendungslösungen]_T...“
 - * „Sie“ ist nicht Teil der KR.

§ A 3.3 Wenig aussagekräftige Verben sowie Modalverben werden i.d.R. nicht annotiert.

- Verben wie „haben“, „arbeiten“, „sind“, „dürfen“, „sollen“, „können“, „müssen“, „wollen“, „mögen“, usw. werden i.d.R. nicht annotiert, es sei denn sie sind für das Verständnis der KR unerlässlich.
 - **Beispiel 16:** „Sie arbeiten mit [CNC-Maschinen]_{AM}“,

- **Beispiel 17:** „Sie haben [mehrjährige Erfahrung in [PHP]_{AM}-Programmierung]_K“
- **Beispiel 18:** „Sie sollten [[Java-Script]_{AM}]_K beherrschen“
 - * Die jeweiligen Verben sind nicht von Bedeutung für das Verständnis der Konzeptkategorie.
- **Gegenbeispiel 1:** „Sie [arbeiten mit Patienten.]_{AB}“
 - * Hier ist das Verb „arbeiten“ entscheidend für das Verständnis der Konzept-Referenz.

§ A 3.4 Artikel werden nicht annotiert.

- **Beispiel 19:** „der [Umgang mit elektronischen Kassen]_T beherrschen“
 - * Der Artikel „der“ wird nicht annotiert.

§ A 3.5 Datums-, Zeit- und Ortsangaben werden nicht annotiert.

- **Beispiel 20:** „Für die Dauer einer [Schwangerschaftsvertretung]_{AB} zunächst bis Ende Februar 2019...“
 - * Angaben zur Befristungsdauer werden nicht annotiert. Das Signalwort genügt.

§ A 3.6 KRen werden getrennt erfasst, wenn diese eindeutig als selbstständige Einheiten aufgefasst werden können.

- **Beispiel 21:** „Sie sind [flexibel]_K und [belastbar]_K“
 - * „flexibel“ und „belastbar“ können als eigenständige KRen des Konzepts Kompetenz aufgefasst werden.

§ A 3.7 Nähere Beschreibungen von KRen (Detailauflistungen) werden separat als eigenständige KRen annotiert.

- Werden KRen näher beschrieben, z.B. durch Klammern, Bindestriche oder andere Formen der Aufzählung gilt § A 3.8 Es wird immer die kleinstmögliche KR annotiert. Es werden also sowohl der Überbegriff als auch die Unterbegriffe getrennt annotiert.
 - **Beispiel 22:** „Sie kümmern sich um den [Kundensupport]_T ([Termine koordinieren]_T, [Beratungsgespräche führen]_T, [Reklamationen annehmen]_T)“
 - **Beispiel 23:** „[gute IT-Kenntnisse]_K, u. a. in den Bereichen [[MS Office]_K]_{AM}, [[DATEV]_K]_{AM} und [[Simba]_K]_{AM}.“
 - **Beispiel 24:** „Sie bringen [gute Kenntnisse in [Designprogrammen]_{AM}]_K mit wie z.B. [CAD]_{AM}.“
- Manche KRen lassen sich nicht getrennt annotieren, da die Überkategorie direkten Bezug zur Unterkategorie hat.
 - **Beispiel 25:** „[Erfahrung im Kundendienst (Annahme und Bearbeitung)]_T“
 - * ohne die Bezugnahme zu „Kundendienst“ wird nicht klar, was angenommen und bearbeitet wird. Es könnten beispielsweise auch Pakete angenommen und bearbeitet werden.
 - **Beispiel 26:** „[Kenntnisse der forstrelevanten Rechtsnormen, z.B. [Bundes- u. Hess. Landeswaldgesetz]_{AM}, [Bundes- u. Hess. Landesnaturschutzgesetz]_{AM}, [PEFC-Zertifizierungssystem]_{AM}]_K.“
 - * Hier bezieht sich die Auflistung der Rechtsnormen auf die Kompetenz „Kenntnisse“. Die Kompetenz lässt sich nicht getrennt annotieren. Die darin enthaltenen Arbeitsmittel jedoch schon (vgl. § A 5.2).

§ A 3.8 KRen werden getrennt erfasst, wenn diese mit einer „Oder“-Verbindung verknüpft sind.

- **Beispiel 27:** „Sie haben eine [abgeschlossene Ausbildung]_K oder aber [Berufserfahrung]_{BE} in dem Bereich“
 - * „Oder“ signalisiert sprachlich Substituierbarkeit von KRen. Damit ließen sich bspw. austauschbare Kompetenzen identifizieren. Es wird jedoch festgelegt, dass diese besonderen Fälle beim Annotieren nicht berücksichtigt werden¹.

§ A 3.9 KRen werden getrennt erfasst, wenn eine Verschachtelung mehrdeutig ist.

- **Beispiel 28:** „Sie sind [sehr lernfähig]_K und [belastbar]_K“
 - * „lernfähig“ und „belastbar“ sind eigenständige Kompetenzen. Jedoch lässt der Text beide inhaltlichen Interpretationen zu, dass sich der Verstärker „sehr“ entweder nur auf „lernfähig“ bezieht oder auch auf „belastbar“ (sprachlich mehrdeutig bzw. ambig). Es wird festgelegt, dass sich der Verstärker nur auf den ersten Teil der Und-Verknüpfung bezieht, daher werden die KRen separat annotiert.
- **Beispiel 29:** „[Medizinische Versorgung] und [Notfallhilfe]“
 - * „Medizinische“ kann sich syntaktisch und semantisch auch auf „Notfallhilfe“ beziehen. Wie oben, wenn nicht eindeutig ist, ob Begriff für beide KRen gilt, werden sie getrennt annotiert.

§ A 4 Eindeutig verschachtelte KRen werden in einer Span annotiert.

- In manchen Fällen kann es vorkommen, dass sich verschiedene KRen auf die gleichen Textteile beziehen und es dadurch nicht möglich ist, diese beim Annotieren zu trennen. KRen können auf drei Arten miteinander verbunden sein. Erstens unverschachtelt und eindeutig (§ A 3.8), zweitens verschachtelt und mehrdeutig (§ A 3.9) und drittens verschachtelt und eindeutig wie in den folgenden Beispielen (§ A 4).
→ Eine genaue Prüfung des Satzes kann nötig sein, welcher der drei Fälle zutrifft.

- **Beispiel 30:** „Sie [kommunizieren offen und vertrauensvoll]_K“
 - * „vertrauensvoll“ bezieht sich hier eindeutig auf die Kommunikation, daher können beide Einheiten nicht getrennt werden. Anders ist es allerdings in folgendem Beispiel.
- **Gegenbeispiel 2:** „Sie [kommunizieren offen]_K und sind [vertrauensvoll]_K“
 - * Hier sind die Kompetenzen zwar ebenfalls durch ein „und“ verbunden, sie besitzen aber jede ihr eigenes Verb, sodass sie syntaktisch unabhängig voneinander sind wie in § A 3.8.
- **Beispiel 31:** „Sie [bearbeiten Produkte und überwachen deren Verkauf]_T“
 - * „Deren“ ist Korreferenz auf „Produkte“, daher muss auch diese annotiert werden, d.h. ganze Span zu markieren. D.h. bei Anaphern und Kataphern ist die Referenz mit zu annotieren (ganzer Span). Bezüge zu vorangegangenen Objekten (?) werden mitannotiert.
- **Beispiel 32:** „[Instandsetzung, Wartung und Pflege von Maschinen und Anlagen]_T“
 - * Instandsetzung, Wartung und Pflege können sich jeweils auf Maschinen und Anlagen beziehen (3 Tätigkeiten an 2 Objekten = 6 Tätigkeitsbeschreibungen). Daher werden alle KRen in einer Span gemeinsam annotiert. Diese Regel gilt sowohl für „und“ als auch „oder“ Verbindungen.

¹Folgeanalysen auf Stellenanzeigen-Level müssten mit bspw. Mustersuche diese Substitute wieder auffinden.

§ A 5 Eine Span kann sich auf mehrere Konzepte gleichzeitig beziehen (Doppelreferenz).

§ A 5.1 Die KRen können dabei vollständig deckungsgleich sein.

- Die Span von zwei KRen kann sich vollständig entsprechen.
 - **Beispiel 33:** „Für den Kundenkontakt vor Ort benötigen Sie ein [[[Fahrzeug]_{AM}]_{AB}].“
 - * „Fahrzeug“ ist sowohl Arbeitsmittel als auch Arbeitsbedingung. Die Spans sind gleich.
 - **Beispiel 34:** „Sie haben die [[[Bäckerlehre]_{BB_P}]_K] erfolgreich abgeschlossen.“
 - * „Bäckerlehre“ beschreibt ein Zertifikat und wird damit den Kompetenzen zugeordnet. Gleichzeitig versteckt sich aber auch mit „Bäcker“ eine Berufsbezeichnung im Substring. Da wir auf Wortebene annotieren, wird in diesem Fall _{BB_P} annotiert (siehe § A 1.6). Dadurch sind die KRen deckungsgleich.

§ A 5.2 Die KRen können dabei ineinander verschachtelt sein.

- Erklärung.
 - **Beispiel 35:** „[[Grundkenntnisse in [MS-Office]_{AM}]_K Anwendungen“
 - * Hier ist einmal die KR „Grundkenntnisse in MS-Office“ als Kompetenz und innerhalb derer die KR „MS-Office“ als Arbeitsmittel zu annotieren. Die Spans sind ineinander verschachtelt.
 - **Beispiel 36:** „[Erstellen von technischen Zeichnungen mit [AutoCAD]_{AM}]_T“
 - * Falls die Tätigkeit ein Arbeitsmittel beinhaltet, ist dieses Teil der Tätigkeits-Span und wird zusätzlich als Arbeitsmittel annotiert. Auch hier sind die Spans ineinander verschachtelt.
- Beispiel 37:** „[Modulübergreifendes Prozesswissen in den [SAP-Kernmodulen MM/PP]_{AM}]_K“

2 Kompetenzen (§ K)

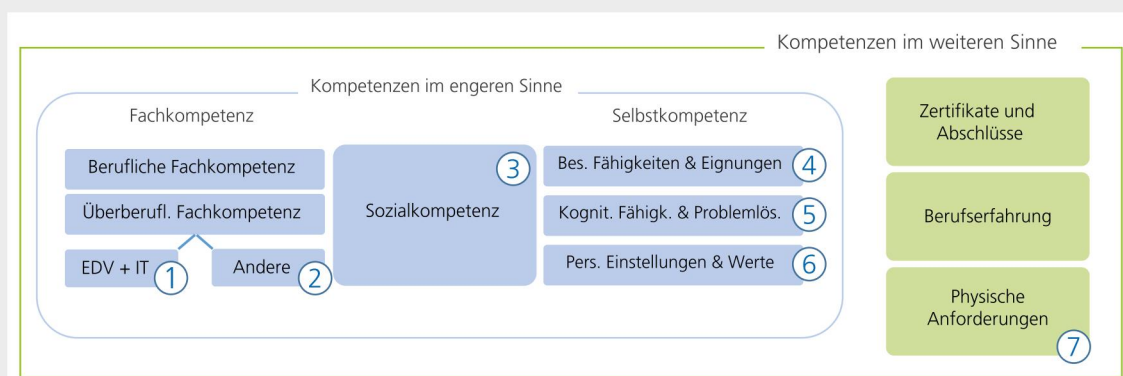
Definition

Im Rahmen der *ersten Iteration* wird zunächst aus der deduktiv-theoretischen Perspektive in das Konzept „Kompetenz“ eingeführt. Danach werden Beispiele aus Stellenanzeigen genannt, um ein möglichst umfassendes Verständnis dafür zu generieren, was unter einer Kompetenz im Rahmen dieses Projektes verstanden wird.

Weinert (2014) versteht unter einer Kompetenz folgendes:

- [D]ie bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können. Als Erträge des schulischen Unterrichts kann man zwischen folgenden Kompetenzen unterscheiden:
 - fachliche Kompetenzen (z. B. physikalischer, fremdsprachlicher, musikalischer Art),
 - fachübergreifende Kompetenzen (z. B. Problemlösen, Teamfähigkeit),
 - Handlungskompetenzen, die neben kognitiven auch soziale, motivationale, volitionale und oft moralische Kompetenzen enthalten und es erlauben, erworbene Kenntnisse und Fertigkeiten in sehr unterschiedlichen Lebenssituationen erfolgreich, aber auch verantwortlich zu nutzen.

Abbildung 16: BIBB-Kompetenzschema



Quelle: eigene Darstellung

Abbildung 1: BIBB Kompetenzschema aus Köhne-Finster et al. 2020: 76

Köhne-Finster et al. (2020: 68) erweiterten diese Kompetenzdefinition um den „Kompetenzbegriff im weiteren Sinne“ für das „BIBB-Kompetenzschema“²:

Neben den Kompetenzen werden weiter „Zertifikate und Abschlüsse“, „Berufserfahrung“ und „physische Anforderungen“ mit einbezogen. Für die Betrachtung der Anforderungen in Stellenanzeigen wird der Kompetenzbegriff im weiteren Sinne (i. w. S.) eingeführt: Anforderungen in Stellenanzeigen sind die Gesamtheit an physischen, psychischen und kognitiven Eigenschaften sowie Berufserfahrung, die ein Arbeitgeber von einer potenziellen Bewerberin oder einem potentiellen Bewerber erwartet bzw. die er voraussetzt, um die Aufgaben der ausgeschriebenen Stelle optimal zu erfüllen.

Im Folgenden werden Konzept-Referenzen bzw. Spans genannt, welche bereits aus Stellenanzeigen extrahiert wurden. Die Aufzählung ist nicht erschöpfend und darf im Zuge weiterer Iterationen ergänzt werden, vor

²Die Herkunft und Struktur des BIBB-Kompetenzschemas ist unter Köhne-Finster et al. 2020 beschrieben.

allem bei strittigen KRen. Beachte auch, dass Kompetenzen in Stellenanzeigen in allen Aggregatsebenen genannt werden können. D.h. Arbeitgeber können sehr abstrakt eine Kompetenz fordern (z.B. „Programmieren“) und/oder sehr konkret (z.B. „Python 3.8 oder höher mit Expertise in pyTorch“). Der besseren Übersichtlichkeit halber sind die KRen grob kategorisiert (angelehnt an Abbildung 1).

- Fachkompetenz
 - Verkaufskompetenz
 - Beratungskompetenz
 - Handwerkliches Geschick (z.B. Leiterplatten löten, Reifen auswuchten, usw.)
 - Denken können (strategisch, analytisch, innovativ, usw.)
 - Produktkenntnisse
 - Hardware- und Software Kenntnisse
 - Büroorganisation-Büromanagement-Kenntnisse
- Sozialkompetenz
 - Empathiefähigkeit und Einfühlungsvermögen
 - Freundlichkeit
 - Kommunikationsfähigkeit
 - gepflegtes Auftreten
 - Zuverlässigkeit
- Selbstkompetenz
 - informelle Sprachkompetenz (mit/ohne bestimmtem Niveau)
 - Generelle gesundheitliche Eignung
 - Flexibilität
 - Bereitschaft für Schichtarbeit/Dienstreisen/Fortbildung/usw.
 - Gewissenhaftigkeit
 - Eigeninitiative und selbständiges Arbeiten
 - Engagement, Freude und Interesse (vgl. § K 1 .2)
 - Leitung und Mitarbeiterführung
 - Koordinationsfähigkeit
 - Durchsetzungsvermögen
- Zertifikate und Abschlüsse (vgl. § K 1 .1)
 - Schulische Abschlüsse (z.B. Hauptschulabschluss, Abi, usw.)
 - Fortbildungsabschlüsse (z.B. Meister, Techniker, usw.)
 - Berufliche Abschlüsse (z.B. Tischler, Studium der Architektur, usw.)
 - formale Sprachzertifikate (z.B. Toeffel Test)
 - Führerscheine
 - Polizeiliche Führungszeugnisse
 - Jagdschein
- Kompetenzbezogene Berufserfahrung (vgl. § K 2)
- Physische Anforderungen
 - Psychische und Physische Belastbarkeit

§ K 1 Grundlegendes

§ K 1.1 Zertifikate werden als Kompetenz annotiert.

- Schul-, Ausbildungs- und Studiumsabschlüsse werden als Kompetenz annotiert, solange sie als Anforderung formuliert sind. Dies gilt auch für andere Zertifikate wie Sprachzertifikate oder Führerscheine. Sie werden nicht annotiert, wenn sie erst bei Ausübung der Stelle zukünftig erworben werden (können).
 - **Beispiel 1:** „Sie verfügen über eine [abgeschlossene [Bäcker-Lehre]_{BB_P}]_K.“
 - * Die Bäcker-Lehre ist hier als Voraussetzung formuliert, daher handelt es sich um ein zu annotierendes Zertifikat unter dem Konzept Kompetenz. Berufsabschluss-Zertifikate oft auch Berufsbezeichnungen (vgl. § A 5 .1). „Abgeschlossen“ wird mit annotiert.
 - **Beispiel 2:** „Sie sind [examiniertes [Gesundheits- und Krankenpfleger]_{BB}]_K in Voll- oder Teilzeit.“
 - **Beispiel 3:** „Sie verfügen über einen [Führerschein der Kl. 3 AB]_K.“
 - * Zusatzinformationen zum Zertifikat wie „examiniert“, „diplomiert“, „staatlich anerkannt“ werden mit annotiert. Auch Details zu Führerscheinen und Sprachzertifikaten.
 - **Beispiel 4:** „Sie haben [Ingenieurwissenschaften mit Fachrichtung Elektrotechnik]_K studiert.“
 - * Auch Studiumsabschlüsse mit Fachrichtungen sind Zertifikate. Sie sind keine Berufsbezeichnungen, vgl. § BB 3.
 - **Gegenbeispiel 1:** „Am Ende der Ausbildung steht das Zertifikat zum Bäcker-Lehrling.“
 - * Das Zertifikat zum Bäcker-Lehrling ist hier nicht als Anforderung, sondern als Produkt der Ausbildung, formuliert.
 - **Gegenbeispiel 2:** „Während der Einarbeitungsphase haben Sie die Möglichkeit einen Stapler-Führerschein zu erwerben.“
 - * Der Stapler-Führerschein wird nicht vorausgesetzt, daher handelt es sich auch nicht um eine Kompetenz.

§ K 1.2 Interesse/Offenheit wird als Kompetenz annotiert.

- Begriffe die auf ein bloßes Interesse am Thema hindeuten („Leidenschaft für...“ „Liebe zu Zweirädern“, „Sie brennen für“, „Bereitschaft“), aber nicht zwingend eine Erfahrung damit voraussetzen, werden als eigenständige Kompetenz erfasst.
 - **Beispiel 5:** „Sie haben [Interesse am Löten]_K.“
 - * Die Person muss nicht zwangsweise Löten können, sie muss nur Interesse daran haben (es zu lernen).
 - **Beispiel 6:** „Sie zeigen [Bereitschaft zur [Weiterbildung]_{AB}]_K.“
 - **Beispiel 7:** „Du ... hast [Freude]_K daran, dich [über neue Technologien und deren Vorzüge zu informieren]_T.“

§ K 2 Zusatzinformationen: Erfahrungsniveaus/Dringlichkeit/Verstärker werden als Teil der Kompetenz annotiert.

§ K 2.1 Begriffe, die das Erfahrungsniveau einer Kompetenz beschreiben, werden immer gemeinsam mit der Kompetenz annotiert.

- **Beispiel 8:** „Sie besitzen [ausgeprägte Erfahrung im Löten]_K.“
 - * Weitere Begriffe für Erfahrungsniveaus:

- (Sehr gute) Kenntnisse in...
- (Einschlägige/Mehrjährige) Erfahrung in...
- Sicherer Umgang/ Sicheres Arbeiten mit...
- Ausgeprägte...
- Eine gute Portion...

§ K 2 .2 Begriffe, die die Dringlichkeit einer Kompetenz beschreiben, werden immer gemeinsam mit der Kompetenz annotiert.

- **Beispiel 9:** „[Erste Erfahrungen im Umgang mit [CNC-Fräsen]_{AM} und [SAP]_{AM} sind wünschenswert]_K.“
 - * Weitere Begriffe für Dringlichkeit:
 - Wünschenswert
 - Idealerweise
 - Mindestens
 - Zwingend erforderlich
 - wichtig ist uns...

§ K 3 Adjektivierte Kompetenzen werden annotiert.

- **Beispiel 10:** „[Zuverlässige]_K, [eigenverantwortliche]_K sowie [selbstständige]_K Arbeitsweise“
 - * „Arbeitsweise“ wird nicht mitannotiert, da die adjektivierte Kompetenz für sich alleine genommen verständlich ist.
- **Beispiel 11:** „„[[Eigenverantwortliche]_K Akquisition und Betreuung von Neu- und Bestandskunden]_T“

3 Tätigkeiten (§ T)

Definition

Im Rahmen der *ersten Iteration* werden hier Tätigkeiten aufgelistet, welche bereits am Text als solche definiert wurden. Diese Induktion ist noch durch theoretische Definitionen und Taxonomien zu ergänzen.

- Administrative Tätigkeiten
 - Koordination von Terminen und Meetings
 - Vor- und Nachbereitung von Workshops
 - Bearbeitung von Kundenaufträgen
 - Buchung von Reisen
 - Verwaltung von Aufträgen
 - Erstellung von Dokumenten
 - Überwachung der Auftragsbearbeitung/ Dienstzeiten
 - Abwicklung und Verbuchung von Konten
- IT-Tätigkeiten
 - Datenbankpflege (z.B. in SAP)
 - Hard- und Software-Beschaffung
 - Programmieren
- Handwerkliche Tätigkeiten
 - Tapezieren /Gipsen/ ...
 - Reinigung von Büroräumen/Industrieanlagen/PKWs
 - Gestaltung von Verkaufsflächen
 - Reparatur
 - Pflanzen- und Landschaftspflege
 - Forstbetriebliche Tätigkeiten
- Maschinen- und anlagenbezogene Tätigkeiten
 - Analyse und Reparatur von Geräten und Baugruppen
 - Anlagenbedienung
 - Durchführung von Tests
 - Montage, Inbetriebnahme, Inspektion und Kalibrierung
 - Mess- und Prüfverfahren durchführen
 - Beseitigung und Dokumentation von Störungen
- Interaktive Tätigkeiten
 - (freundliche) Kundenberatung
 - Kassieren und Abrechnung
 - Verkauf von Waren
 - unterrichten
 - Unterstützung von Kindern und Jugendlichen (z.B. im Schulalltag, bei Bewerbungen, usw.)
 - (enge) Zusammenarbeit mit internen Abteilungen
 - Abstimmung mit anderen Berufsgruppen (z.B. Lehrern, Therapeuten, Fachärzten)
 - Teilnahme an Arbeitskreisen und Gremien
- Logistische Tätigkeiten
 - Auslieferung von Produkten/Geräten/Stoffen
 - Be- und Entladen
 - Verpacken von Gütern und Waren
- Eigenverantwortliche Tätigkeiten (Autonomiebezug)
 - Führungs-/Leitungsaufgaben
 - selbständiges/eigenverantwortliches Arbeiten
 - Überwachen von Prozessen
 - Erstellung von Statistiken/Analysen/Forschungsberichten
- Vorgeschriebene Tätigkeiten (Heteronomiebezug)
 - Arbeiten nach Vorschrift/Vorgabe
 - Auftragskonforme Realisierung von Bauprojekten
 - Untertützung bei ...
 - Erstellung von Statistiken/Analysen/Forschungsberichten

§ T 1 Grundlegendes

§ T 1.1 „Allgemeinfloskeln“ ohne eigentliche Information nicht annotieren

- **Beispiel 1:** „für Unterstützung“
- **Beispiel 2:** „Hilfe im Team“

§ T 1.2 „Arbeiten im Team“ ist eine Arbeitsbedingung und keine Tätigkeitsbeschreibung.

§ T 1.3 Ansprechpartner sein beschreibt eine Tätigkeit und ist daher als solche zu annotieren

- **Beispiel 3:** „[Ansprechpartner für Kunden bei Rückfragen zu Steuersachverhalten]_T“

§ T 2 Zusatzinformation: Autonomie einer Tätigkeit wird mit annotiert.

- Die Autonomie-Information, d.h. eine Info zum Maß der Selbständigkeit einer Tätigkeit, ist oft Teil der Tätigkeitsbeschreibung und wird daher als Teil der Tätigkeit annotiert. Manchmal ist die Autonomiedimension auch nahe an einer Berufsbezeichnung, wird aber bitte als Tätigkeit annotiert, siehe
 - **Beispiel 4:** „[[Eigenverantwortliche]_K Akquisition und Betreuung von Neu- und Bestandskunden]_T“
 - * „Eigenverantwortliche“ deutet auf hohe Autonomie der Tätigkeit hin und ist Teil der Tätigkeits-Span. Gleichzeitig gilt auch § K 3 und die adjektivierte Kompetenz „Eigenverantwortlichkeit“ wird verschachtelt annotiert.
 - **Beispiel 5:** „Sie [koordinieren [eigenverantwortlich]_K Schulungsmaßnahmen]_T.“
 - * „Projektleiter“ beschreibt eine Tätigkeit, nämlich der Leitung der Schulungen. Ähnlich dem nächsten Beispiel.
 - **Beispiel 6:** „Wir suchen ab sofort eine/n [Metallbaumeister/in]_{BB} für eine [[selbstständige]_K Leitung/Führung der Mitarbeiter/innen]_T“
 - **Beispiel 7:** „[Mitarbeit im Lager]_T“
 - * Weitere Begriffe, welche den Autonomiegrad der Tätigkeit beschreiben, sind:
 - Eigenverantwortlich
 - Selbstständiges
 - Mithilfe bei
 - Zuarbeit

§ T 3 Zusatzinformation: Umfang einer Tätigkeit wird mit annotiert.

- Die Umfangs-Information, also eine Info wie umfangreich die Tätigkeit ausfällt, ist oft Teil der Tätigkeitsbeschreibung und wird daher als Teil der Tätigkeit annotiert.
 - **Beispiel 8:** „Ihre Aufgabe: [teilweise Budget- und Ressourcencontrolling]_T.“
 - * „teilweise“ beschreibt den Umfang, also dass diese Tätigkeit keine zentrale Rolle in der auszuübenden Stelle einnimmt.
 - * Weitere Begriffe, welche den Umfang der Tätigkeit beschreiben, sind:
 - Teilweise...
 - Ausführlich(e)...
 - Hauptaufgabe/Nebenaufgabe...

4 Arbeitsmittel (§ AM)

Definition

Es findet sich bei Troll (2002) keine Definition eines Arbeitsmittels. Diese wird implizit vorausgesetzt. Es wird lediglich Folgendes vermerkt:

Als besonders aussagefähig erwies sich das Merkmal „Arbeitsmittel“. Dabei wurden Erwerbstätige schon in mehreren Erhebungen gefragt, mit welchen Werkzeugen, Maschinen oder Anlagen sie bei ihrer Arbeit zu tun haben und welches Arbeitsmittel sie am häufigsten benutzen. Troll (2002, S. 277)

Auch projektintern werden Arbeitsmittel nicht explizit definiert. Es wird nur angemerkt, dass die Arbeitsmittel-Taxonomie iterativ anhand der empirischen Daten (ETB, Stellenanzeigen) ausgebaut wird:

Durch Hinzunahme einer Betrachtung der Arbeitsmittel in der Erwerbstätigenbefragung sowie einer stichprobenartigen Betrachtung der Arbeitsmittel in der Stellenanzeigendatenbank wurde deutlich, dass vor allem digitale Arbeitsmittel/Software bislang unbeachtet blieben. Auf Grundlage von iterativen Auswertungsschleifen der Stellenanzeigendatenbank die Arbeitsmitteltaxonomie stetig erweitert und verfeinert. Güntürk-Kuhl et al. (2017, S. 8)

Auch in der Erwerbstätigenbefragung 2018 ist die Definition eines Arbeitsmittels dem Befragten überlassen:

Was ist für Ihre Tätigkeit als <Tätigkeit aus F100-102 einblenden> das wichtigste Arbeitsmittel? Gemeint ist der Gegenstand, der dafür am wichtigsten ist. FÜR INTERVIEWER: Eine Nennung von Körperteilen ist hier nicht zulässig, es geht um Gegenstände. Bei ungenauer Angabe, wie z. B. „Werkzeug“, bitte nachfragen, ob eine genauere Nennung möglich ist.“ Hall et al. (2018)

Im Rahmen der *ersten Iteration* werden hier Arbeitsmittel aufgelistet, welche bereits am Text als solche definiert wurden:

- Werkzeuge und Maschinen
 - Hebe-/Hubmaschine
 - „Anlage“
 - 3D-Koordinatenmaschine
 - CNC-Drehmaschine
 - Ganzglasanlage
 - Proteus
- Fahrzeuge
 - Firmenfahrzeug
 - Imbiss-Wagen
 - LKW
 - Gabelstapler
- Schutzausrüstung und Arbeitskleidung
 - Sicherheitsglas
 - Stahlkappenschuhe
- Digitale Arbeitsmittel
 - Office (Excel, Word, LibreOffice, usw.)
 - Laptop, Computer, PC
 - Azure
 - Datenbanken
 - EDV
 - Auto CAD (light)
 - SAP
- Potentiell schriftliche Hilfsmittel
 - Filialkonzept
 - Rahmenvorgabe/Qualitätsvorgabe
 - Eisenbahnbetriebsnahmegernehmungsverordnung (EIGV)
 - Regelwerk
 - Gesetzestexte
 - Unfallverhütungsvorschriften
 - technische Zeichnungen

§ AM 1 Grundlegendes

§ AM 1.1 Firmenfahrzeuge sind Arbeitsmittel.

- **Beispiel 1:** „Eine der Position und Erfahrung entsprechende Vergütung, sowie ein [[Firmenfahrzeug]_{AM}]_{AB} - auch zur privaten Nutzung - sind für uns selbstverständlich.“
 - * „Firmenfahrzeug“ wird sowohl als Arbeitsbedingung als auch als Arbeitsmittel annotiert.

§ AM 1.2 Arbeitskleidung ist ein Arbeitsmittel.

§ AM 2 Output, der durch die Tätigkeit produziert wird, ist kein Arbeitsmittel

- Arbeitsmittel sind Hilfsmittel zur Erstellung von Produkten oder Dienstleistungen. Sie sind nicht der Output einer Tätigkeit. Arbeitsmittel sind zu unterscheiden von Produkten/Gütern/Outputs/ Dienstleistungen. Güter, welche bearbeitet, optimiert, verwaltet, gewartet oder repariert werden, sind keine Arbeitsmittel. Auch solche nicht, welche Teil einer Dienstleistung sind. Stellt eine Firma Hämmer her, ist der Hammer kein Arbeitsmittel. Werden dagegen mit einem Hammer Nägel in die Wand geschlagen, handelt es sich um ein Arbeitsmittel. Ein Auto ist für den Taxifahrer ein Arbeitsmittel, für den KFZ Mechaniker, welche Autos repariert, ist es aber Teil der Dienstleistung und daher kein Arbeitsmittel.
 - **Beispiel 2:** „Sie bedienen und warten [Produktionsanlagen]_{AM}“
 - * „Bedienen“ deutet auf eine Nutzung als Hilfsmittel hin. Daher wird die „Produktionsanlage“ als Arbeitsmittel annotiert, auch wenn sie im Vorgang der Wartung einer Dienstleistung (Output) entspricht.
 - **Gegenbeispiel 1:** „Informatikkaufleute beschaffen informations- und telekommunikationstechnische Systeme und verwalten sie.“
 - * Hier sagt der Text, dass die Bereitstellung und die Verwaltung von IKT-Systemen Teil der Arbeit ist. Damit ist das IKT-System Teil einer Dienstleistung und kein Hilfsmittel. Es wird nicht annotiert.
 - **Gegenbeispiel 2:** „Bearbeitung von Steuerelementen“
 - * Steuerelemente werden bearbeitet oder erstellt und sind damit kein Arbeitsmittel, da Teil eines Outputs.

§ AM 3 Inputmaterialien wie Zutaten und andere Rohstoffe sind keine Arbeitsmittel.

- Speisepläne sind Teil der Dienstleistung, Zutaten und Holz sind Rohstoffe bzw. Arbeitsstoffe, daher nicht zu annotieren.
 - **Beispiel 3:** „Sie organisieren die Arbeitsabläufe in der Küche, stellen Speisepläne auf, kaufen Zutaten ein und lagern sie fachgerecht.“
 - **Gegenbeispiel 3:** „Sie bearbeiten Holz mit [Sägemaschinen]_{AM}“

§ AM 4 Schriftliche und potentiell schriftliche Hilfsmittel wie Konzepte, Anleitungen oder Gesetzestexte sind Arbeitsmittel. Zertifikate nicht.

- Auch Standards und Konzepte sind Arbeitsmittel.
 - **Beispiel 4:** „Umsetzung des [Pflegekonzepts]_{AM} und der [Pflegestandards]_{AM}.“
- „Zeichnung“ ist ein Arbeitsmittel.
 - **Beispiel 5:** „Bau und Montage von Maschinen und Anlagen nach [Zeichnung]_{AM}“

- Zertifikate wie Führerscheine sind keine Arbeitsmittel, sie gehören nicht zu „schriftlichen Hilfsmitteln“. Sie sind Teil des Konzepts „Kompetenz“ (vgl. § K 1 .1).
 - **Gegenbeispiel 4:** „Ein [Führerschein ist dringend erforderlich]_K““

5 Arbeitsbedingungen (§ AB)

Definition

Arbeitsbedingungen beschreiben die Bedingungen unter denen potentielle Bewerber bereit sein müssen zu arbeiten, d.h. sie beschreiben den Kontext der Arbeit. Nicht damit gemeint sind jedoch Arbeitsvoraussetzungen wie z.B. der Besitz eines bestimmten Zertifikats. Diese werden unter Kompetenzen erfasst. Arbeitsbedingungen werden in *Iteration 1* rein induktiv definiert. D.h. es bleibt vorerst bei einer groben Aufzählung, welche in späteren Schritten kategorisiert werden muss. Teil von Arbeitsbedingungen sind:

- Arbeitsumfang und Arbeitszeiten
 - Schichtarbeit (z.B. Wochenendarbeit, Nachtdienst, 3-Schicht-System, usw.)
 - Flexible Arbeitszeiten (z.B. Gleitzeit, usw.)
 - Arbeitspensum (z.B. Teilzeit, 30 Stunden Woche, 50% Stelle, usw.)
 - Vertragliche Befristung (unbefristet, „sicherer Arbeitsplatz“, Daueranstellung, Elternzeitvertretung, usw.)
 - „Möglichkeit auf Übernahme“
- Attraktivität Lohn, materielle Güter
 - Lohn („attraktives/überdurchschnittliches/leistungsgerechtes Gehalt“, nach Tarif XY, usw.)
 - Urlaubs- und Weihnachtsgeld
 - Sozialleistungen, betriebliche Altersvorsorge, usw.
 - Dienstwagen, Laptop, Mobiltelefon
 - sozialversicherungspflichtig, Minijob, usw.
- Attraktivität immaterielle Güter
 - familiäres Miteinander, gutes Betriebsklima
 - flache Hierarchien
 - Hilfe beim Anlernen, Ansprechpartner haben
 - familienfreundlich
- Attraktivität Aufstieg
 - Möglichkeit zur Fortbildung/Weiterqualifizierung
 - Aufstiegschancen, „Karriere machen“, usw.
- Tätigkeitsbezogene Arbeitsbedingungen
 - abwechslungsreiche/anspruchsvolle Aufgabe/Tätigkeit, usw.
 - Montage, Außendienst, Reisebereitschaft, „Arbeit beim Kunden vor Ort“ usw.
 - Mobilität
- Bereitschaft etwas zu tun
 - Bereitschaft zur Schichtarbeit
 - Bereitschaft zur Nachtarbeit

§ AB 1 Grundlegendes

§ AB 1.1 Weiterbildungen, Fortbildungen usw. sind Arbeitsbedingungen.

- **Beispiel 1:** „Sie zeigen [Bereitschaft zur [Weiterbildung]_{AB}]_K“

§ AB 1.2 Angaben zur Arbeitszeit/Schichtarbeit sind Arbeitsbedingungen.

- **Beispiel 2:** „die [Schicht geht von 18 bis 22 Uhr, von Montags bis Freitags]_{AB}“
- **Beispiel 3:** „Wir bieten Ihnen [flexible Arbeitszeiten zwischen 6 und 21 Uhr]_{AB}“
- **Beispiel 4:** „Sie arbeiten in [Nachtschicht von 22 bis 04 Uhr]_{AB}“

§ AB 1.3 Befristung ist eine Arbeitsbedingung.

- **Beispiel 5:** „Für die Dauer einer [Schwangerschaftsvertretung]_{AB} zunächst bis Ende Februar 2019“
 - * Angaben zur genauen Befristungsdauer werden nicht annotiert. Das Signalwort genügt.

§ AB 1.4 Teamarbeit ist eine Arbeitsbedingung.

- **Beispiel 6:** „Sie [arbeiten gemeinsam im Team]_{AB}“
 - * Hierbei handelt es sich nicht um eine Tätigkeit.

§ AB 1.5 Kundenkontakt ist eine Arbeitsbedingung.

- **Beispiel 7:** „Sie sind Ansprechpartner für [Kunden]_{AB}“
 - * Hierbei handelt es sich nicht um eine Tätigkeit.

§ AB 2 Formulierungen zur allgemeinen Anforderung der Tätigkeit werden als Arbeitsbedingung annotiert.

- Falls eine Anforderung allgemein formuliert wird und nicht an eine konkrete Tätigkeit geknüpft ist, wird sie als Arbeitsbedingung annotiert.
 - **Beispiel 8:** „Sie suchen eine [anspruchsvolle und herausfordernde Aufgabe]_{AB}?“
Siehe auch „Unklarheiten“ Dokument

§ AB 3 Mobilität ist eine Arbeitsbedingung.

- Informationen zur Mobilität werden als Arbeitsbedingung annotiert, auch wenn die Information implizit formuliert ist. Mobilität bildet also eine gewisse Ausnahme zu § A 1.1.
 - **Beispiel 9:** „[Ein Fahrzeug zum Erreichen des Arbeitsplatzes ist von Vorteil]_{AB}“
 - **Beispiel 10:** „[[Regionale und/oder bundesweite Mobilität]_{AB} und [Umzugsbereitschaft]_{AB}“
 - **Beispiel 11:** „Eine der Position und Erfahrung [entsprechende Vergütung]_{AB}, sowie ein [[Firmenfahrzeug]_{AB}]_{AM} - auch zur privaten Nutzung - sind für uns selbstverständlich.“
 - **Beispiel 12:** „[Sie sind [Ansprechpartner]_T sowohl im Help-Desk als auch beim [Kunden vor Ort]_{AB}“

6 Berufsbezeichnung (§ BB)

Definition

Im Rahmen der *ersten Iteration* werden hier Formulierungen für Berufsbezeichnungen aufgelistet, welche bereits am Text als solche definiert wurden:

- Leitungskraft (m/w/d)
- Mitarbeiter/in (m/w/d)
- Projektleitung (m/w/d)
- Sachbearbeiter (m/w/d) im Fachbereich Personal
- zwei Imbissverkäufer/innen (m/w/d)
- einen Zerspanungsmechaniker- Fachrichtung Drehtechnik (m/w)
- einen examinierten Gesundheits- und Krankenpfleger (m/w/d)
- einen Arzt innere Medizin und/oder Allgemeinmedizin
- einen Bauüberwacher Fahrbahn (m/w/d)
- einen Elektro-Obermonteur/-Bauleiter
- einen Energieanlagenelektroniker (m/w/d)
- einen Fachinformatiker w/m/d Schwerpunkt Microsoft Client, Server
- einen Florist oder Gärtner (m/w/d)
- einen LKW - Fahrer (m/w/d) für den Baustelleneinsatz
- eine ständige Vertretung der Leitungskraft (m/w/d)

§ BB 1 Grundlegendes

§ BB 1.1 Berufsbezeichnungen umfassen die Anzahl der gesuchten Stellen, d.h. es wird „ein/e“ mitannotiert.

- **Beispiel 1:** „Wir suchen [eine/n Koch/Köchin]_{BB} zum nächstmöglichen Zeitpunkt“
- **Beispiel 2:** „Gesucht werden [mehrere LKW-Fahrer]_{BB}“ für unsere Standorte in München und Hamburg.

§ BB 1.2 Berufsbezeichnungen umfassen die vollständige Gender-Formulierung.

- **Beispiel 3:** „Wir suchen [eine/n Koch/Köchin (m/w/d)]_{BB} zum nächstmöglichen Zeitpunkt“

§ BB 2 Fachrichtung oder Spezialisierung werden als Teil der Berufsbezeichnung annotiert, Branchen und Sektoren alleine nicht.

- Fachrichtung, Spezialisierung oder Branche werden zusammen mit der Berufsbezeichnung annotiert, wenn diese in der Nähe von einer Berufsbezeichnung stehen. Dabei spielt es keine Rolle, ob diese frei formuliert sind oder einer formellen Qualifikation entsprechen.
 - **Beispiel 4:** „Wir suchen [einen Kraftfahrzeugmechatroniker/in – Personenkraftwagentechnik]_{BB}“
- Unten beschreibt die Spezialisierung auch eine Tätigkeit, welche Teil der Berufsbezeichnung ist:
 - **Beispiel 5:** „[Zweiradmechaniker/in für [Reparaturen und Montage von Fahrrädern]_T]_{BB}“
 - * Beides wird im jeweiligen Konzept annotiert.
- Die Spezialisierung oder Branche kann auch vor der Berufsbezeichnung stehen. Wichtig ist, dass sie an eine Berufsbezeichnung gekoppelt ist.
 - **Beispiel 6:** „Für den [Bereich Internet/Intranet suchen wir einen IT Administrator]_{BB}“
 - **Beispiel 7:** „Für unseren [Kunden in der Automobilindustrie suchen wir einen KFZ- Mechatroniker]_{BB}“
- Wenn es keinen Bezug zu einer Berufsbezeichnung gibt, annotieren wir Branchen und Spezialisierungen nicht.

- **Gegenbeispiel 1:** „Wir suchen Mitarbeiter im Bereich Tief- Rohrleitungs- und Straßenbau...“
- **Gegenbeispiel 2:** „Wir bieten Ihnen: Eine spannende, abwechslungsreiche und interessante Tätigkeit in den Bereichen Kunst und Kultur...“
- **Gegenbeispiel 3:** „Ein Unternehmen, das sich auf die Montage von Photovoltaikanlagen, Wärmepumpen und Solaranlagen spezialisiert hat, stellt...“

§ BB 3 Berufsbezeichnungen, die innerhalb einer Zertifikatsnennung auftauchen, werden als Berufsbezeichnung annotiert.

- Berufsbezeichnungen sind oft Teil von Zertifikaten. Wenn die Berufsbezeichnung auf ein Zertifikat hindeutet, wird diese auch unter Kompetenzen annotiert:
 - **Beispiel 8:** „Sie besitzen eine [abgeschlossene Berufsausbildung zum [Industriemechaniker (m/w)]_{BB}]_K“
- Zertifikate, welche keinem disjunkten Beruf zuzuordnen sind, werden nicht als Berufsbezeichnung annotiert. Beispielsweise kann ein abgeschlossenes Studium oft keinem konkreten Beruf zugeordnet werden, daher werden abgeschlossene Ausbildungen häufiger zusätzlich eine Berufsbezeichnung beinhalten als ein abgeschlossenes Studium.
 - **Beispiel 9:** „Sie besitzen eine [abgeschlossene Berufsausbildung zum [Informatiker]_{BB}]_K oder haben ein [Studium der Informatik]_K abgeschlossen“
 - * Hier ist nur „Informatiker“ eine Berufsbezeichnung, der Rest sind Zertifikate (Kompetenzen).

7 Berufserfahrung (§ BE)

Definition

Im Rahmen der *ersten Iteration* werden hier Formulierungen für Berufserfahrung aufgelistet, welche bereits am Text als solche definiert wurden:

- Mit Berufserfahrung
 - X-jährige Berufserfahrung
 - vergleichbare Berufserfahrung
- Ohne Berufserfahrung
 - erste Berufserfahrung
 - Berufsanfänger
 - Einstiegsposition
 - Quereinsteiger
 - Nachwuchstalente

§ BE 1 Berufserfahrung wird nur annotiert, wenn sie nicht an eine Kompetenz gekoppelt ist.

- Berufserfahrung wird annotiert, wenn sich die Erfahrung nicht mit einer Kompetenz oder Tätigkeit zusammen annotieren lässt bzw. wenn es um „Berufserfahrung“ im Allgemeinen geht und nicht in Bezug auf spezifische Fähigkeiten. In den folgenden Beispielen fehlt der Bezug zu einer Kompetenz und daher sind die KRen dem Konzept zuzuordnen.
 - **Beispiel 1:** „Eine [entsprechende Berufserfahrung]_{BE}, [Engagement]_K und Liebe zu Zweirädern sind beste Voraussetzungen.“
 - **Beispiel 2:** „[relevante Berufserfahrung]_{BE}...“
 - **Beispiel 3:** „Sie besitzen [Berufserfahrung]_{BE} in dem Bereich.“
 - * „in dem Bereich“ wird nicht annotiert, da kontextbedingt auch ohne explizit-machung klar, dass Berufserfahrung berufs- bzw. branchengebunden ist.
- Weitere Signalwörter für allg. Berufserfahrung sind:
 - * „Quereinsteiger“
 - * „Nachwachstalente“
- Im unteren Beispiel gibt es einen Erfahrungsbezug zu einer konkreten Kompetenz, daher wird diese KR nicht als BE annotiert.
 - **Gegenbeispiel 1:** „Sie besitzen [2 Jahre Erfahrung in der Löt Kunst]_K“