

Machbarkeitsstudie: Teilqualifikationen in Online-Job-Anzeigen (OJA)

Methodenbericht zur automatisierten Extraktion von
Teilqualifikationen für fünf Ausbildungsberufe

Projekt: Aufstieg durch Kompetenzen

Bertelsmann Stiftung

Version

1.0.4

Autoren

Johannes Müller

Erstellt am

19. November 2021

Letzte Änderung am

13. Mai 2022

Inhalt

1	Zusammenfassung	3
1.1	Extraktion von Berufen	3
1.2	Extraktion von Teilqualifikationen	3
2	Methoden	4
2.1	Klassifikation der Berufe 2010 (KldB-2010)	4
2.2	Datensammlung	6
2.3	Konzept-Definition und Annotation	7
2.4	Preprocessing und Extraktionsmodelle	8
2.5	Evaluationsmetriken	11
3	Modelle	12
3.1	KldB-2010: Extraktion	12
3.2	Teilqualifikationen: Koch/Köchin	15
3.3	Teilqualifikationen: Fachkraft im Gastgewerbe	19
3.4	Teilqualifikationen: Tiefbaufacharbeiter:in	22
3.5	Teilqualifikationen: Hochbaufacharbeiter:in	25
3.6	Teilqualifikationen: Ausbaufacharbeiter:in	29
4	Diskussion	33
	Literaturverzeichnis	34
	Tabellenverzeichnis	35

1 Zusammenfassung

Die Extraktions-Modelle der Pilotphase wurden für fünf Pilotberufe entwickelt. In der Gastronomie-Branche bilden die Zielberufe Koch/Köchin und Fachkraft (FK) im Gastgewerbe die Grundlage. Für die Bau-Branche wurden Modelle für die Zielberufe Tiefbaufacharbeiter:in, Hochbaufacharbeiter:in und Ausbaufacharbeiter:in entwickelt.

1.1 Extraktion von Berufen

Das Extraktionsmodell für Berufe nach der Klassifikation der Berufe 2010 (KldB-2010) basiert auf einem regelbasierten Entity-Recognition-Modell. Das Modell nutzt Lexikalisierungen und Suchwörter der Bundesagentur für Arbeit (BA), um Instanzen der Berufe in einem Jobtitel zu finden. Das Modell wurde auf Daten der Jobbörse der BA evaluiert und hat für die fünf Pilotberufe eine Precision von .86 und einen Recall von .61 auf KldB-3-Steller-Ebene. Auf der KldB-5-Steller-Ebene hat das Modell eine Precision von .73 und einen Recall von .52. Die Evaluationsdaten der BA ermöglichen keine vollständige Evaluation, sodass die Güte des Modells mit den Metriken eher unterschätzt wird.

1.2 Extraktion von Teilqualifikationen

Für die Extraktion von Teilqualifikationen wird ein kombiniertes Modell aus einem statistischen Named-Entity-Recognition-Modell (NER) und einem regelbasierten Entity-Recognition-Modell verwendet. Die Modelle wurden pro Pilotberuf mit ca. 1 000 annotierten Dokumenten trainiert. Für die Annotation wurde jeweils das Inter-Annotator-Agreement mithilfe von Krippendorff's Alpha berechnet. Für die Extraktionsmodelle wurden Precision, Recall, F1-Score, Spezifität sowie nominelles Alpha und Beta berechnet. Ein Überblick gibt Tabelle 1.

Tabelle 1: Überblick Evaluationsergebnisse

Zielberuf	Precision	Recall	F1	Anzahl TQ
Koch/Köchin	.72	.91	.80	7
FK Gastgewerbe	.69	.82	.75	6
Tiefbaufacharbeiter:in	.84	.82	.83	5
Hochbaufacharbeiter:in	.88	.83	.85	6
Ausbaufacharbeiter:in	.84	.75	.77	6

2 Methoden

2.1 Klassifikation der Berufe 2010 (KldB-2010)

Die Klassifikation der Berufe wurde entwickelt, um die Berufslandschaft in Deutschland zu systematisieren. Die KldB-2010 ist eine hierarchische Klassifikation mit fünf Gliederungsebenen. Dabei werden Berufe sowohl nach einer horizontalen Dimension (Berufsfachlichkeit) und einer vertikalen Dimension (Anforderungsniveau) strukturiert.

Die Berufe werden mithilfe eines fünf-stelligen Codes numerisch kodiert (Tabelle 2). Die ersten vier Ebenen gruppieren Berufe anhand ihrer Berufsfachlichkeit und die unterste Ebene beinhaltet das Anforderungsniveau.

Tabelle 2: Überblick Klassifikation der Berufe 2010

Stelle	Ebene	Anzahl der Kategorien
1	Berufsbereich	10
2	Berufshauptgruppe	37
3	Berufsgruppe	144
4	Berufsuntergruppe	700
5	Berufsgattung	1286

Dem KldB-5-Steller sind konkrete Berufsbezeichnungen und Berufsbenennungen zugeordnet.

Für die vorliegende Analyse sind vor allem Berufsgruppen (Ebene 3) und die Berufsgattung (Ebene 5) relevant. Bei der Auswahl der Berufsgattung, für welche das Teilqualifikations-Modell relevant ist, wurde dabei jeweils von einem KldB-5-Steller ausgegangen. Auf Basis des zugehörigen 3-Stellers wurden anschließend die relevanten angrenzenden 5-Steller ausgewählt, die entweder auf das Anforderungsniveau „Helfer- und Anlern Tätigkeiten“ oder „fachlich ausgerichtete Tätigkeiten“ hinweisen.

Im Folgenden werden die ausgewählten Berufe abgekürzt, um die Lesbarkeit zu erhöhen. Eine Übersicht der ausgewählten Zielberufe wird in Tabelle 3 dargestellt.

Tabelle 3: Auswahl der Berufsgruppen für die Pilot-Studie und Übersetzung zu ISCO-08

Ausgangs KldB	Bezeichnung KldB	Bezeichnung Zielberuf	Angrenzende KldB-IDs	Übersetzung zu ISCO-08
29302	Köche/Köchinnen (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten	Koch/Köchin	29301, 29382	9411, 9412
63302	Berufe im Gastronomieservice (ohne Spezialisierung) - fachlich ausgerichtete Tätigkeiten	FK Gastgewerbe	63301, 63312, 63322, 63382	5120, 5132, 5246, 5131
32222	Berufe im Straßen- und Asphaltbau - fachlich ausgerichtete Tätigkeiten	Tiefbau-facharbeiter:in	32201, 32202, 32232, 32252	7114, 7123, 8342, 8113
32122	Berufe im Maurerhandwerk - fachlich ausgerichtete Tätigkeiten	Hochbau-facharbeiter:in	32101, 32102, 32112, 32132, 32142, 32152, 32162, 32172	7111, 7112, 9312, 9313, 7125, 7214, 7119
33112	Berufe in der Fliesen-, Platten- und Mosaikverlegung - fachlich ausgerichtete Tätigkeiten	Ausbau-facharbeiter:in	33101, 33102, 33122, 33132	7115, 7121, 7124

Nach einer europaweiten Harmonisierung von Berufstaxonomien entwickelte die Bundesagentur für Arbeit einen Umsteigeschlüssel¹ für den KldB-5-Steller zu der vierstelligen internationalen ISCO-08-Taxonomie. Die ISCO-08-Taxonomie ist ein internationaler Standard, der unter anderem auch von der Firma TextKernel BV genutzt wird. Mithilfe der Übersetzung können die Teilqualifikations-Modelle auch auf Basis von ISCO-Codes genutzt werden (siehe Tabelle 3).

¹ Der Umsteigeschlüssel ist im Statistikportal der Bundesagentur für Arbeit verfügbar.

2.2 Datensammlung

Die Daten für die Annotation wurden von der Jobbörse der Bundesagentur für Arbeit gesammelt. Da der Annotationsdatensatz die Grundlage dieses Vorgangs bildet, müssen 1) die Struktur der Stellenausschreibungen analysiert werden, 2) die Verteilung von Berufen und Kompetenzen² evaluiert werden, und 3) Annahmen im Hinblick auf die Repräsentativität der Stellen für den Online-Stellenausschreibungsmarkt getestet werden.

Datengrundlage

Die Datengrundlage besteht aus Online-Stellenausschreibungen auf der Jobbörse der Bundesagentur für Arbeit. Hier wurden im Juli 2020 81 107 Ausschreibungen für alle Berufe und im November 2020 46 730 Ausschreibungen für die fünf Pilotberufe in strukturierter Form gesammelt. Beide Datensätze weisen eine hohe Datenqualität auf. Die hohe Datenqualität spiegelt sich wider in der Möglichkeit, die Zielberufe³ und die Nebenberufe eindeutig den KldB-2010-Berufen zuordnen zu können, sowie in der sauberen Formatierung von Freitexten. Letzteres ermöglicht es, Stellenanzeigen für den Annotationsdatensatz einfach und reproduzierbar zu deduplizieren und zu segmentieren.

Verteilung der Berufe und Kompetenzen

Die Analyse der Kompetenzen in den Metadaten der Stellenausschreibungen zeigt, dass der Datensatz alle Zielberufe in ausreichendem Umfang abdeckt. Dies trifft für alle fünf KldB-5-Steller, sowie die angrenzenden KldB-5-Steller unter demselben KldB-3-Steller zu. Die Stellenausschreibungen weisen weiterhin eine ausreichende Varianz von Kompetenzen auf, um eine Erkennung von unterschiedlichen Teilqualifikationen zu ermöglichen.

Struktur der Stellenausschreibungen

Mit durchschnittlich 150 Wörtern pro Stellenausschreibungen und drei bis vier zugewiesenen Kompetenzen bietet der im November erstellte Datensatz eine gute Grundlage zum Annotieren. Ein Vergleich mit einer Stichprobe aus dem TextKernel-Korpus zeigt, dass die hohe Datenqualität der Jobbörse-Daten die Voraussetzungen für transferierbare Modelle und Algorithmen erfüllt.

² Die standardisierten Kompetenzen der Bundesagentur für Arbeit dienen dabei als ein Proxy für die Teilqualifikationen.

³ Arbeitgeber:innen können bei der Freischaltung einer Stellenausschreibung einen Hauptberuf und (optional) Alternativberufe angeben. Haupt- und Nebenberufe stammen jeweils aus der Datenbank der Kennziffern (DKZ) der Bundesagentur für Arbeit. Über diese lassen sie sich dem KldB-5-Steller zuordnen.

Repräsentativität

Die Verteilung verschiedener Indikatoren (Textlänge, Berufe, Kompetenzen) wurde analysiert, um starke Stichprobenverzerrungen auszuschließen. Es wurden keine starken Verzerrungen festgestellt.

Annotationskorpus

Für die Annotation wurden jeweils ca. 1 000 Dokumente über eine Stichprobenziehung vorbereitet. Die Dokumente wurden zunächst gefiltert nach den Zielberufen (angegeben als Hauptberuf in der Stellenausschreibung) und der Länge (mindestens 10 Wörter) des Freitexts. Anschließend wurde per Zufallsziehung der Annotationskorpus erstellt.

2.3 Konzept-Definition und Annotation

Für die Annotation wurden ca. 1 000 Dokumente pro Beruf annotiert. Dabei annotieren 3 bis 10 Annotator:innen Wortsequenzen in Freitexten, die eine Teilqualifikation beschreiben. Eine Wortsequenz kam in den meisten Fällen als eine der folgenden Wortkombinationen vor:

- Substantiv + Verb (z. B. „Speisen zubereiten“)
- Verb + Substantiv (z. B. „kochen Speisen“)
- Substantiviertes Verb + Substantiv (z. B. „Zubereitung von Speisen“)
- Deskriptives Substantiv (z. B. „Speisenzubereitung“)

Definition Teilqualifikation

Die Kompetenzmodelle für die Berufe wurden von der Bertelsmann Stiftung (2018) ausführlich beschrieben. Aus der Beschreibung der Teilqualifikationen wurden die Regeln abgeleitet. Eine Teilqualifikation liegt demnach vor, wenn eine Wortsequenz ...

- ... wörtlich mit einem Teilqualifikations-Label übereinstimmt.
- ... synonym mit einem Teilqualifikations-Label verwendet werden.
- ... eine Anforderung beschreibt, die Teil einer Teilqualifikation ist.
- ... eine Gruppe von Anforderungen beschreibt, die eine Teilqualifikation repräsentieren.

- ... eine Gruppe von Anforderungen beschreibt, die eine Teilqualifikation beinhalten.

Eine Teilqualifikation liegt nicht vor, wenn eine Wortsequenz ...

- ... sich explizit auf das Unternehmen der Stellenausschreibung bezieht und nicht auf die Anforderungen an den/die Bewerber:in.
- ... sich nicht explizit oder implizit auf eine Teilqualifikation bezieht.

Annotationsprozess

Der Annotationsprozess sah dabei folgendermaßen aus:

Erläuterung der Teilqualifikationen: Mit Expert:innen werden Annotations-Handreichungen erstellt mit Regeln zur Identifizierung einer Teilqualifikation. Weiterhin werden die Inhalte der Teilqualifikationen vorgestellt.

Erste Überprüfungsrunde: Alle Annotator:innen annotieren 50 Dokumente. Mithilfe von Krippendorffs' Alpha wird anschließend die Übereinstimmung berechnet. Anschließend werden die Regeln geschärft sowie Unklarheiten gemeinsam besprochen.

(Optional) Zweite Überprüfungsrunde: Ist die Annotationsübereinstimmung noch nicht ausreichend (Krippendorffs' Alpha < 0.6) wird eine zweite Überprüfungsrunde durchgeführt. Die Annotator:innen annotieren alle 50 Dokumente. Mithilfe von Krippendorffs' Alpha wird anschließend die Übereinstimmung berechnet.

Hauptannotation: 1 000 Dokumente werden auf die Annotator:innen aufgeteilt. Dabei werden 25 Prozent der Dokumente doppelt verteilt, sodass die Inter-Annotator-Reliabilität berechnet werden kann. Um den Annotationsprozess zu vereinfachen wird eine Vor-Annotation mithilfe des regelbasierten Modells vorgenommen.

Adjudikation: Die 25 Prozent der Dokumente die von zwei Annotator:innen annotiert wurden, werden genutzt um die Inter-Annotator-Reliabilität mittels Krippendorffs' Alpha zu berechnen. Anschließend werden die Annotationskonflikte durch Expert:innen aufgelöst und der bereinigte Datensatz dem Gold-Standard hinzugefügt.

2.4 Preprocessing und Extraktionsmodelle

Für das Preprocessing wird ein trainiertes Textmodell („de_core_news_lg“) der deutschen Sprache von Explosion.ai genutzt (Honnibal & Montani 2017). Dieses übernimmt alle Pre-Processing Schritte: [Tokenisierung](#), [Lemmatisierung](#) und ein [Tok2Vec](#) Embedding.

Alle Modelle wurden mithilfe der Python Library SpaCy 3.1 trainiert. Folgende Modell-Komponenten wurden für die Pilotierung genutzt:

Regelbasiertes-Named-Entity-Recognition

Die regelbasierte Entity-Recognition sucht in einem Text nach Instanzen von bestimmten Entitäten und extrahiert die Position der Instanzen. Dabei wird kein statistisches Modell trainiert. Für die Pilotierung werden zwei unterschiedliche Versionen genutzt:

Der Phrase-Matcher (**matcher**) kann genutzt werden, wenn bereits eine große Terminologieliste mit Einzel- oder Mehrfach-Token-Phrasen vorhanden ist, deren exakte Instanzen in den Dokumenten gefunden werden sollen. Bei dem Abgleich werden Groß- und Kleinschreibung ignoriert.

Der Entity-Ruler (**ruler**) ist nicht so effizient wie der Phrase-Matcher, da er über einzelne Token-Attribute hinweg vergleicht. Der Vorteil des Entity-Rulers ist allerdings, dass gesuchte Token nicht nur über die Terminologie selbst, sondern auch über lexikalische Attribute, linguistische Merkmale, Operatoren, Mengenzugehörigkeit und umfangreiche Vergleiche gesucht werden können. So kann beispielsweise die Phrase „Zubereitung von Speisen“ als Pattern beschrieben werden: „Zubereitung ODER Zubereiten ODER Kochen“, gefolgt von ein oder zwei optionalen Token, gefolgt von „Speisen ODER Essen ODER Gerichte“. So können unterschiedliche Versionen der Phrase gefunden werden.

Statistisches Named-Entity-Recognition

Die Named-Entity-Recognition (**ner**) Komponente der SpaCy-Pipeline ist ein Deep-Learning-Modell, das auf Convolutional-Neural-Network (CNN) und einem transition-based Parsing aufbaut. Für eine bessere Extraktion können auch das statistische NER-Modell mit dem regelbasierten RULER-Modell verbunden werden. Für die Extraktion der Teilqualifikationen wird der RULER genutzt, um eine Vorannotation zu erstellen und anschließend generalisiert ein statistische NER-Modell die Extraktion. Werden RULER- und NER-Modell gemeinsam genutzt, können sowohl Precision als auch Recall erhöht werden.

Supervised Classification

Aus den annotierten Daten (auf Wortsequenzebene) kann ein Klassifikationskorpus generiert werden. Auf diesem kann ein Supervised-Classification-Modell (**textcat**) trainiert werden. Hierfür wird die Ensemble-Architektur von SpaCy verwendet, welche ein lineares „Bag-of-Words“ Modell (BoW) mit einem Convolutional-Neural-Network (CNN) verbindet.

2.5 Evaluationsmetriken

Für die Evaluation der Modelle werden unterschiedliche Metriken berechnet. Dabei werden die Teilqualifikations-Modelle sowohl auf Dokument-Ebene als auch auf Wortsequenz-Ebene evaluiert. Auf Wortsequenz-Ebene können jeweils nur Precision, Recall und F1-Score berechnet werden.

Precision

$$\frac{\textit{True Positives}}{\textit{True Positives} + \textit{False Positives}}$$

Recall (Sensitivität)

$$\frac{\textit{True Positives}}{\textit{True Positives} + \textit{False Negatives}}$$

F1-Score

$$\frac{\textit{Precision} * \textit{Recall}}{\textit{Precision} + \textit{Recall}} * 2$$

**True-Negative-Rate
(Spezifität)**

$$\frac{\textit{True Negatives}}{\textit{True Negatives} + \textit{False Positives}}$$

**False-Positive-Rate
(Nominelles Alpha)**

$$\frac{\textit{False Positives}}{\textit{False Positives} + \textit{True Negatives}}$$

**False-Negative-Rate
(Nominelles Beta)**

$$\frac{\textit{False Negatives}}{\textit{False Negatives} + \textit{True Positives}}$$

3 Modelle

Die Darstellung und Präsentation der Modelle im Folgenden orientiert sich an den von Mitchell et al. (2019) vorgeschlagenen „Model Cards for Model Reporting“. Neben der Darstellung der Evaluationsmetriken wird auch der Anwendungsbereich für die Modelle festgelegt und auf Herausforderungen bei der Anwendung hingewiesen.

3.1 KldB-2010: Extraktion

Das Berufsextraktionsmodell für die ca. 1 250 Berufe nach der Klassifikation der Berufe 2010 baut auf einem Phrase-Matcher-Modell auf. Dabei werden Phrasen aus einem Wörterbuch gegen den Freitext verglichen. Es handelt sich um ein regelbasiertes und nicht um ein statistisches Modell.

Die Wörterbücher wurden mithilfe von Suchwortlisten der Bundesagentur für Arbeit, sowie Lexikalisierungen der Berufsbezeichnungen erstellt. Da Suchwörter unterschiedlichen Berufen zugeordnet sein können, entsteht Ambiguität in der Extraktion. Diese kann a priori über ein Filtern der Suchwörter nach Zuordnungswahrscheinlichkeit eingegrenzt werden und über ein Disambiguationsverfahren nach der Extraktion teilweise aufgelöst werden.

Bestes Modell

Das beste Modell (`large_custom_1`) benutzt ausschließlich Suchwörter die maximal vier Berufen zugeordnet sind (siehe Tabelle 4 und Tabelle 5). Zusätzlich wurden 20 Prozent des Datensatzes analysiert und die Suchwörter, welche die größte Zahl falsch positiver Ergebnisse verursachen, ausgeschlossen. Die Evaluation erfolgt auf den restlichen 80 Prozent des Datensatzes.

Intendierte Nutzung

Das Modell wurde getestet auf Stellenausschreibungen, für die der mögliche KldB-Bereich auf die Zielberufe beschränkt wurde bzw. auf alle Berufe, die unter den KldB-3-Steller der Berufe fallen (z. B. „320XX“ für den Beruf Koch/Köchin). Für die Anwendung des Modells sollte deshalb eine Eingrenzung über bekannte Informationen vorgenommen werden, z. B. über den ISCO-Code. Es gibt allerdings keine Anhaltspunkte, warum das Modell bei einem ungefilterten Korpus eine schlechtere Performanz haben sollte, da die Ambiguität der Suchwörter hauptsächlich innerhalb eines KldB-3-Steller-Bereichs entsteht.

Die Extraktionsmodelle sollten in der derzeitigen Version hauptsächlich auf dem Jobtitel und nicht auf dem Freitext der Stellenausschreibung genutzt werden. Für die Extraktion im Freitext sind weitere Disambiguationsverfahren notwendig. Wenn das Modell dennoch auf dem Freitext angewendet werden soll, sollte ein Modell genutzt werden mit hoher Präzision, z. B. das Modell `large_1`. Dieses hat eine hohe Präzision, aber eine geringere Sensitivität.

Trainings- und Evaluationsdaten

Für die Evaluation der Berufsbezeichnungen wurde eine Stichprobe mit 46 730 Stellenanzeigen von der Jobbörse der Bundesagentur für Arbeit aufbereitet. Die Daten wurden dabei für den Zielbereich gefiltert. Wenn ein:e Arbeitgeber:in auf der Jobbörse eine Stelle inseriert, kann sie einen Jobtitel in einem Freitext-Feld vergeben (vorhanden für 38 860 Ausschreibungen). Weiterhin können aus der Datenbank der BA passende Berufsbezeichnungen ausgewählt werden: Einen Hauptberuf-8-Steller und mehrere Nebenberuf-8-Steller. Diese werden für diese Studie jeweils auf KldB-5-Steller- und KldB-3-Steller-Ebene aggregiert.

Grundsätzlich hat der Datensatz eine hohe Datenqualität und kann durch die indirekte Annotation durch den/die Inserent:in als Gold-Korpus für die Evaluation der Extraktion genutzt werden. Bei der Evaluation sollten aber folgende Herausforderungen berücksichtigt werden:

1. Der Jobtitel muss nicht zwangsläufig eine Berufsbezeichnung nach der KldB-2010-Taxonomie beschreiben. „Bauhelfer (m/w/d)“ kann zum Beispiel einen Helfer im Hochbau oder Tiefbau beschreiben.
2. Ein Jobtitel lässt sich nicht immer eindeutig auflösen. Wenn ein Arbeitgeber z. B. eine „Servicekraft“ sucht, kann sich dies auf eine Helfer- oder eine Fachkraftstelle beziehen. Diese Ambiguität führt dazu, dass die Präzision des Modells auf KldB-5-Steller-Ebene möglicherweise unterschätzt wird.
3. Der Jobtitel muss nicht direkt einen Hinweis auf einen Beruf enthalten (z. B. „Fachkraft gesucht“).

Die Angaben der Arbeitgeber sind außerdem nicht immer korrekt – so können Berufe zu einer falschen Berufsbezeichnung zugeordnet sein.

Evaluationsergebnisse

Ausschlaggebend für die Güte des Modells sind neben den Extraktionsmetriken (Precision, Recall, F1) auch die Ambiguität und die Konsistenz des Modells. Die Konsistenz der Erstellung des Modells und die a-priori-Ambiguität (Ontological Ambiguity) werden bei der Auswahl der Suchwörter festgelegt. Die höchste Konsistenz hat dabei ein Modell, in welchem alle Suchwörter der BA genutzt werden, ohne diese zu filtern. Die Extraktionsambiguität (Extraction Ambiguity) wird mithilfe der Evaluationsdaten bestimmt. Es wird berechnet, wie viele Suchbegriffe jeweils 1) ausschließlich True Positives (keine Ambiguität), 2) sowohl True Positives als auch False Positives und 3) mehrere False Positives identifizieren.

$$\text{Extraction Ambiguity} = \frac{\text{Entities}_{\text{ambiguous}}}{\text{Entities}_{\text{total}}}$$

Im Folgenden werden die Ergebnisse für die Evaluation auf der 3-Steller-Ebene (Tabelle 4) und 5-Steller-Ebene (Tabelle 5) der KIdB-Klassifikation präsentiert. Der erste Wert gibt jeweils die Evaluation für den von einer Arbeitgeber:in angegebenen Hauptberuf an. Der Wert in Klammern gibt den Wert für alle angegebenen Berufe (Haupt- und Alternativberufe) an.

Tabelle 4: Evaluationsergebnisse KIdB-Extraktion (KIdB-3-Steller)

Modell	P	R	F1	Specificity	Extraction Ambiguity	Ontological Ambiguity allowed	Consistency
small	.91 (.91)	.30 (.31)	.45 (.46)	1.00 (1.00)	.00	.00	High
large_1	.22 (.22)	.76 (.76)	.35 (.35)	1.00 (1.00)	.66	.00	High
large_2	.80 (.60)	.60 (.67)	.69 (.63)	1.00 (1.00)	.60	.10	High
large_custom_1	.82 (.63)	.63 (.70)	.71 (.66)	1.00 (1.00)	.44	.10	Medium
large_custom_2	.86 (.76)	.62 (.65)	.71 (.67)	1.00 (1.00)	.35	.25	Medium

Tabelle 5: Evaluationsergebnisse KIdB-Extraktion (KIdB-5-Steller)

Modell	P	R	F1	Specificity	Extraction Ambiguity	Ontological Ambiguity allowed	Consistency
small	.87 (.74)	.30 (.26)	.45 (.39)	1.00 (1.00)	.00	0	High
large_1	.11 (.08)	.68 (.55)	.20 (.14)	1.00 (1.00)	.66	0	High
large_2	.30 (.33)	.60 (.32)	.40 (.33)	1.00 (1.00)	.60	.10	High
large_custom_1	.69 (.39)	.53 (.65)	.60 (.49)	1.00 (1.00)	.44	.10	Medium
large_custom_2	.74 (.51)	.53 (.61)	.61 (.55)	1.00 (1.00)	.35	.25	Medium

3.2 Teilqualifikationen: Koch/Köchin

Das Extraktionsmodell für den Zielberuf Koch/Köchin (KldB-3-Steller: 293) besteht aus drei Komponenten: 1) Ein Pattern-Matcher (Entity-Ruler), 2) ein Named-Entity-Recognition-Modell und 3) ein Text-Klassifikationsmodell. Das Ziel der Modelle ist die Extraktion von Teilqualifikationen in deutschen Online-Stellenausschreibungen. Für den Beruf Koch/Köchin wurden sieben Teilqualifikationen definiert: „H1: Spül- und Reinigungsarbeiten durchführen“, „H2: Salate und Vorspeisen vorbereiten und herstellen“, „H3: Beilagen und Suppen vorbereiten und herstellen“, „H4: Süßspeisen herstellen“, „H5: Fleisch- und Fischgerichte herstellen“, „H6: Küchenbetrieb leiten“, „H7: Im Service arbeiten“.

Eine Besonderheit des Kompetenzmodells ist, dass es Anforderungen gibt, die mehrere Teilqualifikationen beinhalten. Die Anforderungen „Gerichte zubereiten“ oder „Speisen zubereiten“, bei denen nicht weiter spezifiziert wird, welche Gerichte zubereitet werden sollen, wurden vier Teilqualifikationen (H2, H3, H4, H5) zugeordnet. Für die Anforderung „Speisen zubereiten“ wurden deshalb für die Patterns und die Extraktion eine neue Kategorie erstellt „Kochen allgemein“. Bei der Extraktion wird diese wieder getrennt und die jeweiligen Teilqualifikationen hinzugefügt.

Für die Teilqualifikationen des/der Kochs/Köchin wurden insgesamt 43 individuelle Tokenmuster erstellt. Die Substantiv-Verb-Kombinationen sowie deskriptiven Substantive, wurden aus den Beschreibungen der Teilqualifikationen und ihrer Prozesse initialisiert und anschließend mit Annotationen aus den Überprüfungsrunden ergänzt. Anschließend wurden die Token jeweils um Synonyme und lexikalische Varianten händisch ergänzt. Das **ruler**-Modell evaluiert den tokenisierten Text gegen die lemmatisierten Token und Muster.

Die Modelle wurden zwischen dem 31.10.2020 und 31.07.2021 entwickelt und evaluiert.

Annotation

Für die Erstellung des Korpus wurden zwischen dem 01.11.2020 und 24.12.2020 insgesamt 969 Dokumente von sechs Annotator:innen annotiert. Die Annotation des Berufs war gleichzeitig auch die erste Pilotierung der Annotationsregeln, sodass die Ergebnisse der Annotationsübereinstimmung geringer sind als bei den anderen Berufen. Außerdem wurde noch keine Vor-Annotation für den Annotationsprozess genutzt. Die Patterns für den Entity-Ruler wurden für den/die Koch/Köchin mithilfe des Trainingsdatensatzes erstellt. Einen Überblick über die Annotationsergebnisse findet sich in Tabelle 6.

Tabelle 6; Annotationsmetriken für die Berufsgruppe Koch/Köchin

	Teilqualifikation	Häufigkeit (absolut)	Häufigkeit (prozentual)	Krippendorffs' Alpha
1	H1: Spül- und Reinigungsarbeiten durchführen	300	31.0%	.74
2	H2: Salate und Vorspeisen vorbereiten und herstellen	550	56.8%	.70
3	H3: Beilagen und Suppen vorbereiten und herstellen	501	51.7%	.69
4	H4: Süßspeisen herstellen	423	43.7%	.61
5	H5: Fleisch- und Fischgerichte herstellen	450	46.4%	.52
6	H6: Küchenbetrieb leiten	74	7.6%	.27
7	H7: Im Service arbeiten	73	7.5%	.14

Intendierte Nutzung

Die Extraktionsmodelle wurden auf dem Freitext von Online-Stellenausschreibungen trainiert. Dementsprechend können Teilqualifikation-Entitäten auf Wortsequenz-Ebene extrahiert werden. Das Modell sollte ausschließlich für Stellenausschreibungen für den Beruf des Kochs/Köchin (KldB-3-Steller: 293) verwendet werden. Das Modell wurde auf Ausschreibungen aus dem folgenden KldB-Bereich trainiert: 29301, 29302, 29382.

Die extrahierten Entitäten können auf Dokument-Ebene aggregiert und als Klassifikations-Modell genutzt werden. Die unausgeglichene Daten (Teilqualifikationen H6 und H7 kommen relativ selten vor) sollten bei der Analyse und Interpretation beachtet werden.

Für eine einheitliche Extraktion der Teilqualifikationen sollte das `ner_ruler`-Modell genutzt werden (siehe 4. Diskussion).

Tabelle 7: Evaluation Teilqualifikationen Koch/Köchin auf Dokumentebene ner_ruler Modell

Entity	Precision	Recall	F1	TP/FP/ FN/TN	Spezifizität	Nom. Alpha	Nom. Beta
Micro-Score	.72	.91	.80				
Macro-Score	.68	.92	.77				
H1: Spül- und Reinigungsarbeiten durchführen	.88	.95	.92	TP: 60 FP: 8 FN: 3 TN: 133	.94	.06	.05
H2: Salate und Vorspeisen vorbereiten und herstellen	.72	.84	.78	TP: 104 FP: 40 FN: 20 TN: 40	.50	.50	.16
H3: Beilagen und Suppen vorbereiten und herstellen	.73	.92	.82	TP: 97 FP: 36 FN: 8 TN: 63	.64	.36	.08
H4: Süßspeisen herstellen	.72	.93	.81	TP: 100 FP: 39 FN: 8 TN: 57	.59	.41	.07
H5: Fleisch- und Fischgerichte herstellen	.70	.93	.80	TP: 101 FP: 44 FN: 8 TN: 51	.54	.46	.07
H6: Küchenbetrieb leiten	.44	.82	.57	TP: 14 FP: 18 FN: 3 TN: 169	.90	.10	.18
H7: Im Service arbeiten	.50	1.00	.67	TP: 13 FP: 13 FN: 0 TN: 178	.93	.07	.00

Evaluationsergebnisse

Im Tabelle 7 werden die Metriken nach Entitäten mit Anzahl der True Positives (TP), False Positives (FP) und False Negatives (FN) dargestellt – beispielhaft für das `ner_ruler`-Modell. Die Evaluation wurde auf der Dokumentebene durchgeführt. Die Anforderung „Kochen allgemein“ wurde aufgelöst und die entsprechenden Entitäten hinzugefügt. Der Evaluationsdatensatz besteht aus 204 Dokumenten.

In Tabelle 8 sind die Ergebnisse der unterschiedlichen Modelle auf Wortsequenzebene dargestellt. Für die Modelle wird jeweils die Precision, der Recall und die F1-Statistik berechnet. Der erste Wert gibt jeweils das Klassifikationsergebnis (aggregiert auf das Dokument) an; der Wert in Klammern gibt die Ergebnisse für die Entitäten-Extraktion an. Die Werte sind jeweils die Micro-Scores⁴. Die Kategorie „Kochen allgemein“ wird als eigene Klasse vorhergesagt und erst im Anschluss wieder in ihre zugeordneten Klassen (in diesem Fall Teilqualifikationen) aufgetrennt. Für die Evaluation in Tabelle 7 wurde die zusätzliche Klasse aufgelöst – beispielhaft für das `ner_ruler`-Modell. Für die Evaluationen in Tabelle 8 wurden die Modelle direkt evaluiert (mit „Kochen allgemein“ als zusätzlicher Klasse).

Tabelle 8: Modell-Vergleich Teilqualifikationen Koch/Köchin auf Dokumentebene (in Klammern: Wortsequenzebene)

	Modell	Precision	Recall	F1
1	ruler	.70 (.57)	.87 (.78)	.78 (.66)
2	ner	.68 (.62)	.60 (.52)	.64 (.57)
3	ner_ruler	.65 (.53)	.86 (.79)	.74 (.63)
4	textcat	Micro: .74 Macro: .67	Micro: .75 Macro: .60	Micro: .75 Macro: .61

⁴ Der Macro-Score berechnet jeweils die Metrik unabhängig für jede Klasse und nimmt dann den Durchschnitt. Für den Micro-Score werden die Entitäten aller Klassen zusammen betrachtet, um die durchschnittliche Metrik zu berechnen.

3.3 Teilqualifikationen: Fachkraft im Gastgewerbe

Das Extraktionsmodell für den Beruf der Fachkraft im Gastgewerbe (KldB-3-Steller: 633) besteht erneut aus drei Komponenten: 1) Eine Pattern-Matcher (Entity-Ruler), 2) ein Named-Entity-Recognition-Modell und 3) ein Text-Klassifikationsmodell. Das Ziel der Modelle ist die Extraktion von Teilqualifikationen in deutschen Online-Stellenausschreibungen. Für den Beruf der Fachkraft im Gastgewerbe wurden sechs Teilqualifikationen definiert: „H1: Reinigungsarbeiten durchführen (Housekeeping)“, „H2: Veranstaltungen vorbereiten und im Warenlager arbeiten“, „H3: In der Küche helfen“, „H4: Im Service arbeiten“, „H5: An der Rezeption arbeiten“, „H6: Service leiten“.

Für die Teilqualifikationen der Fachkraft im Gastgewerbe wurden insgesamt 39 individuelle Tokenmuster erstellt. Die Substantiv-Verb-Kombinationen, sowie deskriptiven Substantiven, wurden aus den Beschreibungen der Teilqualifikationen und ihrer Prozesse initialisiert und anschließend mit Annotationen aus den Überprüfungen ergänzt. Anschließend wurden die Token jeweils um Synonyme und lexikalische Varianten händisch ergänzt. Das **ruler**-Modell evaluiert den tokenisierten Text gegen die lemmatisierten Token und Muster.

Die Modelle wurden zwischen dem 31.10.2020 und 31.07.2021 entwickelt und evaluiert.

Tabelle 9: Annotationsmetriken für die Berufsgruppe FK Gastgewerbe

	Teilqualifikation	Häufigkeit (absolut)	Häufigkeit (prozentual)	Krippendorffs' Alpha
1	H1: Reinigungsarbeiten durchführen (Housekeeping)	57	9.9%	.47
2	H2: Veranstaltungen vorbereiten und im Warenlager arbeiten	106	18.4%	.40
3	H3: In der Küche helfen	231	40.0%	.86
4	H4: Im Service arbeiten	464	80.4%	.75
5	H5: An der Rezeption arbeiten	30	5.2%	.93
6	H6: Service leiten	92	15.9%	.55

Annotation

Für die Erstellung des Korpus wurden zwischen dem 01.04.2021 und 31.05.2021 insgesamt 748 Dokumente von zwei Annotator*innen annotiert. Nach der Überprüfungsrunde wurden die Patterns für die Vor-Annotation und das Entity-Ruler-Modell erstellt.

Intendierte Nutzung

Die Extraktionsmodelle wurden auf dem Freitext von Online-Stellenausschreibungen trainiert. Dementsprechend können Teilqualifikation-Entitäten auf Wortsequenz-Ebene extrahiert werden. Weiterhin können die extrahierten Entitäten auf Dokument-Ebene aggregiert werden und als Klassifikations-Modell genutzt werden.

Das Modell wurde auf Stellenausschreibungen für den folgenden KldB-Bereich trainiert und sollte nur auf dem entsprechenden 3-Steller (633) verwendet werden: 63301, 63302, 63312, 63322, 63382.

Die unausgeglichene Daten (Teilqualifikationen 1, 2, 5 und 6 kommen relativ selten vor), sollten bei der Analyse und Interpretation beachtet werden. Nur ca. jede zehnte Stellenausschreibung für die Fachkraft im Gastgewerbe bezieht sich im Trainings- und Evaluationsdatensatz auf eine Berufsausübung im Hotel-Bereich, was zu einer möglichen Unterrepräsentation von Teilqualifikationen in diesem Bereich führen kann. Weiterhin ist die Differenzierung von Teilqualifikation 2 („Veranstaltungen vorbereiten“) nicht trennscharf zu Teilqualifikation 4 („Im Service arbeiten“) möglich, da große Teile des Service ebenfalls die Vorbereitung beinhaltet. Für die Annotation wurde aber in diesem Fall die Entscheidung getroffen, im Zweifel Teilqualifikation 4 und nicht Teilqualifikation 2 zu annotieren. Dies kann ebenfalls zu einer Unterrepräsentation von Teilqualifikation 2 in den extrahierten Daten führen.

Für eine einheitliche Extraktion der Teilqualifikationen sollte das **ner_ruler**-Modell genutzt werden (siehe 4. Diskussion).

Evaluationsergebnisse

In Tabelle 10 werden die Metriken nach Entitäten mit Anzahl der True Positives, False Positives, und False Negatives, dargestellt – beispielhaft für das **ner_ruler**-Modell. Die Evaluation wurde auf der Dokumentenebene durchgeführt. Der Evaluationsdatensatz besteht aus 152 Dokumenten.

Tabelle 10: Evaluation Teilqualifikationen FK Gastgewerbe auf Dokumentebene ner_ruler-Modell

Entity	Precision	Recall	F1	TP/FP/ FN/TN	Spezifizität	Nom. Alpha	Nom. Beta
Micro-Score	.69	.83	.75				
Macro-Score	.48	.63	.52				
H1: Reinigungsarbeiten durchführen (Housekeeping)	.34	1.00	.51	TP: 10 FP: 19 FN: 0 TN: 123	.87	.13	.00
H2: Veranstaltungen vorbereiten und im Warenlager arbeiten	.43	.38	.40	TP: 6 FP: 8 FN: 10 TN: 128	.94	.06	.62
H3: In der Küche helfen	.77	.68	.72	TP: 97 FP: 36 FN: 8 TN: 11	.23	.77	.08
H4: Im Service arbeiten	.84	.97	.90	TP: 106 FP: 20 FN: 3 TN: 23	.53	.47	.03
H5: An der Rezeption arbeiten	.50	.67	.57	TP: 4 FP: 4 FN: 2 TN: 142	.97	.03	.33
H6: Service leiten	.47	.70	.56	TP: 14 FP: 18 FN: 3 TN: 117	.87	.13	.18

Im Tabelle 11 sind die Ergebnisse der unterschiedlichen Modelle auf Dokument- und Wortsequenzebene dargestellt. Für die Modelle wird jeweils die Precision, der Recall, die F1-Statistik berechnet. Der erste Wert gibt jeweils das Klassifikationsergebnis (aggregiert auf das Dokument) an. Der Wert in Klammern gibt die Ergebnisse für die Entitäten-Extraktion (Wortsequenzebene) an.

Tabelle 11: Evaluation Teilqualifikationen FK Gastgewerbe auf Dokumentenebene (in Klammern: Wortsequenzebene)

	Modell	Precision	Recall	F1
1	ruler	.75 (.68)	.76 (.49)	.76 (.57)
2	ner	.76 (.70)	.72 (.54)	.74 (.61)
3	ner_ruler	.69 (.62)	.83 (.61)	.75 (.61)
4	textcat	Micro: .78 Macro: .26	Micro: .65 Macro: .28	Micro: .71 Macro: .27

3.4 Teilqualifikationen: Tiefbauarbeiter:in

Das Extraktionsmodell für den Beruf des Tiefbauarbeiter:in (KldB-3-Steller: 322) besteht aus drei Komponenten: 1) Eine Pattern-Matcher (Entity-Ruler), 2) ein Named-Entity-Recognition-Modell und 3) ein Text-Klassifikationsmodell. Das Ziel der Modelle ist die Extraktion von Teilqualifikationen in deutschen Online-Stellenausschreibungen. Für den Beruf der Tiefbauarbeiter:in wurden fünf Teilqualifikationen definiert: „H1: Vermessungs- und Erdbauarbeiten durchführen“, „H2: Schächte setzen, Rohrleitungen verlegen und instand setzen“, „H3: Pflasterflächen herstellen“, „H4: Verkehrsflächen instand setzen“, „H5: Einfache Hochbauarbeiten durchführen“.

Für die Teilqualifikationen der/des Tiefbauarbeiterin/Tiefbauarbeiters wurden insgesamt 27 individuelle Tokenmuster erstellt. Die Substantiv-Verb-Kombinationen sowie deskriptiven Substantiven, wurden aus den Beschreibungen der Teilqualifikationen und ihrer Prozesse initialisiert und anschließend mit Annotationen aus den Überprüfungsrounden ergänzt. Anschließend wurden die Token jeweils um Synonyme und lexikalische Varianten händisch ergänzt. Das **ruler**-Modell evaluiert den tokenisierten Text gegen die lemmatisierten Token und Muster.

Die Modelle wurden zwischen dem 31.03.2021 und 31.05.2021 entwickelt und evaluiert.

Annotation

Für die Erstellung des Korpus wurden zwischen dem 01.02.2021 und 31.03.2021 insgesamt 1 041 Dokumente von acht Annotator:innen annotiert. Nach der Überprüfungsrounde wurden die Patterns für die Vor-Annotation und das Entity-Ruler-Modell erstellt.

Tabelle 12: Annotationsmetriken für die Berufsgruppe Tiefbaufacharbeiter:in

	Teilqualifikation	Häufigkeit (absolut)	Häufigkeit (prozentual)	Krippendorffs' Alpha
1	H1: Vermessungs- und Erdbauarbeiten durchführen	372	35.7%	.68
2	H2: Schächte setzen, Rohrleitungen verlegen und instand setzen	474	45.5%	.73
3	H3: Pflasterflächen herstellen	391	37.6%	.84
4	H4: Verkehrsflächen instand setzen	341	32.8%	.77
5	H5: Einfache Hochbauarbeiten durchführen	76	7.3%	.36

Intendierte Nutzung

Die Extraktionsmodelle wurden auf dem Freitext von Online-Stellenausschreibungen trainiert. Dementsprechend können Teilqualifikation-Entitäten auf Wortsequenz-Ebene extrahiert werden. Weiterhin können die extrahierten Entitäten auf Dokument-Ebene aggregiert werden und als Klassifikations-Modell genutzt werden.

Das Modell wurde auf Stellenausschreibungen für den folgenden KldB-Bereich trainiert und sollte nur auf dem entsprechenden 3-Steller (322) verwendet werden: 32201, 32202, 32222, 32232, 32252, 32293.

Die entitätsspezifischen Extraktionsergebnisse sind mit F1-Scores von .77 bis .95 sehr gut. Ausschließlich Teilqualifikation 5 („Einfache Hochbauarbeiten“) weist eine große Ambiguität auf, weshalb die Ergebnisse für diese Teilqualifikation mit Vorsicht analysiert und interpretiert werden sollten.

Für eine einheitliche Extraktion der Teilqualifikationen sollte das `ner_ruler`-Modell genutzt werden (siehe 4. Diskussion).

Evaluationsergebnisse

In Tabelle 13 werden die Metriken nach Entitäten mit Anzahl der True Positives, False Positives und False Negatives dargestellt – beispielhaft für das `ner_ruler`-Modell. Der Evaluationsdatensatz besteht aus 206 Dokumenten.

Tabelle 13: Evaluation Teilqualifikationen Tiefbaufacharbeiter:in auf Dokumentebene `ner_ruler`-Modell

Entity	Precision	Recall	F1	TP/FP/ FN/TN	Spezifizität	Nom. Alpha	Nom. Beta
Micro-Score	.84	.82	.83				
Macro-Score	.66	.63	.64				
H1: Vermessungs- und Erdbauarbeiten durchführen	.80	.78	.79	TP: 56 FP: 14 FN: 16 TN: 120	.90	.10	.22
H2: Schächte setzen, Rohrleitungen verlegen und instand setzen	.84	.83	.83	TP: 67 FP: 13 FN: 14 TN: 112	.90	.10	.17
H3: Pflasterflächen herstellen	.98	.92	.95	TP: 79 FP: 2 FN: 7 TN: 118	.98	.02	.08
H4: Verkehrsflächen instand setzen	.75	.80	.77	TP: 53 FP: 18 FN: 13 TN: 122	.87	.13	.20
H5: Einfache Hochbauarbeiten durchführen	.50	.67	.57	TP: 6 FP: 4 FN: 8 TN: 188	.98	.02	.57

In Tabelle 14 sind die Ergebnisse der unterschiedlichen Modelle dargestellt. Für die Modelle wird jeweils die Precision, der Recall, die F1-Statistik berechnet. Der erste Wert gibt jeweils das Klassifikationsergebnis (aggregiert auf das Dokument) an; der Wert in Klammern gibt die Ergebnisse für die Entitäten-Extraktion an. Die Werte sind jeweils die Micro-Scores.

Tabelle 14: Evaluation Teilqualifikationen Tiefbaufacharbeiter:in auf Dokumentebene (in Klammern: Wortsequenzebene)

	Modell	Precision	Recall	F1
1	ruler	.89 (.80)	.71 (.61)	.79 (.69)
2	ner	.86 (.72)	.79 (.68)	.82 (.70)
3	ner_ruler	.84 (.70)	.82 (.73)	.83 (.71)
4	textcat	Micro: .81 Macro: .70	Micro: .69 Macro: .58	Micro: .75 Macro: .63

3.5 Teilqualifikationen: Hochbaufacharbeiter:in

Das Extraktionsmodell für den Beruf des/der Hochbaufacharbeiter:in (KldB-3-Steller: 321) besteht aus drei Komponenten: 1) Eine Pattern-Matcher (Entity-Ruler), 2) ein Named-Entity-Recognition-Modell und 3) ein Text-Klassifikationsmodell. Das Ziel der Modelle ist die Extraktion von Teilqualifikationen in deutschen Online-Stellenausschreibungen. Für den Beruf Hochbaufacharbeiter:in wurden sechs Teilqualifikationen definiert: "H1: Einfache einschalige Baukörper erstellen", "H2: Bauwerke erschließen", "H3: Baukörper mit tragenden Funktionen herstellen", "H4: Baukörper abdichten und dämmen", "H5: Einfache Ausbauarbeiten durchführen", "H6: Besondere Bauteile und Sichtmauerwerk herstellen".

Eine Besonderheit des Kompetenzmodells ist, dass es Anforderungen gibt, die mehrere Teilqualifikationen beinhalten. Die Anforderung „mauern“ oder „Mauern erstellen“, welche nicht weiter spezifiziert wird, gehört zu zwei Teilqualifikationen (H1 und H3). Für die Anforderung „Mauern herstellen“, wurde deshalb für die Patterns und die Extraktion eine neue Kategorie erstellt „Mauern allgemein“. Bei der Extraktion wird diese wieder getrennt und die jeweiligen Teilqualifikationen hinzugefügt.

Für die Teilqualifikationen der/des Hochbaufacharbeiterin/Hochbaufacharbeiters wurden insgesamt 33 individuelle Tokenmuster erstellt. Die Substantiv-Verb-Kombinationen sowie deskriptiven Substantiven, wurden aus den Beschreibungen der Teilqualifikationen und ihrer Prozesse initialisiert und anschließend mit Annotationen aus den Überprüfungsrounden ergänzt. Anschließend wurden die Token jeweils um Synonyme und lexikalische Varianten händisch ergänzt. Das ruler-Modell evaluiert den tokenisierten Text gegen die lemmatisierten Token und Muster.

Die Modelle wurden zwischen dem 01.06.2021 und 31.06.2021 entwickelt und evaluiert

Annotation

Für die Erstellung des Korpus wurden zwischen dem 01.04.2021 und 31.05.2021 insgesamt 965 Dokumente von drei Annotator:innen annotiert. Nach der Überprüfungsrunde wurden die Patterns für die Vor-Annotation und das Entity-Ruler-Modell erstellt.

Tabelle 15: Annotationsmetriken für die Berufsgruppe Hochbaufacharbeiter:in

	Teilqualifikation	Häufigkeit (absolut)	Häufigkeit (prozentual)	Krippendorffs' Alpha
1	H1: Einfache einschalige Baukörper erstellen	225	22.3%	.82
2	H2: Bauwerke erschließen	215	22.3%	.83
3	H3: Baukörper mit tragenden Funktionen herstellen	207	21.5%	.92
4	H4: Baukörper abdichten und dämmen	80	8.3%	.89
5	H5: Einfache Ausbauarbeiten durchführen	201	20.8%	.86
6	H6: Besondere Bauteile und Sichtmauerwerk herstellen	81	8.4%	.70

Intendierte Nutzung

Die Extraktionsmodelle wurden auf dem Freitext von Online-Stellenausschreibungen trainiert. Dementsprechend können Teilqualifikation-Entitäten auf Wortsequenz-Ebene extrahiert werden. Weiterhin können die extrahierten Entitäten auf Dokument-Ebene aggregiert werden und als Klassifikations-Modell genutzt werden.

Das Modell wurde auf Stellenausschreibungen für den folgenden KldB-Bereich trainiert und sollte nur auf dem entsprechenden 3-Steller (321) verwendet werden: 32101, 32102, 32112, 32122, 32123, 32132, 32142, 32152, 32162, 32172.

Für eine einheitliche Extraktion der Teilqualifikationen sollte das **ner_ruler**-Modell genutzt werden (siehe 4. Diskussion).

Evaluationsergebnisse

Tabelle 16: Evaluation Teilqualifikationen Hochbaufacharbeiter:in auf Dokumentenebene ner_ruler-Modell

Entity	Precision	Recall	F1	TP/FP/ FN/TN	Spezifizität	Nom. Alpha	Nom. Beta
Micro-Score	.88	.83	.85				
Macro-Score	.86	.82	.84				
H1: Einfache einschalige Baukörper erstellen	.90	.88	.89	TP: 36 FP: 4 FN: 5 TN: 97	.96	.04	.12
H2: Bauwerke erschließen	.80	.80	.80	TP: 32 FP: 8 FN: 8 TN: 94	.92	.08	.20
H3: Baukörper mit tragenden Funktionen herstellen	.92	.85	.88	TP: 33 FP: 3 FN: 6 TN: 100	.97	.03	.15
H4: Baukörper abdichten und dämmen	.70	.78	.74	TP: 7 FP: 3 FN: 2 TN: 130	.98	.02	.22
H5: Einfache Ausbauarbeiten durchführen	.91	.73	.81	TP: 29 FP: 3 FN: 11 TN: 99	.97	.03	.28
H6: Besondere Bauteile und Sichtmauerwerk herstellen	.86	.80	.83	TP: 12 FP: 2 FN: 3 TN: 125	.98	.02	.20

In Tabelle 16 werden die Metriken nach Entitäten mit Anzahl der True Positives, False Positives und False Negatives dargestellt – beispielhaft für das `ner_ruler`-Modell. Für die Evaluation wurde auf der Dokumentenebene die kombinierte Kategorie "Mauerarbeiten allgemein" aufgelöst und die entsprechenden Entitäten für die einzelnen Teilqualifikationen hinzugefügt. Der Evaluationsdatensatz besteht aus 142 Dokumenten.

In Tabelle 17 sind die Ergebnisse der unterschiedlichen Modelle dargestellt. Für die Modelle wird jeweils die Precision, der Recall, die F1-Statistik berechnet. Der erste Wert gibt jeweils das Klassifikationsergebnis (aggregiert auf das Dokument) an. Der Wert in Klammern gibt die Ergebnisse für die Entitäten-Extraktion auf Wortsequenzebene an. Die Werte sind jeweils die Micro-Scores. Abweichungen zu Tabelle 16 sind dadurch zu erklären, dass die Kategorie „Mauerarbeiten allgemein“ nicht aufgelöst wurde.

Tabelle 17: Evaluation Teilqualifikationen Hochbaufacharbeiter:in auf Dokumentenebene (in Klammern: Wortsequenzebene)

	Modell	Precision	Recall	F1
1	ruler	.98 (.95)	.67 (.56)	.80 (.70)
2	ner	.82 (.77)	.73 (.62)	.77 (.69)
3	ner_ruler	.85 (.81)	.79 (.70)	.82 (.75)
4	textcat	Micro: .85 Macro: .82	Micro: .69 Macro: .66	Micro: .76 Macro: .73

3.6 Teilqualifikationen: Ausbaufacharbeiter:in

Das Extraktionsmodell für den Beruf des/der Ausbaufacharbeiters/Ausbaufacharbeiterin (KldB-3-Steller: 331) besteht aus drei Komponenten: 1) Ein Pattern-Matcher (Entity-Ruler), 2) ein Named-Entity-Recognition-Modell und 3) ein Text-Klassifikationsmodell. Das Ziel der Modelle ist die Extraktion von sogenannten Teilqualifikationen in deutschen Online-Stellenausschreibungen. Für den Beruf Ausbaufacharbeiter:in wurden sechs Teilqualifikationen definiert: „H1: Wandfliesen im Dünnbett verlegen“, „H2: Bodenfliesen im Dünnbett verlegen“, „H3: Zubehör und zusätzliche Materialien einbauen“, „H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen“, „H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen“, „H6: Einfache Hochbauarbeiten durchführen“.

Eine Besonderheit des Kompetenzmodells ist, dass es Anforderungen gibt, die mehrere Teilqualifikationen beinhalten. Die Anforderung „Fliesen legen“ oder „Fliesenlegen“, welche nicht weiter spezifiziert, welche Fliesen verlegt werden, gehört zu vier Teilqualifikationen (1, 2, 3 und 4). Für die Anforderung „Fliesenlegen“, wurde deshalb für die Patterns und die Extraktion eine neue Kategorie erstellt „Fliesenlegen Allgemein“. Bei der Extraktion wird diese wieder getrennt und die zugehörigen Teilqualifikationen als Entitäten hinzugefügt.

Für die Teilqualifikationen der/des Ausbaufacharbeiterin/Ausbaufacharbeiters wurden insgesamt 33 individuelle Tokenmuster erstellt. Die Substantiv-Verb-Kombinationen sowie deskriptiven Substantiven, wurden aus den Beschreibungen der Teilqualifikationen und ihrer Prozesse initialisiert und anschließend mit Annotationen aus den Überprüfungsrounden ergänzt. Anschließend wurden die Token jeweils um Synonyme und lexikalische Varianten händisch ergänzt. Das **ruler**-Modell evaluiert den tokenisierten Text gegen die lemmatisierten Token und Muster.

Die Modelle wurden zwischen dem 01.06.2021 und 31.06.2021 entwickelt und evaluiert.

Annotation

Für die Erstellung des Korpus wurden zwischen dem 01.04.2021 und 31.05.2021 insgesamt 894 Dokumente von vier Annotator:innen annotiert. Nach der Überprüfungsrounde wurden die Patterns für die Vor-Annotation und das Entity-Ruler-Modell erstellt.

Tabelle 18: Annotationsmetriken für die Berufsgruppe Ausbaufacharbeiter:in

	Teilqualifikation	Häufigkeit (absolut)	Häufigkeit (prozentual)	Krippendorffs' Alpha
1	H1: Wandfliesen im Dünnbett verlegen	290	32.4%	.70
2	H2: Bodenfliesen im Dünnbett verlegen	304	34.0%	.67
3	H3: Zubehör und zusätzliche Materialien einbauen	281	31.4%	.62
4	H4: Spezielle Fliesen- und Plattenverlege- arbeiten durchführen	264	29.5%	.71
5	H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen	193	21.6%	.80
6	H6: Einfache Hochbauarbeiten durchführen	26	2.9%	.55

Intendierte Nutzung

Die Extraktionsmodellen wurden auf dem Freitext von Online-Stellenausschreibungen trainiert. Dementsprechend können Teilqualifikation-Entitäten auf Wortsequenz-Ebene extrahiert werden. Weiterhin können die extrahierten Entitäten auf Dokument-Ebene aggregiert werden und als Klassifikations-Modell genutzt werden.

Das Modell wurde auf Stellenausschreibungen für den folgenden KldB-Bereich trainiert und sollte nur auf dem entsprechenden 3-Steller (331) verwendet werden: 33101, 33102, 33112,33122, 33132, 33133, 33193.

Für eine einheitliche Extraktion der Teilqualifikationen sollte das **ner_ruler**-Modell genutzt werden (siehe 4. Diskussion).

Evaluationsergebnisse

In Tabelle 19 werden die Metriken nach Entitäten mit Anzahl der True Positives, False Positives und False Negatives dargestellt - beispielhaft für das **ner_ruler**-Modell. Für die Evaluation wurde auf der Dokumentenebene die kombinierte Kategorie „Fliesenlegen allgemein“ aufgelöst und die entsprechenden Entitäten für die einzelnen Teilqualifikationen hinzugefügt. Der Evaluationsdatensatz besteht aus 113 Dokumenten.

Tabelle 19: Evaluation Teilqualifikationen Ausbaufacharbeiter:in auf Dokumentenebene ner_ruler Modell

Entity	Precision	Recall	F1	TP/FP/ FN/TN	Spezifizität	Nom. Alpha	Nom. Beta
Micro-Score	.84	.75	.80				
Macro-Score	.81	.73	.77				
H1: Wandfliesen im Dünnbett verlegen	.88	.78	.83	TP: 45 FP: 6 FN: 13 TN: 49	.89	.11	.22
H2: Bodenfliesen im Dünnbett verlegen	.79	.77	.78	TP: 48 FP: 13 FN: 14 TN: 38	.75	.25	.23
H3: Zubehör und zusätzliche Materialien einbauen	.81	.75	.78	TP: 50 FP: 12 FN: 17 TN: 34	.74	.26	.25
H4: Spezielle Fliesen- und Plattenverlegearbeiten durchführen	.86	.70	.77	TP: 48 FP: 8 FN: 21 TN: 36	.82	.18	.36
H5: Dämm-, Trocken-, Putz- und Estricharbeiten ausführen	.85	.83	.84	TP: 29 FP: 5 FN: 6 TN: 73	.94	.06	.17
H6: Einfache Hochbauarbeiten durchführen	.60	.50	.55	TP: 3 FP: 2 FN: 3 TN: 105	.98	.02	.50

In Tabelle 20 sind die Ergebnisse der unterschiedlichen Modelle dargestellt. Für die Modelle wird jeweils die Precision, der Recall, die F1-Statistik berechnet. Der erste Wert gibt jeweils das Klassifikationsergebnis (aggregiert auf das Dokument) an. Der Wert in Klammern gibt die Ergebnisse für die Entitäten-Extraktion an. Die Werte sind jeweils die Micro-Scores. Abweichungen zu Tabelle 19 sind dadurch zu erklären, dass die Kategorie „Fliesenlegen allgemein“ nicht aufgelöst wurde.

Tabelle 20: Evaluation Teilqualifikationen Ausbaufacharbeiter:in auf Dokumentebene (in Klammern: Wortsequenzebene)

	Modell	Precision	Recall	F1
1	ruler	.84 (.80)	.46 (.37)	.60 (.51)
2	ner	.75 (.65)	.67 (.56)	.71 (.60)
3	ner_ruler	.74 (.66)	.72 (.62)	.73 (.64)
4	textcat	Micro: .82 Macro: .68	Micro: .76 Macro: .65	Micro: .79 Macro: .66

4 Diskussion

Die vorliegende Analyse evaluierte die Möglichkeiten und Grenzen der Extraktion von Teilqualifikationen aus Online-Stellenausschreibungen.

Eine Voraussetzung für die gezielte Extraktion von Teilqualifikationen ist zunächst die Einschränkung auf Zielberufe, für welche die Teilqualifikationen entwickelt wurden. Das Modell für die zugrundeliegenden Klassen der KldB-2010 wurde für die Zielberufe optimiert. Für die Entwicklung eines Modells, welches alle 1286 KldB-5-Steller klassifizieren kann, ist ein Ansatz notwendig, bei welchem das Modell global optimiert werden kann. Die Erstellung eines Goldstandards für die Evaluation von KldB-Extraktions-Modellen bleibt dabei aber auch in Zukunft eine Herausforderung: Zum einen sind Angaben von Arbeitgeber:innen oft ungenau und zum anderen ist eine eindeutige Zuordnung zu einer Klasse der KldB-2010 nicht immer möglich.

Die Extraktion der Teilqualifikationen funktioniert grundsätzlich sehr gut über einen hybriden Ansatz, bei dem ein regelbasierter Ansatz um ein statistisches Modell ergänzt wird. Regelbasierte Modelle weisen zunächst eine hohe Precision auf, da Wortsequenz-Muster zielgenau formuliert werden können. Die Schwäche von regelbasierten Modellen ist, dass sie nicht über die vordefinierten Muster hinaus generalisieren können. Diese Schwäche kann aber durch das statistische Modell teilweise ausgeglichen werden, welches über die semantische Struktur der Trainingsbeispiele nicht gesehene Wortsequenzen erkennen kann. Um eine einheitliche Extraktion der Teilqualifikationen über die Zielberufe hinweg zu gewährleisten, sollte deshalb jeweils das `ner_ruler`-Modell verwendet werden.

Die Analyse zeigt, dass die Definition der Teilqualifikationen – und ihre Nähe zur tatsächlichen Tätigkeit in dem jeweiligen Beruf – einen großen Einfluss auf die Güte der Modelle haben. Dies zeigt sich insbesondere bei dem Modell für die FK Gastgewerbe: Stellen für die FK Gastgewerbe können in unterschiedlichen Bereichen ausgeschrieben werden, die jeweils einen sehr unterschiedlichen Kontext und Tätigkeiten aufweisen. Um diese Herausforderung zu mitigieren, müsste die Anzahl der Dokumente aus unterschiedlichen Kontexten gewichtet werden (im Fall der FK Gastgewerbe z. B. Hotelrezeption, Hotelreinigung, Küchenhilfe, Restaurant-Service) oder es müssten insgesamt deutlich mehr Dokumente annotiert werden.

Die Evaluation der Modelle zeigt, dass Teilqualifikationen in Stellenausschreibungen benannt sind und ihre Extraktion mit guten bis sehr guten Ergebnissen möglich ist. Über die Verbindung mit der KldB-Extraktion können die Modelle auf allen Stellenausschreibungen angewandt werden, die einen Job-Titel und einen Freitext aufweisen.

Literaturverzeichnis

Bertelsmann Stiftung (2018): „Projektinformationen zu MYSKILLS – Berufliche Kompetenzen erkennen.“

Feilmayr, Christina, and Wolfram WöB. (2016): „An analysis of ontologies and their success factors for application to business.“ Data & Knowledge Engineering 101: 1-23.

Hazem, A. and Daille, B. (2018): „Word embedding approach for synonym extraction of multi-word terms.“ In Proceedings of the Eleventh International Conference on Language Resources and Evaluation.

Honnibal M, Montani I. (2017): „spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.“

Maynard, Diana, Yaoyong Li, and Peters, W. (2008): „NLP Techniques for Term Extraction and Ontology Population.“. 107-127.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... and Gebru, T. (2019): „Model cards for model reporting“. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229).

Wang, Bin, Angela Wang, Fenxiao Chen, Yuncheng Wang, and C-C. Jay Kuo. (2019): „Evaluating word embedding models: methods and experimental results.“ APSIPA transactions on signal and information processing 8.

Tabellenverzeichnis

Tabelle 1: Überblick Evaluationsergebnisse	3
Tabelle 2: Überblick Klassifikation der Berufe 2010	4
Tabelle 3: Auswahl der Berufsgruppen für die Pilot-Studie und Übersetzung zu ISCO-08	5
Tabelle 4: Evaluationsergebnisse KldB-Extraktion (KldB-3-Steller)	14
Tabelle 5: Evaluationsergebnisse KldB-Extraktion (KldB-5-Steller).....	14
Tabelle 6; Annotationsmetriken für die Berufsgruppe Koch/Köchin.....	16
Tabelle 7: Evaluation Teilqualifikationen Koch/Köchin auf Dokumentebene ner_ruler Modell.	17
Tabelle 8: Modell-Vergleich Teilqualifikationen Koch/Köchin auf Dokumentebene (in Klammern: Wortsequenzebene).....	18
Tabelle 9: Annotationsmetriken für die Berufsgruppe FK Gastgewerbe	19
Tabelle 10: Evaluation Teilqualifikationen FK Gastgewerbe auf Dokumentebene ner_ruler-Modell	21
Tabelle 11: Evaluation Teilqualifikationen FK Gastgewerbe auf Dokumentebene (in Klammern: Wortsequenzebene)	22
Tabelle 12: Annotationsmetriken für die Berufsgruppe Tiefbaufacharbeiter:in	23
Tabelle 13: Evaluation Teilqualifikationen Tiefbaufacharbeiter:in auf Dokumentebene ner_ruler-Modell	24
Tabelle 14: Evaluation Teilqualifikationen Tiefbaufacharbeiter:in auf Dokumentebene (in Klammern: Wortsequenzebene).....	25
Tabelle 15: Annotationsmetriken für die Berufsgruppe Hochbaufacharbeiter:in	26
Tabelle 16: Evaluation Teilqualifikationen Hochbaufacharbeiter:in auf Dokumentebene ner_ruler-Modell	27
Tabelle 17: Evaluation Teilqualifikationen Hochbaufacharbeiter:in auf Dokumentebene (in Klammern: Wortsequenzebene).....	28
Tabelle 18: Annotationsmetriken für die Berufsgruppe Ausbaufacharbeiter:in.....	30
Tabelle 19: Evaluation Teilqualifikationen Ausbaufacharbeiter:in auf Dokumentebene ner_ruler Modell	31
Tabelle 20: Evaluation Teilqualifikationen Ausbaufacharbeiter:in auf Dokumentebene (in Klammern: Wortsequenzebene).....	32