

Stellenanzeigen-Analyse in der sozialwissenschaftlichen Forschung

Johanna Binnewitt

Bundesinstitut für Berufsbildung

Übung „Sprachverarbeitung“, 03.06.2025

- Johanna Binnewitt
- Wissenschaftliche Mitarbeiterin am Bundesinstitut für Berufsbildung in Bonn
- Master Informationsverarbeitung, Uni Köln
- Promotionsstudentin am Institut für Digital Humanities (IDH), Uni Köln

- Ressortforschungseinrichtung des Bundes
- über 800 Mitarbeitende, größtenteils aus Sozial- und Wirtschaftswissenschaften, Wirtschaftspädagogik und Psychologie
- Erforschung und Weiterentwicklung von beruflicher Aus- und Weiterbildung
 - Anpassung der 328 anerkannten Ausbildungsberufe
 - weitere Themen: Berufsorientierung, Qualifizierungsbedarfe, Fachkräftesicherung, ...
- empirische quantitative Sozialforschung arbeitet üblicherweise mit Befragungsdaten

BIBB/BAuA-Erwerbstätigenbefragung (ETB)

- Befragung von Erwerbstätigen
- ca. alle 6 Jahre
- aktuellste Befragung: 2024 (ca. 20.000 Erwerbstätige)
- Themen u.a.: Arbeitsanforderungen, Arbeitsaufgaben und Arbeitsbedingungen

BIBB/BAuA-Erwerbstätigenbefragung (ETB)

- Befragung von Erwerbstätigen
- ca. alle 6 Jahre
- aktuellste Befragung: 2024 (ca. 20.000 Erwerbstätige)
- Themen u.a.: Arbeitsanforderungen, Arbeitsaufgaben und Arbeitsbedingungen

BIBB-Betriebspanel zu Qualifizierung und Kompetenzentwicklung

- Betriebsbefragung von mind. 3500 Betrieben in Deutschland
- seit 2011 jährlich durchgeführt
- Themen u.a.: betriebliche Nachfrage nach Fach- und Nachwuchskräften sowie Aus- und Weiterbildungsgeschehen

- Wie verändern sich Arbeitsanforderungen durch den **technologischen Wandel**?

- Wie verändern sich Arbeitsanforderungen durch den **technologischen Wandel**?
- Welche Qualifikationen werden für die **sozial-ökologische Transformation** benötigt?

- Wie verändern sich Arbeitsanforderungen durch den **technologischen Wandel**?
- Welche Qualifikationen werden für die **sozial-ökologische Transformation** benötigt?
- Welche Berufsgruppen fallen durch den **demographischen Wandel** weg und wie wirkt sich das auf den Arbeitsmarkt aus?

- Wie verändern sich Arbeitsanforderungen durch den **technologischen Wandel**?
- Welche Qualifikationen werden für die **sozial-ökologische Transformation** benötigt?
- Welche Berufsgruppen fallen durch den **demographischen Wandel** weg und wie wirkt sich das auf den Arbeitsmarkt aus?
- Wie beeinflusst Rekrutierungsverhalten von Unternehmen **Gender-stereotype Besetzung** von Stellen?

limitierte Aktualität

- Befragungen sind zeitlich begrenzt
- Feldphasen benötigen Vor- und Nachbereitungszeit

limitierte Aktualität

- Befragungen sind zeitlich begrenzt
- Feldphasen benötigen Vor- und Nachbereitungszeit

steigende Kosten

- Personalkosten bei Befragungsinstituten
- Personalkosten bei Betrieben

limitierte Aktualität

- Befragungen sind zeitlich begrenzt
- Feldphasen benötigen Vor- und Nachbereitungszeit

steigende Kosten

- Personalkosten bei Befragungsinstituten
- Personalkosten bei Betrieben

sinkende Rücklaufquoten

- immer weniger Personen nehmen an Studien teil

limitierte Aktualität

- Befragungen sind zeitlich begrenzt
- Feldphasen benötigen Vor- und Nachbereitungszeit

steigende Kosten

- Personalkosten bei Befragungsinstituten
- Personalkosten bei Betrieben

sinkende Rücklaufquoten

- immer weniger Personen nehmen an Studien teil

geringe Granularität

- Abfrage von Arbeitsaufgaben oder Kompetenzen häufig auf grobe Kategorien beschränkt

- Stellenanzeigen
- Curricula (Modulhandbücher, Ausbildungsordnungen)
- Weiterbildungsanzeigen
- ...

- Beschreibung einer zukünftigen Stelle
- Arbeitsaufgaben und Anforderungen aus Arbeitgeber-Perspektive

Zu den Beständen der SUB Göttingen gehören ein umfangreicher Altbestand mit wertvollen Inkunabeln, historischen Drucken und mittelalterlichen Handschriften, zahlreiche wissenschaftliche Nachlässe und das historische Bibliotheksarchiv. Für diese und weitere Materialien ist die Abteilung Spezi­alsammlungen und Bestandserhaltung zuständig, die sie unter Anwendung aktueller Metadatenstandards und damit in Hinsicht auf ihre Nutzung in den Digital Humanities erschließt und öffentlich, auch in Form von wissenschaftlichen Publikationen und Datenbanken, zugänglich macht.

Ihre Aufgaben:

- Katalogisierung von historischen Drucken des 15. bis 19. Jahrhunderts
- Erschließung von Handschriften und Nachlässen in den entsprechenden Datenbanken
- Mitarbeit in der Benutzung der Abteilung Spezi­alsammlungen und Bestandserhaltung
- Mitarbeit bei der Digitalisierung des historischen Altbestand und der Sondersammlungen der SUB

Erforderliche Fähigkeiten und Kenntnisse:

- Laufbahnprüfung für den gehobenen Dienst an Bibliotheken (Diplom-Bibliothekar*in) oder

erfolgreich abgeschlossenes Studium als Diplom-Bibliothekar*in (FH) oder Bachelor im Bereich Bibliotheks- und Informationsmanagement

- Erfahrungen in der Bearbeitung historischer Bestände und Sondersammlungen
- Erfahrungen in der Erfassung von Provenienzen
- sichere Kenntnisse bibliothekarischer Regelwerke (z.B. RDA, RNAB; Erfahrungen in der Erschließung alter Drucke sind von Vorteil)
- Kenntnisse in einer modernen Fremdsprache (z.B. Englisch, Französisch, Italienisch)

Wünschenswerte Fähigkeiten und Kenntnisse:

- grundlegende Kenntnisse der lateinischen Sprache
- gute kommunikative Fähigkeiten, freundliches Auftreten und starke Serviceorientierung
- Verantwortungsbereitschaft, ein gutes Zeitmanagement und Teamfähigkeit

Die Universität Göttingen strebt in den Bereichen, in denen Frauen unterrepräsentiert sind, eine Erhöhung des Frauenanteils an und fordert daher qualifizierte Frauen nachdrücklich zur Bewerbung auf. Sie versteht sich zudem als familienfreundliche Hochschule und fördert die Vereinbarkeit von Wissenschaft/Beruf und Familie. Der beruflichen Teilnehmenden sind die vielfältigen Möglichkeiten der Weiterbildung, insbesondere die Möglichkeit der Teilnahme an

- über 80 Millionen Stellenanzeigen seit 2015
- Quellen: Jobportale & Firmenwebsites
- Scraping und Metadaten-Anreicherung durch Firma Bullhorn (ehemals Textkernel)
- Weitere Vorverarbeitung, Textverarbeitung und Analyse im BIBB

- Stellenanzeigen werden unter Umständen auf mehreren Jobportalen gleichzeitig geschaltet
- jede Anzeige wird in Deutschland schätzungsweise über 2-5 Kanäle veröffentlicht
- Struktur der Anzeige kann sich dabei unterscheiden
- Erkennung von Dubletten als wesentliche Vorverarbeitung
- Kombination aus Dokumentenähnlichkeit (Doc2Vec + Kosinus-Ähnlichkeit) und strukturierten Metadaten [Gerlach, 2021]

- Korpus umfasst reguläre Stellen, Ausbildungsplätze, duale Studienplätze, Praktika, ...
- je nach Forschungsfrage kann nur ein Teil der Anzeigen relevant sein
- automatische Unterscheidung von Ausbildungs-Stellenanzeigen und anderen Stellenanzeigen [Krüger, 2023]
- Klassifikation von Stellenanzeigen-Texten mithilfe von Transformer-Modellen

Vorverarbeitung: Berufsklassifikation

- Beruf spielt in vielen statistischen Auswertungen eine zentrale Rolle
- Klassifikation der Stellenanzeigen in gängige Berufsklassifikationen
- Gruppierung von Berufen anhand von Tätigkeitsprofilen
- Erkennung von Berufsbezeichnungen

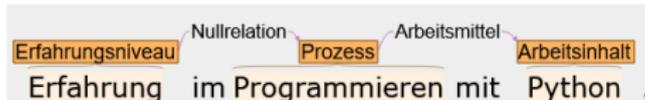


Klassifikation der Berufe 2010 (KIdB)

- Arbeitsaufgaben und Anforderungen als Hauptinteresse bei der Stellenanzeigen-Analyse
- Umsetzung als **Informationsextraktion**
- Modellierung von Anforderungen über Entitäten und Relation
[Krüger et al., 2025]

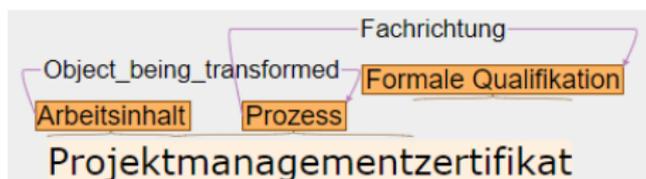
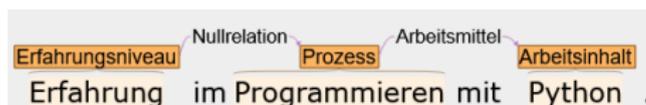
Skills-Extraktion aus Stellenanzeigen

- Arbeitsaufgaben und Anforderungen als Hauptinteresse bei der Stellenanzeigen-Analyse
- Umsetzung als **Informationsextraktion**
- Modellierung von Anforderungen über Entitäten und Relation [Krüger et al., 2025]



Skills-Extraktion aus Stellenanzeigen

- Arbeitsaufgaben und Anforderungen als Hauptinteresse bei der Stellenanzeigen-Analyse
- Umsetzung als **Informationsextraktion**
- Modellierung von Anforderungen über Entitäten und Relation [Krüger et al., 2025]



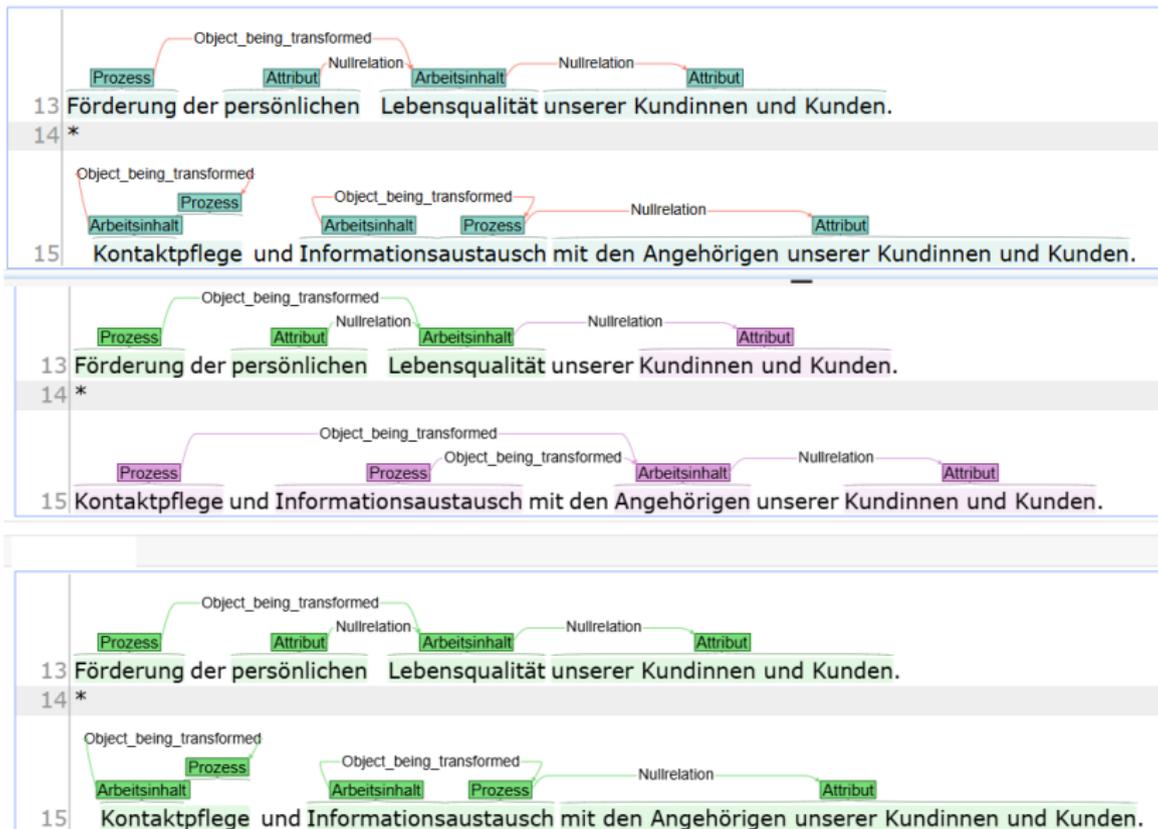
- Entwicklung einer Annotationsguideline für Anforderungen in Stellenanzeigen
- Ziele der Guideline
 - Einheitliches Verständnis von Anforderungen in Stellenanzeigen
 - konsistente Strukturen für Machine Learning-Algorithmen

- Entwicklung einer Annotationsguideline für Anforderungen in Stellenanzeigen
- Ziele der Guideline
 - Einheitliches Verständnis von Anforderungen in Stellenanzeigen
 - konsistente Strukturen für Machine Learning-Algorithmen

Annotationsworkflow

- Trainingsphase
 - 19 Mitarbeitende annotieren testweise 3 Stellenanzeigen
 - Annotationen werden mit Goldstandard verglichen
- Produktivphase
 - 500 Anzeigen werden von zwei Personen unabhängig annotiert
 - Kuration von einer dritten Person

Annotation und Kuration in Inception



Inter-Annotator Agreement (Krippendorff's alpha)

- Entitäten: Krippendorff's alpha 0.88
- Relationen: Krippendorff's alpha 0.8

Inter-Annotator Agreement (Krippendorff's alpha)

- Entitäten: Krippendorff's alpha 0.88
- Relationen: Krippendorff's alpha 0.8

Entitäten (Krippendorff's alpha)

Typ	IAA
Attitüde	0.87
Formale Qualifikation	0.87
Erfahrungsniveau	0.85
Berufsbezeichnung	0.83
Prozess	0.78
Arbeitsinhalt	0.75
Attribut	0.60
Branche	0.55

Relationen (Krippendorff's alpha)

Typ	IAA
Negation	0.90
Dringlichkeit	0.78
Alternative	0.75
Koordination	0.75
Object Being Transformed (OBT)	0.72
Fachrichtung	0.68
Zusammengehörige Entitätsteile	0.67
Autonomiegrad	0.62
Detail	0.62
Arbeitsmittel	0.61
Nullrelation	0.52



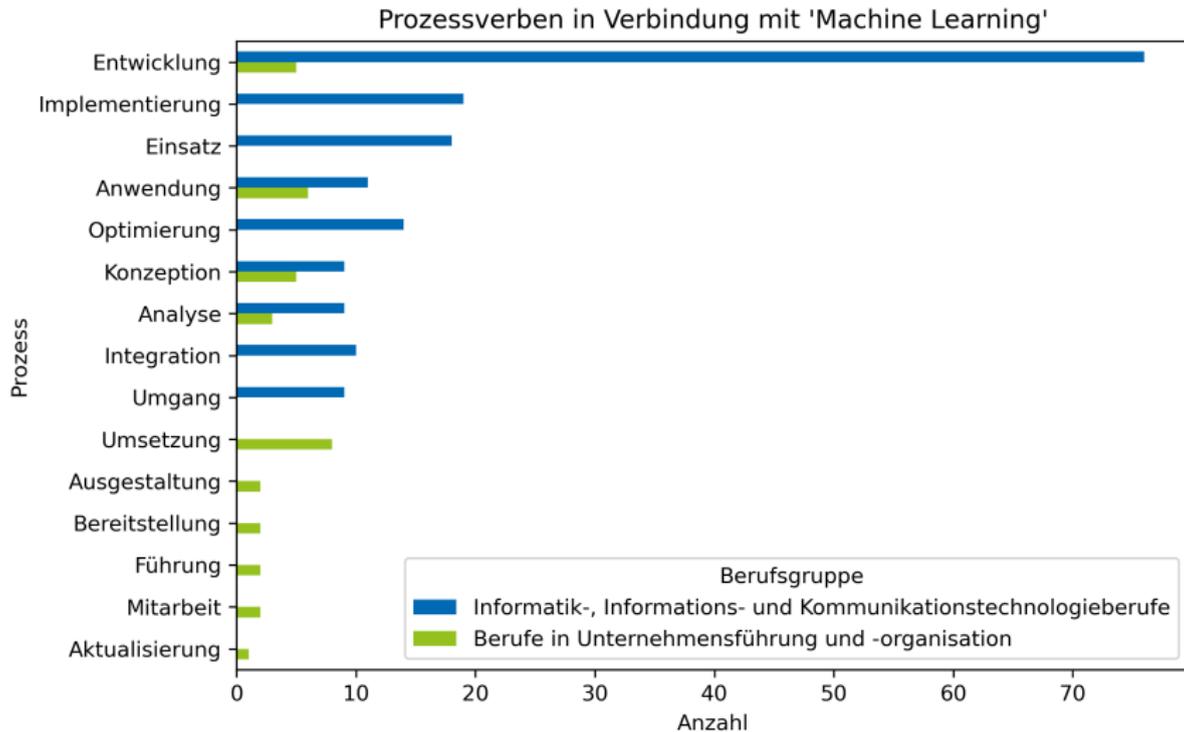
- fine-tuning verschiedener Transformer-Modelle (u.a. GermanBERT, jobBERT-de, XLM-RoBERTa)
- Transformer-Modelle können besonders gut Kontext-abhängige Bedeutungen erkennen



- fine-tuning verschiedener Transformer-Modelle (u.a. GermanBERT, jobBERT-de, XLM-RoBERTa)
- Transformer-Modelle können besonders gut Kontext-abhängige Bedeutungen erkennen
- Bestes Modell: XLM-RoBERTa
 - Entity: F1-Score 0.85
 - Relations: F1-Score 0.91

Ausblick: Inhaltliche Analysen

Welche Prozesse werden in Verbindung mit dem Arbeitsinhalt „Machine Learning“ beschrieben?



- laufend Themen und Daten für Abschlussarbeiten
- Dauerausschreibung für SHK-Stelle am BIBB
- weitere Informationen gerne im Nachgang an die Sitzung



Gerlach, A. (2021).

Entwicklung eines Verfahrens zur Dublettenerkennung bei Stellenanzeigen. Ein Methodenvergleich.
Master Thesis, Universität zu Köln, Köln.



Krüger, K. (2023).

Ausklasser - a classifier for German apprenticeship advertisements.

In Communication Papers of the 18th Conference on Computer Science and Intelligence Systems, Annals of Computer Science and Information Systems, pages 171–178. PTI.



Krüger, K., Akbik, A., Binnewitt, J., Ehmann, K., Grüner, L., and Winnige, S. (2025).

Improving Online Job Advertisement Analysis via Compositional Entity Extraction.