

Anpassung des Modells für die Extraktion berufsspezifischer Kompetenzen für die aktualisierte VerBIS Taxonomie

Bertelsmann Stiftung

Version

1

Autor*innen

Josephina Thiele & Rahkakavee Baskaran

Annotation und inhaltliche Beratung

Angela Pilgrim, Gunvald Herdin, Jana Fingerhut, Kerstin Ködding

Erstellt am

28. Oktober 2025

Letzte Änderung am

28. Oktober 2025

The logo for 'effect' features a stylized ampersand symbol followed by the word 'effect' in a bold, lowercase, sans-serif font.

Inhalt

1	Aktualisierung des Modells.....	3
2	Aktualisierung der Taxonomie.....	4
3	Manuelle Modellevaluierung	4
3.1	Verfahren	4
3.2	Datensatz	5
3.3	Ergebnisse	5
4	Vergleichende und explorative Analyse	6
4.1	Allgemeines Sample	6
4.2	Themenspezifische Sample	7
4.3	Ähnlichkeitsmetrik	9
4.4	Kompetenzen und Tätigkeitsbereiche	11
4.5	Visuelle Projektion	12
4.6	Anwendung.....	14
5	Fazit	14
6	Appendix	16
6.1	Annotationsregeln.....	16
6.2	Erweiterte Annotationsergebnisse	17
6.3	Erweiterte Analyse der themenspezifischen Sample	17
6.4	Erweiterung des repräsentativen Sampels	18
6.5	Erweiterte Ähnlichkeitsanalyse	18
	Literatur.....	21

1 Aktualisierung des Modells

Die Extraktions-Pipeline dient der Identifizierung und Kategorisierung berufsspezifischer Kompetenzen aus deutschsprachigen Stellenausschreibungen gemäß der VerBIS (Vermittlungs-, Beratungs- und Informationssystem der Bundesagentur für Arbeit) Taxonomie der Bundesagentur für die Arbeit (BA).¹ Der zweistufige Prozess umfasst die Phasen der Extraktion und Normalisierung der Kompetenzen nach VerBIS. [1]

Die VerBIS-Taxonomie wurde seit dem Training des Modells aktualisiert. Damit das Modell weiterhin nutzbar ist und die Extraktionen angepasst an den aktuellen Stand der Fachkompetenzen ist, muss das Modell ebenfalls aktualisiert werden. Die Anforderungen an das aktualisierte Modell sind dabei, dass sowohl weiterhin die bestehenden Kompetenzen als auch die neu hinzugekommene Kompetenzen gefunden werden. Außerdem sollte der Umfang der Modellparameter, die neu trainiert werden, wie auch die händischen Annotationen möglichst geringgehalten werden. Dafür können drei verschiedene Methoden herangezogen werden. Die erste Option ist, das Modell nicht neu zu trainieren, sondern nur die Taxonomie auszutauschen. Das ist möglich, da davon auszugehen ist, dass die neu hinzugekommenen Kompetenzen semantisch ähnlich zu den Kompetenzen der alten Taxonomie sind, und diese daher ähnlich gut von dem Modell aus den Stellenausschreibungen extrahiert werden. Der Vorteil hierbei ist, dass das Modell nicht neu trainiert werden muss, allerdings auch mit voraussichtlich schlechteren Ergebnissen gerechnet werden sollte. Für ein erneutes Training gibt es zunächst die Möglichkeit, ein Fine-Tuning des bestehenden Modells auf der neuen Taxonomie vorzunehmen. Dies hält den Trainingsaufwand geringer als die Erneuerung des gesamten Modells und es ist zu erwarten, dass die Extraktionen der neuen Kompetenzen präziser und korrekter sind als ohne erneutes Training. Die letzte Möglichkeit besteht darin, das gesamte Modell neu auf der Taxonomie zu trainieren. Diese Option erzielt voraussichtlich gute Ergebnisse in der Qualität der Extraktion sowohl neuer als auch alter Kompetenzen, ist allerdings auch mit einem erheblichen Neuaufwand verbunden. Hinzu kommt, dass für alle Optionen kein händisch annotierter Goldstandard für die neuen Kompetenzen existiert, sondern lediglich für die Kompetenzen der alten Taxonomie. Für alle drei Optionen muss die Extraktion auf einem Sample händisch annotiert werden, um die Richtigkeit der Ergebnisse zu überprüfen. Im Fall von Option drei sei jedoch bemerkt, dass hier die Evaluation auf einem größeren Sample sinnvoll wäre, da es sich um die Erstellung eines gesamten neuen Modells handelt.

Wir entschieden uns, die Optionen stufenweise abzuwiegen und mit einer explorativen Analyse der niedrigschwelligsten Methode zu beginnen. Dabei explorierten wir die Extraktion von 10000 Postings und beobachteten, dass zum einen alte Kompetenzen weiterhin gefunden werden, also auch neue Kompetenzen, in den Ergebnissen der Extraktions-Pipeline auftauchen. Darauf basierend, entschieden wir uns, dass Modell

¹ Die VerBIS Taxonomie kann im Download Portal der BA heruntergeladen werden [7]



nicht neu zu trainieren. In folgendem Bericht werden zum einen die Ergebnisse der händischen als auch der statistischen Analyse des Modells mit der aktualisierten Taxonomie dargelegt.

2 Aktualisierung der Taxonomie

Die VerBIS-Taxonomie besteht aus 273 Kompetenzgruppen, denen jeweils verschiedenen Fachkompetenzen zugeordnet werden. In der aktualisierten Taxonomie wurden die Bezeichnungen von 7 Kompetenzgruppen geändert; inhaltlich entsprechen diese jedoch weiterhin den bisherigen Gruppen. Insgesamt wurden 30 Konzepte entfernt und 847 neue Konzepte ergänzt, sodass die Taxonomie nun 8.785 Kompetenzen umfasst. Knapp ein Viertel (24,2%) der neuen Konzepte entfällt auf die folgenden fünf Gruppen: „Energieversorgung“, „EDV-Dienstleistungen“, „Programmiersprachen, Programm-bibliotheken, Entwicklungsumgebungen“, „Marketing, Vertrieb, Verkauf“ sowie „Betriebswirtschafts-, Finanz- und Controllingsoftware“.

3 Manuelle Modellevaluierung

In der händischen Analyse soll mit der Hilfe von Expert*innenwissen die Extraktion des Modells bewertet werden. Dafür wurde ein Sample auf Satzebene den Annotierenden bereitgestellt und die Ergebnisse der Bewertung analysiert. Das Verfahren der Annotation, sowie eine Übersicht des Samples und der Ergebnisse werden im Folgenden erläutert.

3.1 Verfahren

Die erste Stufe des Modells, die Lokalisierung der Kompetenzen, hat sich mit der Aktualisierung der Taxonomie nicht verändert, weshalb diese nicht mehr evaluiert wird. Zur Bewertung der Kompetenzextraktion aus Sätzen, die in der ersten Stufe als „enthält Kompetenz“ klassifiziert wurden, wurde ein Datensatz mit insgesamt 6.840 Sätzen annotiert. Nach einer Testrunde mit 100 zusätzlichen Sätzen annotierten zwei Personen den Datensatz in drei Annotationsrunden mit jeweils 1.200 Sätzen. Dabei überschritten sich 120 Sätze für beide Annotierenden. Jede Zeile des Datensatzes enthielt einen Satz aus einer Stellenausschreibung sowie ein extrahiertes Konzept des aktualisierten

Fachkompetenzmodells. Die Annotierenden bewerteten anschließend, ob die angegebene Kompetenz tatsächlich im Satz enthalten war. Vor Beginn der Annotation wurden Annotationsregeln definiert, ähnlich wie bei der ursprünglichen Evaluation des Modells [2]. Nach der Testrunde und nach jeder Annotationsrunde wurden diese Regeln anhand der Abweichungen bei den sich überschneidenden Sätzen weiter verfeinert. Die finalen Annotationsregeln sind im Appendix unter 6.1 zu finden.

3.2 Datensatz

Der Annotationsdatensatz umfasst 6.840 zufällig ausgewählte Sätze aus dem Annotationsdatensatz der Evaluation des ursprünglichen Modells. Abbildung 1 zeigt die Kompetenzen, die vom aktualisierten Modell am häufigsten extrahiert wurden. Es ist erkennbar, dass diese weitgehend mit den am häufigsten extrahierten Kompetenzen aus dem repräsentativen Sample aus 4.1 übereinstimmen. Abweichungen lassen sich dadurch erklären, dass der Annotationsdatensatz deutlich kleiner ist und die Kompetenzen auf Satz anstatt auf Postingebebene betrachtet werden, wodurch Unterschiede in der relativen Häufigkeit einzelner Kompetenzen auftreten können.

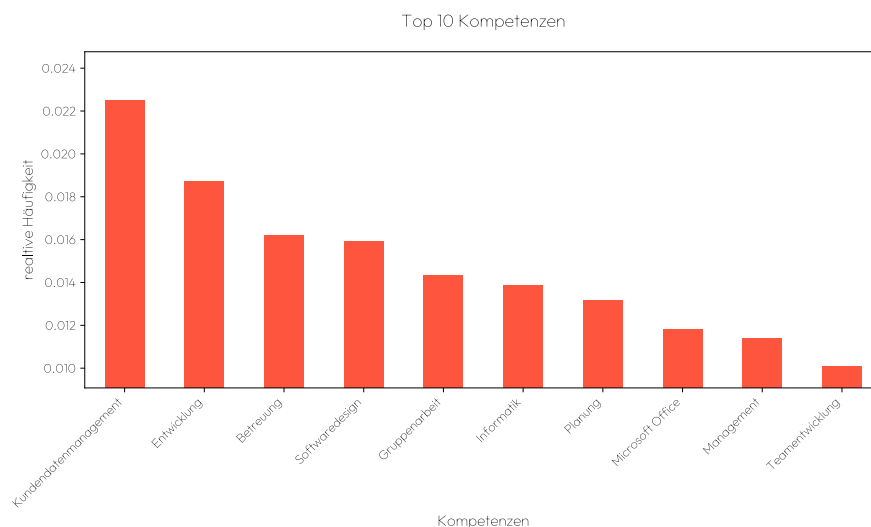


Abbildung 1: Top 10 Kompetenzen Annotationssample

3.3 Ergebnisse

Die Testrunde ergab eine hohe Übereinstimmung zwischen den Annotierenden von 80% und einem Cohens-Kappa-Wert von 0,66, weshalb die Annotation mit Runde zwei bis vier fortgesetzt werden konnte. Diese ergaben eine durchschnittliche Übereinstimmung von

85% und ein Cohens-Kappa-Wert von 0,70. Aus der Bewertung der Extraktion des Modells, konnte eine durchschnittliche Präzision von 0,56 für das aktualisierte Modell ermittelt werden. Für diese wurden die Annotationen der Sätze benutzt, die nur von einer Annotierenden bewertet wurden und jenen, für die die Einschätzung der beiden Annotierenden identisch war. Detaillierte Werte zur Übereinstimmung und Präzision der einzelnen Runden und Annotierenden sind im Appendix 6.2 aufgeführt. Eine mögliche Erklärung für die durchschnittlich niedrigere Präzision des aktualisierten Modells als die Makro-Präzision von 0,63 des ursprünglichen Modells [2] sind die restriktiveren Annotationsregeln. Beispielsweise wird eine Kompetenz, die erst während des Jobs erlernt werden soll, bei der alten Annotation als richtig gewertet, während sie bei der neuen Annotation als falsch eingestuft wird. An dieser Stelle sei bemerkt, dass durch die händische Annotation nur False Positive und True Positive Extraktionen ermittelt werden können, und somit als Metrik nur die Präzision berechnet werden kann.

Insgesamt lässt sich aus dem Ergebnis der händischen Evaluation schließen, dass das Modell Kompetenzen extrahiert, die tatsächlich in den Sätzen der Jobanzeigen vorkommen.

4 Vergleichende und explorative Analyse

Im Folgenden wird die Aktualisierung des Modells durch den Austausch der Taxonomie genauer analysiert. Zunächst wird eine Übersicht der Extraktion anhand eines repräsentativen Samples sowie themenspezifischer Stichproben dargestellt. Anschließend werden die Ergebnisse der alten und der neuen Taxonomie miteinander verglichen, indem die Ähnlichkeit zwischen den extrahierten Kompetenzen berechnet wird. Schließlich werden die Jobpostings Tätigkeitsfeldern zugeordnet, um die extrahierten Kompetenzen zu visualisieren und in einen inhaltlichen Kontext zu setzen. Diese Analysen dienen der Validierung und Bewertung des aktualisierten Modells.

4.1 Allgemeines Sample

Aus einem repräsentativen Sample der Textkernel-Jobpostings aus den Jahren 2018 bis 2024 [3] wurden 10.000 Postings zufällig ausgewählt, um die Extraktion des

Fachkompetenzmodells unter Verwendung der neuen Taxonomie zu analysieren.² Abbildung 2 zeigt die zehn am häufigsten erkannten Kompetenzen – einmal basierend auf der alten Taxonomie (links) und einmal auf der neuen Taxonomie (rechts). Neu hinzugekommene Fachkompetenzen sind farblich hervorgehoben. Es zeigt sich eine hohe Übereinstimmung zwischen den beiden Extraktionen: Neun der zehn am häufigsten extrahierten Kompetenzen unter der alten Taxonomie finden sich auch in den Top 10 der Extraktion mit der neuen Taxonomie wieder. Als neue Kompetenz kommt „Kundendatenmanagement“ hinzu, die erst in der aktualisierten Taxonomie eingeführt wurde. Dies zeigt, dass das Modell auch neue Kompetenzen zuverlässig identifizieren kann, obwohl es nicht explizit mit der neuen Taxonomie trainiert wurde.

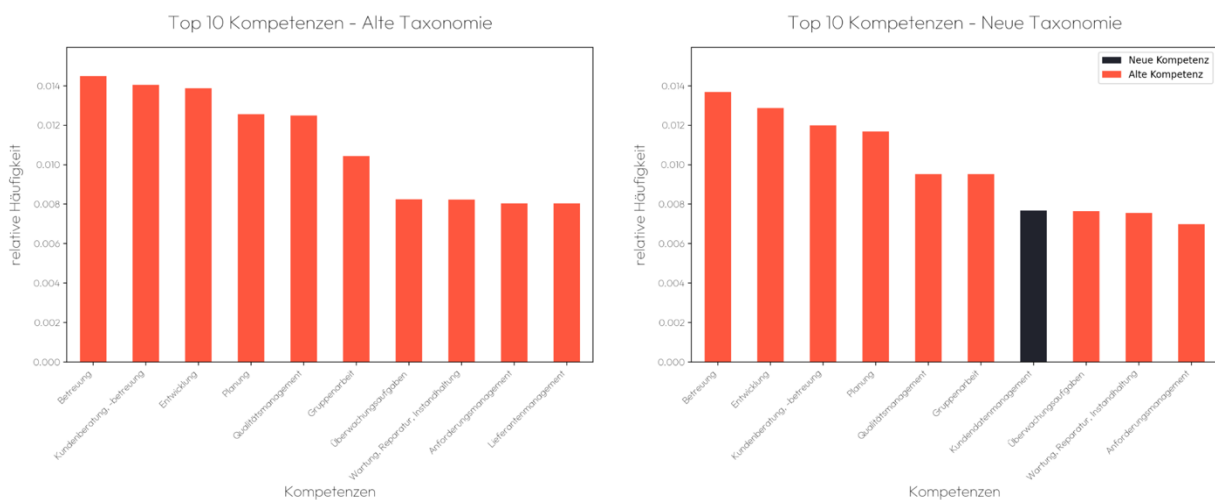


Abbildung 2: Top 10 Kompetenzen – Allgemeines Sample

4.2 Themenspezifische Sample

Die aktualisierte VerBIS-Taxonomie enthält insbesondere neue Kompetenzen aus Bereichen, die sich in den vergangenen Jahren stark weiterentwickelt haben, darunter Erneuerbare Energien und Künstliche Intelligenz. Mithilfe der Modelle zur Klassifizierung von Online-Stellenausschreibungen im Bereich Erneuerbare Energien bzw. Künstliche Intelligenz [4] wurden weitere Sample generiert. Nicht alle Stellenausschreibungen haben eine Extraktion für Erneuerbare Energie oder Künstliche Intelligenz. Für eine präzisere Analyse wurden deshalb bei der Auswahl ausschließlich Stellenausschreibungen berücksichtigt, die Konzepte des jeweiligen Modells enthalten. Aus dieser Auswahl wurden

² Die regelmäßige Erweiterung des Samples wird in Appendix 6.4 erläutert.

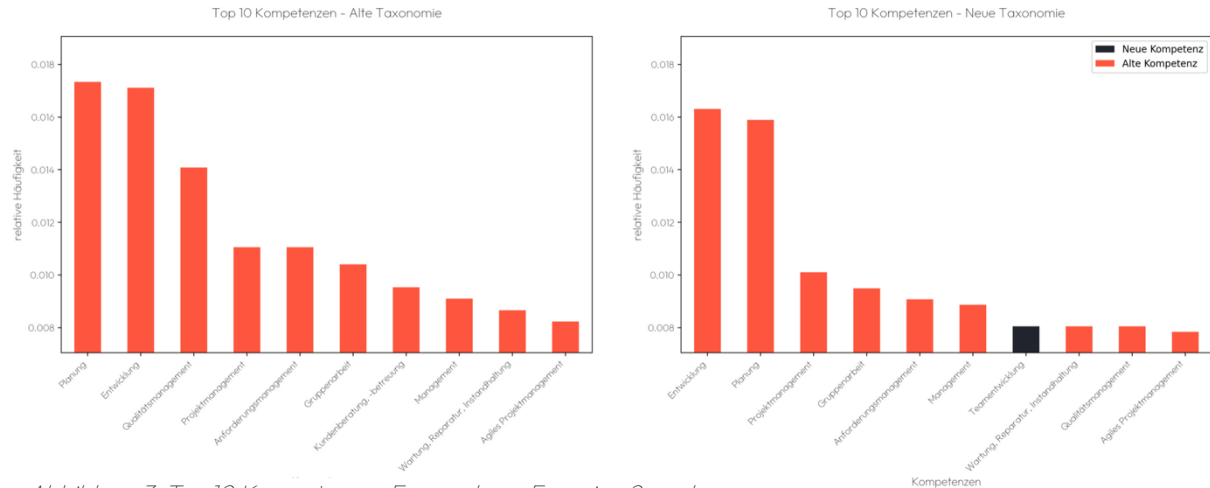


Abbildung 3: Top 10 Kompetenzen Erneuerbare Energien Sample

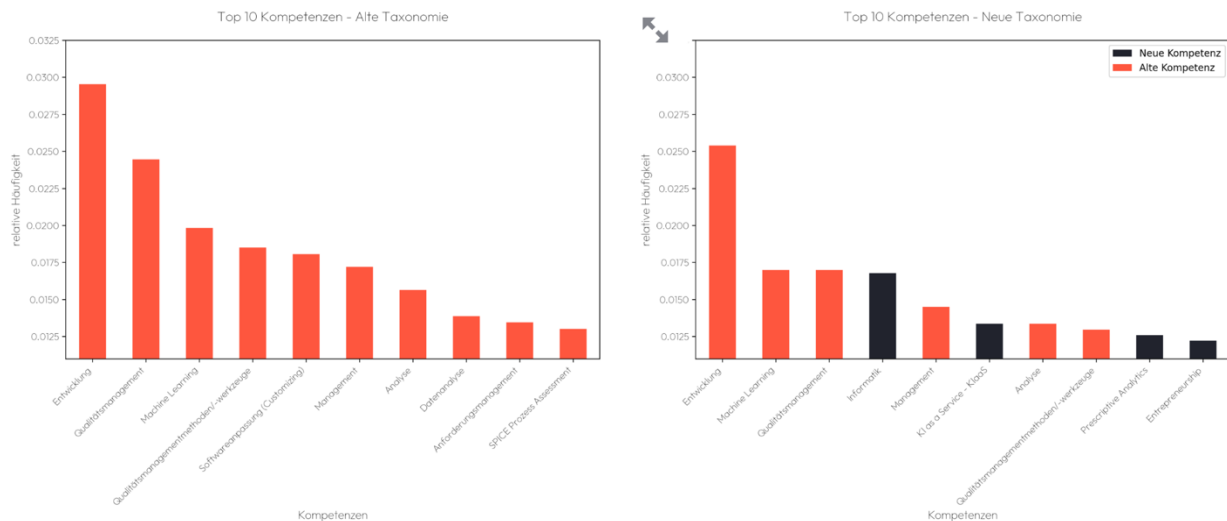


Abbildung 4: Top 10 Kompetenzen Künstliche Intelligenz Sample

dann jeweils 500 Jobpostings zwischen 2018 und 2024 zufällig für die themenspezifische Analyse herausgenommen. Die Ergebnisse der zehn am häufigsten extrahierten Kompetenzen – einmal basierend auf der alten (links) und einmal auf der neuen (rechts) Taxonomie – sind in Abbildung 3 für das Erneuerbare-Energien-Sample und in Abbildung 4 für das Künstliche-Intelligenz-Sample dargestellt. Es zeigt sich, dass insbesondere in Ausschreibungen aus dem Bereich Künstliche Intelligenz neue Kompetenzen aus diesem Themenfeld häufig gefunden werden. Das bestätigt, dass das neue Modell in der Lage ist die neu hinzugekommenen Kompetenzen zu finden und bildet daher die Fachkompetenzen der Stellenausschreibungen in Berufsbereichen, die sich in den letzten Jahren stark gewandelt haben, korrekt ab. Im Appendix 6.3 sind zudem die am häufigsten neu extrahierten Kompetenzen gesondert aufgeführt.

4.3 Ähnlichkeitsmetrik

Eine Möglichkeit, die Extraktion mithilfe der neuen Taxonomie mit derjenigen auf Basis der alten Taxonomie zu vergleichen, besteht darin, die Ähnlichkeit zwischen den Ergebnissen zu berechnen. Hierfür wurde der Jaccard-Index für Mengen verwendet, der für zwei Mengen A und B wie folgt definiert ist:

$$\text{Jaccard-Index: } J(A,B) = |A \cap B| / |A \cup B|$$

Da in der neuen Taxonomie viele Kompetenzen enthalten sind, die in der alten nicht vorkommen, wurden die Ergebnisse der Extraktion auf dem 10.000er-Sample auf jene Kompetenzen gefiltert, die in der Schnittmenge beider Taxonomien liegen. Diese Filterung stellt sicher, dass die Ähnlichkeitsmetrik die Qualität des Modells unter der neuen Taxonomie angemessen bewertet, da neu hinzugekommene Kompetenzen sonst zu einem niedrigeren Ähnlichkeitswert führen würden. Die Extraktion neuer Kompetenzen wird hingegen als Indikator für eine gute Performance des Modells und die Übertragbarkeit auf die neue Taxonomie gewertet. Tabelle 1 zeigt beispielhaft, welche Fälle bei dem Vergleich der beiden Extraktionen auftreten können. Die Kompetenzen aus den Spalten „Hinzugefügte Kompetenzen“, sowie „Herausgenommene Kompetenzen“ werden für die Berechnung der Ähnlichkeitsmetrik nicht verwendet.

Die nach der Filterung verbleibenden Kompetenzen wurden pro Posting zu Mengen ohne Duplikaten aggregiert, und anschließend wurde der Jaccard-Index je Posting berechnet. In Abbildung 5 ist die Verteilung der Ähnlichkeitswerte dargestellt. Die x-Achse zeigt die Werte des Jaccard-Indexes (zwischen 0 – keine Übereinstimmung – und 1 – identische Extraktion), die Balkenhöhe die Häufigkeit der jeweiligen Werte, normiert auf die Gesamtzahl der Postings. Zusätzlich ist farblich hervorgehoben, für welche Postings Kompetenzen herausgefiltert wurden, weil sie entweder neu in der aktualisierten Taxonomie hinzugefügt sind oder aus der alten Taxonomie herausgenommen wurden. In Tabelle 1 kennzeichnet die erste Spalte, ob es sich um ein Posting mit Kompetenzen, die nur in einer der beiden Taxonomien vorkommt, handelt oder nicht.

Posting mit Kompetenzen, die nur in einer Taxonomie vorkommen	Neue Extraktion	Alte Extraktion	Hinzugefügte Kompetenzen	Herausgenommene Kompetenzen
Ja	[Betreuung, Angebotsmanagement, Drohnentechnik, Training]	[Drohnenführerschein (Kenntnisnachweis gemäß § 21a Abs. 4 LuftVO), Betreuung, Angebotsmanagement, Training]	[Drohnen- technik]	[Drohnen- führerschein (Kenntnisnach-weis gemäß § 21a Abs. 4 LuftVO)]
Ja	"Netzwerktech- nik MS Internet Information Server", "Informatik", "Anwenderberatung, Anwender- Support (IT)", "Außendienst", "Systembe- treuung, System- administration, System- verwaltung", "Kundendienst", "Versorgungstechnik", "Elektro- installation"]	"Netzwerktechnik MS Internet Information Server", "Anwender- beratung, Anwender- Support (IT)", "Außendienst", "Systembetreuung, Systemad- ministration, Systemverwaltung", "Kundendienst", "Versorgungs- technik", "Elektroinstallation"	"Informatik"	/
Nein	"Social- Media Kommunikation", "Forstwirt- schaft", "Waldpflege- arbeiten", "Internet- kompetenz", "HTML-Editor und Webdesign-Software Frontpage"	"Waldpflege- arbeiten", "Forstwirtschaft"	/	/

Tabelle 1: Filterung der Ergebnisse für die Ähnlichkeitsmetrik

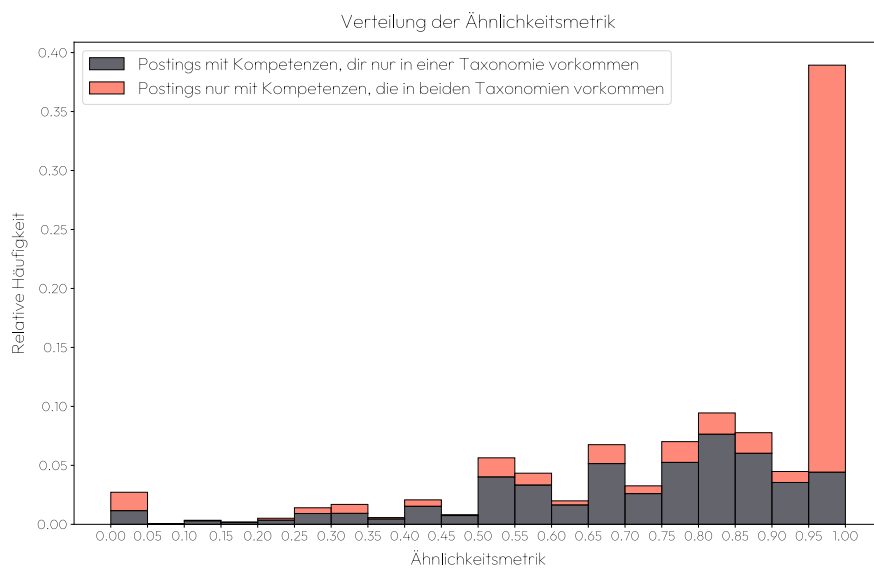


Abbildung 5: Verteilung der Ähnlichkeitsmetrik mit Filterung der Ergebnisse

Aus dem Diagramm geht hervor, dass knapp 40 % der Postings einen Ähnlichkeitswert zwischen 0,95 und 1,0 erreichen. 50 % der Postings weisen einen Wert von mindestens 0,86 auf, und 75 % einen Wert von mindestens 0,67. Es ist deutlich zu erkennen, dass der Anteil an Postings, bei denen Kompetenzen herausgefiltert wurden, im Vergleich zu Postings ohne Filterung größer für Ähnlichkeitswerte unter 0,95 ist. Dies spiegelt sich auch in den durchschnittlichen Ähnlichkeitswerten wider: Im gesamten Sample beträgt der Durchschnitt 0,79, bei Postings mit herausgefilterten Kompetenzen 0,70 und bei Postings ohne Filterung 0,88. Die überwiegend hohen Ähnlichkeitswerte bestätigen, dass die Qualität der Extraktion mit dem Austausch der Taxonomien erhalten bleiben. Die unterschiedlichen Durchschnittswerte für Postings mit Kompetenzen, die nur in einer Taxonomie enthalten sind, zu denen mit nur Kompetenzen, die in beiden Taxonomien enthalten sind, ist ein Indikator dafür, dass hier alte Kompetenzen mit Neuen ersetzt wurden. Da die Extraktion von neuen Kompetenzen erwünscht ist und gerade aktualisierte Kompetenzen in der Taxonomie sich auch in der Extraktion widerspiegeln sollen, bewerten wir diese Ergebnisse ebenfalls als positiv für die Evaluation des neuen Modells.

4.4 Kompetenzen und Tätigkeitsbereiche

Zur besseren Einschätzung der Gesamtqualität wird im Folgenden – neben dem direkten Vergleich der Extraktionsergebnisse – der Zusammenhang zwischen den in den Stellenausschreibungen genannten Berufen und den geforderten Kompetenzen hergestellt.

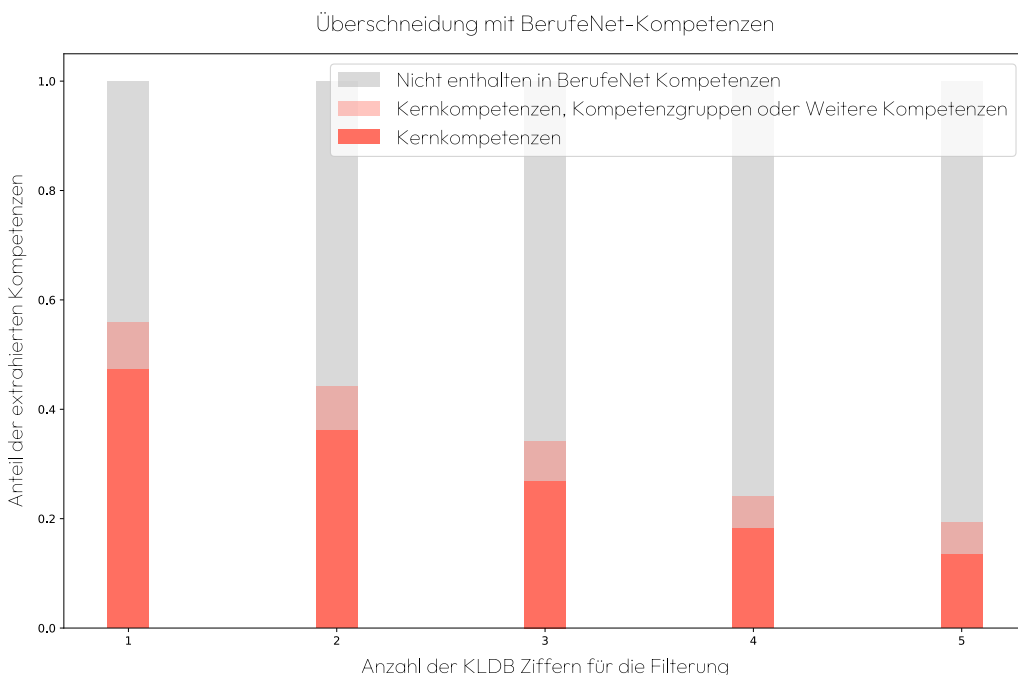


Abbildung 6: Überschneidung der Extraktionen mit den BerufeNet-Kompetenzen

Hierfür wird das Modell für die Klassifizierung der Berufe [1] benutzt, welches Jobtitel den Berufen aus der KLDB-Taxonomie [5] zuordnet. In dieser erhält jeder Beruf eine fünfstellige KLDB-Nummer, wobei jede Ziffer eine Ebene der beruflichen Klassifikation repräsentiert: Die erste Ziffer steht für die grösste Einordnung, die letzte für die spezifischste. Die Bundesagentur für Arbeit (BA) stellt für verschiedene Berufe aus der KLDB-Taxonomie die zugeordneten Kompetenzen, unterteilt in Kernkompetenzen, Kompetenzgruppen und weitere Kompetenzen, bereit [6]. Das 10.000 Sample wurde mit der Extraktion des Modells zur Klassifizierung der Berufe angereichert, um die Ergebnisse des Fachkompetenzen-Modells mit den definierten Kompetenzen der BA zu vergleichen. Abbildung 6 zeigt die Überschneidung nach verschiedenen Kriterien. Die KLDB-Ziffern dienen dabei als Filterkriterium, unter dem die extrahierten Kompetenzen mit den BA-Kompetenzen verglichen wurden. Eine Analyse auf Basis der ersten KLDB-Ziffer bedeutet, dass die extrahierten Kompetenzen eines Postings mit allen Kompetenzen verglichen werden, die die BA für Berufe mit derselben ersten KLDB-Ziffer definiert. Farbliche Hervorhebungen zeigen, ob eine extrahierte Kompetenz zu den Kernkompetenzen, zu einer Kompetenzgruppe oder zu weiteren Kompetenzen gehört bzw. nicht enthalten ist. Bei einem Vergleich auf Basis nur einer KLDB-Ziffer liegen 47 % der extrahierten Kompetenzen in den zugehörigen Kernkompetenzen und 56 % in einer der drei Kompetenzkategorien. Wie erwartet, sinkt die Übereinstimmung, je restriktiver der Filter gewählt wird – also je mehr Stellen der KLDB-Nummer für die Übereinstimmung herangezogen werden.

4.5 Visuelle Projektion

Die Extraktion der Fachkompetenzen wird im Folgenden grafisch visualisiert. Dabei wird angenommen, dass Berufe ähnliche Kompetenzprofile aufweisen. Folglich sollten Stellenausschreibungen mit demselben KLDB-Beruf auch untereinander ähnlich sein. Unter dieser Annahme sollten sich bei einer guten Gesamtpformance der Extraktion Cluster bilden, die Stellenausschreibungen desselben Berufs umfassen.

Zur besseren Vergleichbarkeit mit der grafischen Validierung der alten Extraktion wurde die Methodik aus Teil 6 in [1] reproduziert, wobei der Parameter ‚min_dist‘ auf 0,01 gesetzt wurde, da dies zu visuell besser interpretierbaren Ergebnissen führte.

In der Projektion Abbildung 7 ist erkennbar, dass Berufe wie „Gesundheits-, Krankenpfleger(o.S.)-Fachkraft“ und „Verkauf (ohne Produktspezial.)-Fachkraft“ deutlich trennbare Cluster bilden. Dies wurde für erstgenannten Beruf bereits in der

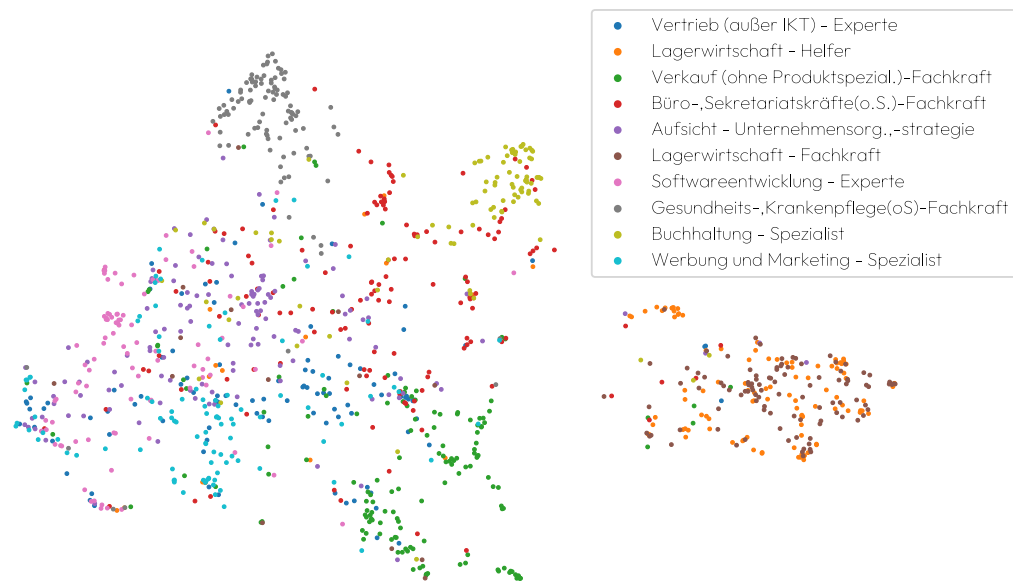


Abbildung 7: UMAP Visualisierung für Kompetenzen-Embeddings

Visualisierung der Extraktion mit der alten Taxonomie beobachtet. Der Beruf „Verkauf (ohne Produktspezial.)-Fachkraft“ war in der früheren Visualisierung weniger klar trennbar und vermischte sich dort mit anderen Berufsfeldern – von denen einige in der aktuellen Visualisierung nicht enthalten sind. Eine mögliche Erklärung hierfür ist, dass diese fehlenden Berufe thematisch näher an „Verkauf (ohne Produktspezial.)-Fachkraft“ liegen als die nun abgebildeten Berufe. Für Berufe, die in beiden Visualisierungen enthalten sind, deren Cluster sich aber in der neuen Darstellung besser trennen lassen, kann eine geringere Fallzahl eine Rolle spielen, sodass sich die Extraktionen zufällig besser separieren.

Die Berufe „Lagerwirtschaft – Helfer“ und „Lagerwirtschaft – Fachkraft“ überlappen in der Projektion stark, was zu erwarten ist, da sie ähnliche Kompetenzanforderungen aufweisen. Zusammen bilden sie jedoch – wie bereits in der alten Visualisierung – ein gut trennbares Cluster.

Der Beruf „Buchhaltung – Spezialist“ ist in beiden Visualisierungen trennbar, überlappt jedoch teilweise mit „Büro-, Sekretariatskräfte(o.S.)-Fachkraft“. Letzterer überschneidet sich ebenfalls in beiden Projektionen mit weiteren Berufsfeldern. Eine mögliche Erklärung dafür ist, dass diese Tätigkeit ein sehr breites Spektrum an Kompetenzen erfordert und daher häufig Überschneidungen mit anderen Berufen aufweist [1]. Gleiches gilt für „Aufsicht – Unternehmensorg., -strategie“, das sich in beiden Visualisierungen mit anderen Berufsfeldern überdeckt.

Die Berufe „Werbung und Marketing – Spezialist“ sowie „Softwareentwicklung – Experte“ sind teilweise als Cluster erkennbar, zeigen jedoch ebenfalls deutliche Überschneidungen. Auch hier könnte ein breites Anforderungsspektrum und insbesondere eine gemeinsame Basis an digitalen Kompetenzen eine plausible Erklärung darstellen.

Insgesamt zeigt die visuelle Projektion, dass sich die extrahierten Kompetenzen eines Berufsfelds oft ähnlich sind und wiederum abgrenzbar zu den Kompetenzen anderer Berufe sind. Auch im Vergleich zur visuellen Projektion der alten Extraktion, können wir feststellen, dass die Sichtbarkeit der Cluster nicht abgenommen hat und somit die Aussagekraft der Extraktion erhalten geblieben ist.

4.6 Anwendung

Eine mögliche Anwendung der Filterung aus 4.4 ist die UMAP-Visualisierung auf den gefilterten Daten. Abbildung 8 zeigt diese für die Extraktion gefiltert nach allen fünf KLDB-Ziffern und nach Kernkompetenzen und Kompetenzgruppen. Es ist deutlich erkennbar, dass alle 10 Berufe nun Cluster bilden, die besser trennbar sind als in der Visualisierung der Extraktionen ohne Filterung nach dem BerufeNet-Mapping [6].

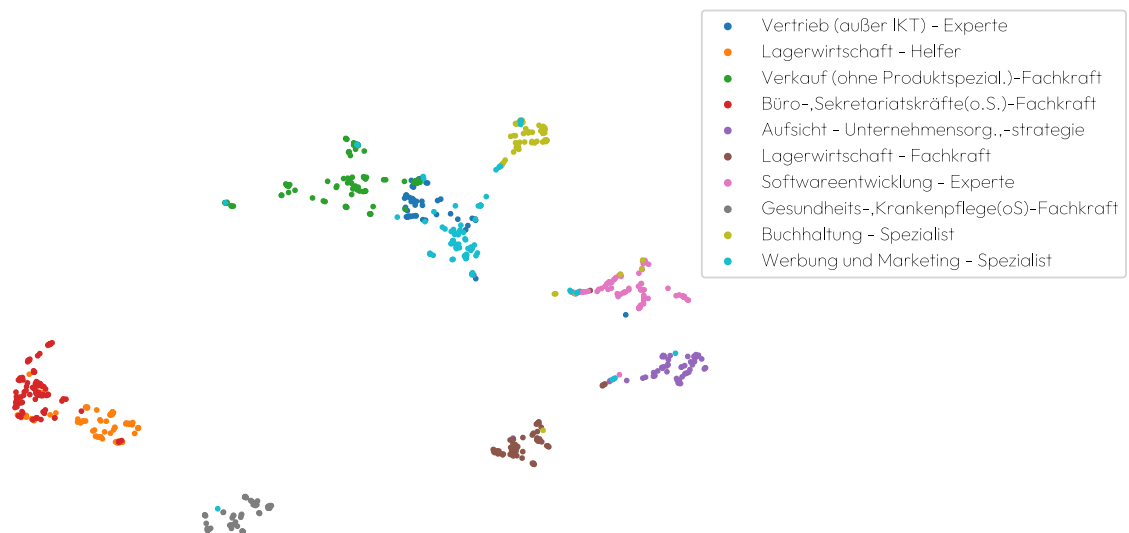


Abbildung 8: UMAP Visualisierung der gefilterten Kompetenz-Embeddings

5 Fazit

Die Analyse des Modells zeigt, dass durch die Aktualisierung des Fachkompetenzenmodells mit der neuen Taxonomie ohne neues Training, die Qualität der Ergebnisse erhalten bleibt. Mit Hilfe der manuellen Annotation der Extraktion des neuen Modells, wurde Expert*innenwissen mit eingebunden, um die diese zu bewerten. Der daraus resultierende Wert von 0,56 für die Präzision des Modells, bestätigt die Korrektheit der extrahierten Kompetenzen. Neben der händischen Analyse bietet die Auswertung der Extraktionen auf

dem repräsentativen Sample eine Übersicht über das neue Modell und die auf den themenspezifischen Samplen belegen, dass das Modell in der Lage ist die neu hinzugekommenen Kompetenzen zu finden. Die Ähnlichkeitsanalyse bestätigt, dass ein Großteil der extrahierten Kompetenzen des neuen Modells mit denen des alten Modells übereinstimmen. Mit Hilfe der Verknüpfung der Extraktion der Fachkompetenzen mit den Tätigkeitsbereichen, konnten die Ergebnisse in Kontext gebracht werden und mit Kompetenzen, die von der BA für die unterschiedlichen Tätigkeitsbereiche definiert sind, verglichen werden. Die visuelle Projektion ergab, dass sich die extrahierten Kompetenzen oft in Cluster für die verschiedenen Berufe abgrenzen lassen. Insgesamt konnten wir damit die Aktualisierung des Modell validieren und die Ergebnisse auf einem repräsentativen Sample analysieren.

6 Appendix

6.1 Annotationsregeln

Dateiaufbau:

Die Excel-Datei enthält drei Spalten:

- evaluation_id – interne ID, bitte nicht beachten
- sentence – ein Satz aus einer Jobanzeige
- result – eine Fachkompetenz, die vom Modell extrahiert wurde
- evaluation – hier bitte eintragen:
 - 1 → passt
 - 0 → passt nicht
 - 1 → unsicher

Inhaltliche Hinweise:

- Nicht zu tief analysieren – wenn etwas grundsätzlich passen könnte, lieber 1 geben.
- Die Kompetenz muss nicht wortwörtlich genannt sein – auch sinngemäße Aussagen zählen.
- Wenn Vermutung da, dass das Wissen benötigt wird, eine 1 geben
- Gedanklich verschiedene Definitionen der Fachkompetenz zulassen (Beispiel Betreuung, Pflege im IT Bereich etc.)
- Komplette Falsches raus
- Jede Zeile unabhängig bewerten

Inhaltliche Ergänzungen nach den ersten Annotationsrunden:

- falls Begriffe nicht bekannt: zunächst googlen/ AI Chatbot fragen, falls dann immer noch unklar → '-1'
- Zeilen unabhängig bewerten: besonders, wenn ein Kontext nicht explizit erwähnt wird, nicht mit dem Kontext der vorherigen Zeile bewerten
- Kompetenz mit Spezifizierung in Klammern: Wenn der Hauptbegriff nicht enthalten ist → '0'; wenn der Hauptbegriff enthalten ist und der Satz nicht dem widerspricht, was in Klammern steht → '1'
- Wenn aus dem Satz nicht klar hervorgeht, ob es sich um die Beschreibung des Unternehmens oder der Person handelt → im Zweifel um die Person, also '1'; genauso aber, wenn klar hervorgeht, dass es sich um das Unternehmen handelt (auch wenn die Kompetenz drinsteht) → '0'

- wenn Kompetenzen erst erlernt werden sollen: '0'
- bei unbekannten Begriffen googlen oder z.B. AI-Chatbot fragen, ob ein Zusammenhang besteht
- wenn Kompetenz wörtlich vorkommt: '1'

6.2 Erweiterte Annotationsergebnisse

In Tabelle 2 werden die Übereinstimmung und die Werte für das Cohens Kappa der einzelnen Runden aufgeschlüsselt. Tabelle 3 bildet die Ergebnisse für die ermittelte Präzision der einzelnen Runden und Annotierenden ab.

Runde	Übereinstimmung	Cohens Kappa
1	80%	0,66
2	87%	0,71
3	85%	0,70
4	86%	0,71

Tabelle 2: Übereinstimmung zwischen den Annotierenden

Runde	Annotierende	Präzision
1	1	0,64
1	2	0,62
2	1	0,58
2	2	0,61
3	1	0,54
3	2	0,58
4	1	0,46
4	2	0,58

Tabelle 3: Ergebnisse - Präzision des neuen Modells

6.3 Erweiterte Analyse der themenspezifischen Sample

In Abbildung 9 werden die Kompetenzen dargestellt, die nur in der neuen Taxonomie vorkommen und am häufigsten extrahiert wurden für das Erneuerbare Energien Sample links und das Künstliche Intelligenz Sample rechts.

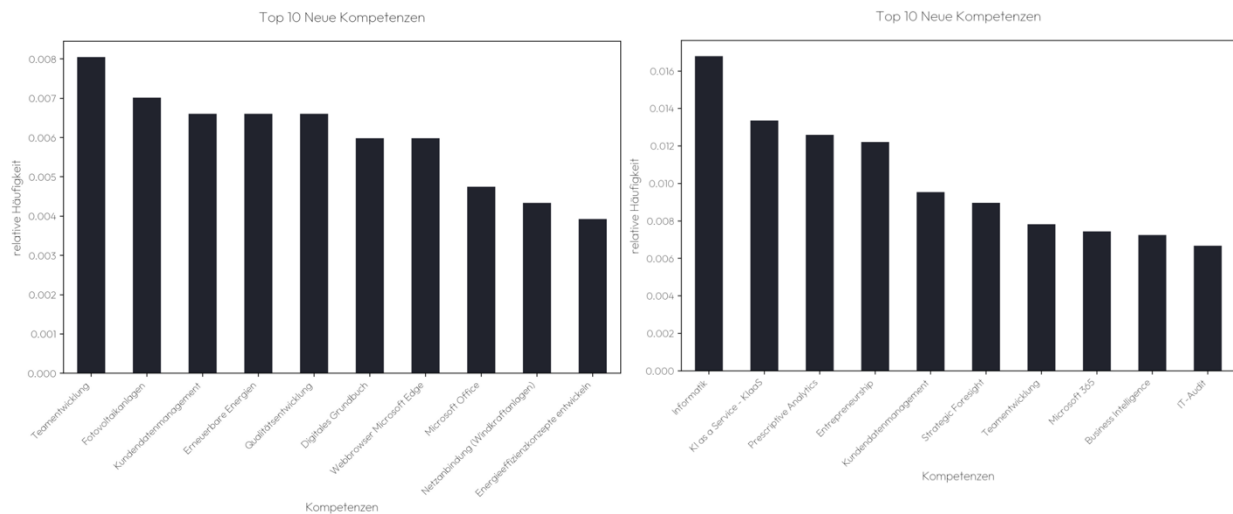


Abbildung 9: Themenspezifische Sample Top 10 neue Kompetenzen

6.4 Erweiterung des repräsentativen Sampls

Das repräsentative Sample wird regelmäßig – halbjährlich oder jährlich – um Postings aus neueren Zeiträumen erweitert. Dabei wird darauf geachtet, dass der Anteil der Postings aus den einzelnen Zeitabschnitten konstant bleibt, um Verzerrungen in den Analysen zu vermeiden. Die Erweiterung erfolgt dabei monatsweise, wobei jeweils 19 % der verfügbaren Postings aufgenommen werden. Dies entspricht zum einen der Verteilung des ursprünglichen Samples und stellt zum anderen sicher, dass die Daten aus früheren Zeiträumen unverändert bleiben, sodass Analysen mit den bisherigen Ergebnissen vergleichbar bleiben.

6.5 Erweiterte Ähnlichkeitsanalyse

Wie in der Ähnlichkeitsanalyse in Abschnitt 4.3 beschrieben, wurden die Ergebnisse zur Berechnung der Ähnlichkeitsmetrik auf jene Kompetenzen gefiltert, die in beiden Taxonomien vorkamen. Im Folgenden werden die daraus resultierenden Unterschiede näher betrachtet.

Berücksichtigt man ausschließlich die Postings, bei denen sich die extrahierten Kompetenzen durch die Filterung geändert haben, ergeben sich für die Quantile und den Durchschnittswert niedrigere Werte als für jene Postings, bei denen keine Kompetenzen herausgefiltert wurden (Tabelle 4). Dieser Unterschied lässt sich auch in den jeweils separaten Verteilungen in Abbildung 10 erkennen. Dies bestätigt, dass die Ähnlichkeit bei Postings, für die neue Kompetenzen extrahiert wurden, geringer ausfällt. Eine mögliche

Erklärung hierfür ist, dass bei diesen Postings alte Kompetenzen durch Neue ersetzt wurden und die ursprünglichen Konzepte daher in den gefilterten Ergebnissen nicht mehr enthalten sind.

	mit Postings mit gefilterten Kompetenzen	ohne Postings mit gefilterten Kompetenzen
25%	0,43	0,83
50%	0,6	1,00
75%	0,75	1,00
Durchschnitt	0,57	0,88

Tabelle 4: Quantile und Durchschnitt der Ähnlichkeitsmetrik für Postings mit und ohne gefilterten Kompetenzen

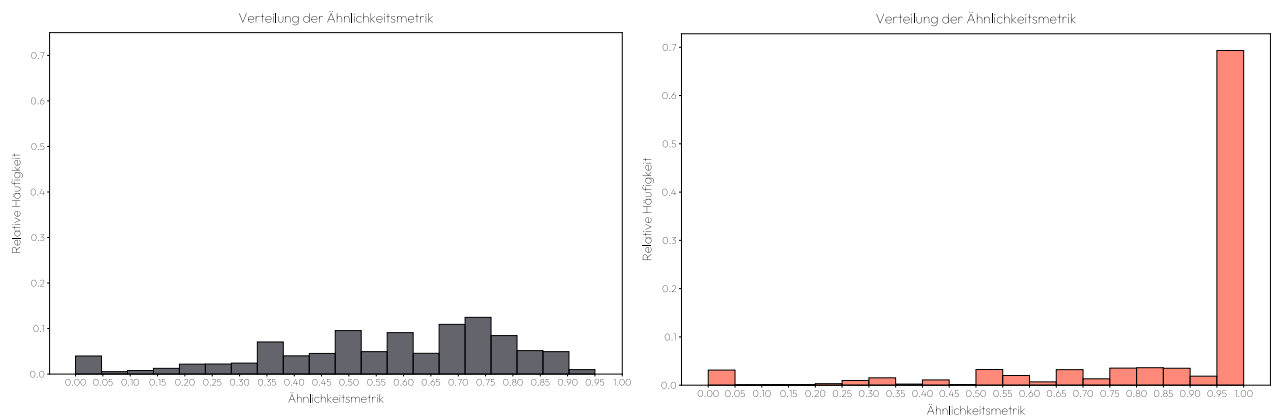


Abbildung 10: Die Verteilung der Ähnlichkeitsmetrik nur für Postings, bei denen keine Kompetenzen herausgefiltert wurden (rechts) und nur für Postings, bei denen Kompetenzen herausgefiltert wurden (links).

Abbildung 11 zeigt die Verteilung der Ähnlichkeitsmetrik ohne Filterung. Dabei ist zu erkennen, dass lediglich knapp 35 % – anstatt 40 % – der Extraktionen eine sehr hohe Übereinstimmung (0,95 bis 1,0) aufweisen. Insgesamt verschiebt sich die Verteilung hin zu niedrigeren Ähnlichkeitswerten. Dies ist erwartbar, da in diesem Fall auch Extraktionen berücksichtigt werden, bei denen das Modell neue Kompetenzen identifiziert hat, die möglicherweise eine frühere Kompetenz sinngemäß ersetzen und dennoch negativ in die Ähnlichkeitsberechnung eingehen.

Neben der Verteilung der Ähnlichkeitsmetrik für das gesamte Sample werden im Folgenden exemplarische Postings mit unterschiedlichen Ähnlichkeitswerten dargestellt – einmal für die gefilterte Variante (Tabelle 5) und einmal für die unfilterte Variante (Tabelle 6).

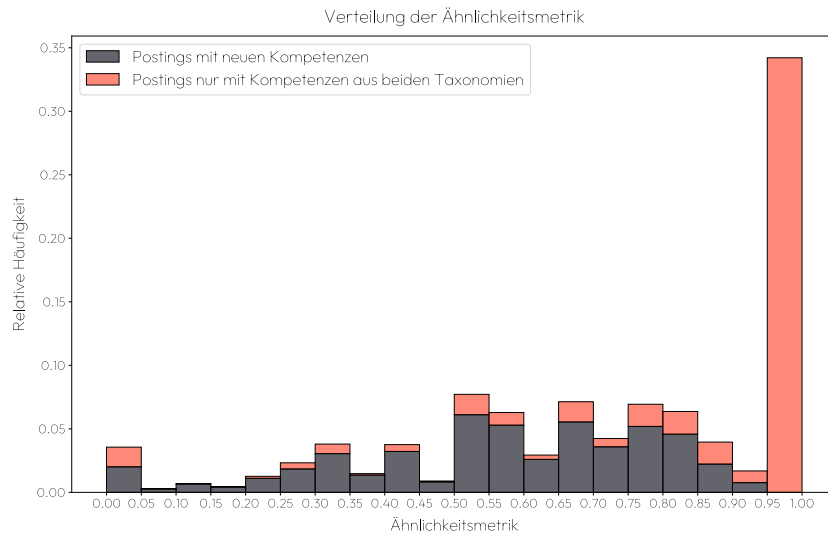


Abbildung 11: Verteilung der Ähnlichkeitsmetrik ohne Filterung der Ergebnisse

Ähnlichkeit	Neue Taxonomie	Alte Taxonomie
0	"Evaluation", "Softwareanpassung (Customizing)"	"IT-Architekturmanagement"
0,25	"Inbetriebnahme"	"Wärmepumpen", "Kraftwerkstechnik", "Inbetriebnahme", "Elektroinstallation"
0,5	"Speisekarten, -pläne zusammenstellen", "Einzelunterricht"	"Speisekarten, -pläne zusammenstellen", "Einzelunterricht", "Allergologie", "Präsentation"
0,86	"Daten-, Texterfassung", "3D-Design", "GIS (Geo-Info-System)", "Bestandspflege (Archiv, Bibliothek)", "Geoinformationssystem PolyGIS", "IATA-Bestimmungen"	"Daten-, Texterfassung", "3D-Design", "GIS (Geo-Info-System)", "Bestandspflege (Archiv, Bibliothek)", "Geoinformationssystem PolyGIS", "IATA-Bestimmungen"

Tabelle 5: Beispiele Ähnlichkeitsmetrik der Ergebnisse gefiltert nach Kompetenzen, die in beiden Taxonomien existieren

Ähnlichkeit	Neue Taxonomie	Alte Taxonomie	Neue Kompetenzen
0,6	'Drohntechnik', 'Betrieung', 'Angebotsmanagement', 'Training'	'Betrieung', 'Angebotsmanagement', 'Drohnführerschein (Kenntnisnachweis gemäß § 21a Abs. 4 LuftVO)', 'Training'	'Drohntechnik'
0,67	'Vertretung', 'Gruppenarbeit', 'Gruppen-, Teamleitung', 'Teamentwicklung', 'Case Tools anwenden, bereitstellen'	'Vertretung', 'Gruppenarbeit', 'Gruppen-, Teamleitung', 'Organisationsberatung', 'Case Tools anwenden, bereitstellen'	'Teamentwicklung'
0,88	'Kundendienst', 'Anwenderberatung, Anwender-Support (IT)', 'Netzwerktechnik MS Internet Information Server', 'Außendienst', 'Elektroinstallation', 'Versorgungstechnik', 'Systembetreuung, Systemadministration, Systemverwaltung', 'Informatik'	'Anwenderberatung, Anwender-Support (IT)', 'Netzwerktechnik MS Internet Information Server', 'Außendienst', 'Elektroinstallation', 'Versorgungstechnik', 'Systembetreuung, Systemadministration, Systemverwaltung', 'Kundendienst'	'Informatik'
0	'Corporate Design', 'Employee Experience'	'Optionen, Futures'	'Employee Experience'
0,5	'Busse', 'Arztbriefe schreiben', 'Betrieung', 'Patientenbetreuung'	'Betrieung', 'Patientenbetreuung'	'Arztbriefe schreiben'

Tabelle 6: Beispiele Ähnlichkeitsmetrik der Ergebnisse ohne Filterung

Literatur

- [1] Baskaran R., Müller J., „Classification of German job titles in online job postings using the KldB 2010 taxonomy,“ 2023. [Online]. Available: https://www.and-effect.com/publications/2022-11-21_technical_report_kldb.pdf. [Zugriff am 19.08.2025].
- [2] Baskaran R., Müller J., „Extraktion berufsspezifischer Kompetenzen nach VerBIS aus deutschsprachigen Online-Jobanzeigen,“ 9 Juni 2022. [Online]. Available: https://and-effect.com/publications/2024-07-11_Methodenbericht_Fachkompetenzen.pdf. [Zugriff am 19.08.2025].
- [3] „Textkernel,“ 2022. [Online]. Available: <https://www.textkernel.com/de/>. [Zugriff am 19.08.2025].
- [4] „IW Köln und Bertelsmannstiftung,“ 2024. [Online]. Available: <https://huggingface.co/IW-BSt>. [Zugriff am 19.08.2025].
- [5] „Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 Band 1: Systematischer und alphabetischer Teil mit Erläuterungen.,“ 2020.
- [6] Bundesagentur für Arbeit, „Berufenet,“ 2025b. [Online]. Available: <https://web.arbeitsagentur.de/berufenet/>. [Zugriff am 19.08.2025].
- [7] Bundesagentur für Arbeit, „Download Portal Arbeitsagentur,“ 2025a. [Online]. Available: <https://download-portal.arbeitsagentur.de>. [Zugriff am 19.08.2025].