

Was kann ich eigentlich... (werden)? Detektion und Vergleich von Kompetenz-Profilen in der Digitalen Wirtschaft

Sabrina Ziebarth, Nils Malzahn, H. Ulrich Hoppe
Institut für Informatik und Angewandte Kognitionswissenschaft
Universität Duisburg-Essen
Lotharstr. 63/65
47048 Duisburg
{ziebarth, malzahn, hoppe}@collide.info

Abstract: Die digitale Wirtschaft hat aufgrund der rasanten Produkt- und damit verbundenen Technologie-Innovationen Probleme qualifiziertes Personal zu finden. Ein Teil des Problems besteht in der fehlenden Standardisierung von Berufsprofilen, die mit den genannten Innovationen einhergehen. Das bedeutet, dass es bei der Rekrutierung von Personen und der Personaleinsatzplanung Schwierigkeiten gibt, die aus der Vielzahl unterschiedlicher Benennungen für im Prinzip gleichartige Tätigkeitsprofile resultieren. Das vorliegende Papier zeigt Methoden auf, welche einerseits die Orientierung bei der beruflichen Weiterentwicklung unterstützen und andererseits die Verstetigung von Berufsprofilen in sich dynamisch entwickelnden Branchen beschleunigen sollen.

1 Einleitung

Die berufliche (Neu-) Orientierung, die Planung von beruflicher Weiterentwicklung oder anderer Lernvorhaben sowie die Vorbereitung des (Wieder-) Eintritts ins Erwerbsleben stellen nur einige Situationen dar, in denen sich die Frage nach den eigenen Kompetenzen und deren Einsatzmöglichkeiten im Berufsleben stellt. Bei der Suche nach geeigneten Stellen oder Weiterbildungszielen muss häufig die genaue Bezeichnung des eigenen bzw. angestrebten Berufs/der Stelle bekannt sein. Dies wird besonders in dem Bereich der Digitalen Wirtschaft durch einen Mangel an standardisierten Berufs- und Stellenprofilen erschwert. Unternehmen neigen dazu neue oder firmenindividuelle Berufsbezeichnungen, Positionen und Funktionen zu erfinden, so dass es zu einer Stellenbezeichnung häufig verschiedene Aufgaben- und Tätigkeitsbeschreibungen gibt [Ho09].

Der Mangel an standardisierten Profilen ist u. a. deren Kurzlebigkeit in einer sehr dynamischen Branche geschuldet. Bis Verbände und Bildungsträger sich auf Berufsprofile geeinigt haben, sind diese oft schon obsolet. Zur Unterstützung einer zeitnahen Standardisierung von aktuellen Berufsprofilen, stellen wir Methoden zur semi-automatischen Extraktion von Berufs-/Stellenprofilen aus Stellenanzeigen vor (siehe Abschnitt 4). Durch den automatischen Vergleich der in Stellenanzeigen vorgefundenen

Kompetenzprofile mit ihren, in den in Stellenanzeigen genannten Bezeichnern, kann der Prozess einer einheitlichen Benennung unterstützt und beschleunigt werden.

Durch die Zuordnung des eigenen Kompetenzprofils zu den Berufsprofilen können Bewerber die üblichen Bezeichnungen für ihr eigenes Profil identifizieren. Dies erleichtert die Schlüsselwortsuche in gängigen Stellenbörsen. Ferner besteht die Möglichkeit auf Basis des eigenen Kompetenzprofils anstelle von Berufsbezeichnungen im hier vorgestellten System nach geeigneten Stellen zu suchen.

In Abschnitt 2 folgt ein Überblick über die Themen „Kompetenz“, „Kataloge“ und „Profile“. Basierend darauf werden in Abschnitt 3 verschiedene Verfahren zum Vergleich von Kompetenz-Profilen erläutert und verglichen. Abschnitt 4 und 5 stellen Verfahren und Ergebnisse zur semi-automatischen Extraktion von Berufsprofilen aus Stellenanzeigen durch Clustering-Verfahren vor. Abschnitt 6 führt die Ergebnisse zusammen und gibt einen Ausblick auf zukünftige Arbeiten.

2 Grundlagen: Kompetenzen, Kataloge, Profile

Kompetenzprofile beinhalten typischerweise eine Liste von Kompetenzen zusammen mit einer Einschätzung deren bei einer Person vorhandenen bzw. für eine Stelle notwendigen Ausprägungen anhand einer gegebenen Skala [PAS09]. Eine einfache und daher auch häufig eingesetzte Möglichkeit zum Erstellen von Personen-Profilen besteht in der Selbstprofilierung, also der Selbsteinschätzung der eigenen Kompetenzen. Dieses Verfahren birgt die Gefahr der Über- oder Unterschätzung der eigenen Fähigkeiten. Mit Hilfe von Fremdeinschätzung durch Vorgesetzte, Kollegen und Freunde können objektivere Profile entstehen [ER03]. Werden Kompetenzprofile im Rahmen des betrieblichen Kompetenz-Managements erhoben, so basieren sie häufig auf einem Kompetenzkatalog (oder einer Kompetenzontologie), in welchem die strategisch wichtigen Kompetenzen des Unternehmens modelliert sind [KS06]. Auch bei der Profilierung zur persönlichen Dokumentation im Rahmen von lebenslangem Lernen oder zur eigenen beruflichen (Neu-)Orientierung bzw. Planung der beruflichen Weiterentwicklung, kommen vordefinierte Kompetenzkataloge zum Einsatz [Ke07]. Andere Ansätze in diesem Bereich, wie z.B. ProfilPass¹, verzichten auf einen Kompetenzkatalog und erlauben die freie Wahl von Kompetenzen.

Kompetenzen können in verschiedene Kompetenzarten unterteilt werden, z.B. in Selbst-, Sach-, Methoden- und Sozialkompetenz [Ro71]. Während Sach- oder auch Fachkompetenzen stark domänenabhängig sind, bündeln Methodenkompetenzen jene Fähigkeiten und Fertigkeiten, die es domänenübergreifend erlauben Problemlösungsstrategien zu entwickeln und durchzuführen, Sozialkompetenz beinhaltet Kompetenzen, welche in der Zusammenarbeit mit anderen benötigt werden und Selbstkompetenz die Kompetenzen, welche sich mit Selbstwahrnehmung, -darstellung und Reflexion beschäftigen [LN01].

¹ <http://www.profilpass-online.de/>

Kompetenzkataloge unterteilen Kompetenzen häufig nach Kompetenzart und ordnen sie hierarchisch entsprechend ihres Abstraktionsgrades an [KS06]. Bei der Modellierung von Kompetenzkatalogen (oder -ontologien), wie z. B. der KOWIEN²-Kompetenzontologie [ABD04] oder [Sc09], fällt auf, dass besonders fachliche aber auch methodische Kompetenzen wesentlich feiner unterteilt werden als soziale und persönliche, welche sich häufig in einer Hierarchie-Ebene befinden und nicht weiter untergliedert sind. Dies liegt insofern nahe, als dass einerseits mehr fachliche und methodische Kompetenzen in Stellenanzeigen genannt werden und diese andererseits Berufsprofile weit besser diskriminieren als die sog. „Softskills“ [ZMH09].

3 Vergleich von Kompetenz-Profilen

Beim Vergleich von Kompetenz-Profilen treten mehrere Schwierigkeiten auf. Bei freier Eingabe von Kompetenzen können verschiedene Bezeichner für eine Kompetenz angegeben werden, so bezeichnen z. B. „Teamarbeit“, „Teamwork“, „Arbeit in Teams“ und „teamfähig“ die gleiche Kompetenz. Außerdem gibt es semantische Beziehungen zwischen Kompetenzen, die berücksichtigt werden sollten. Wird in einem Profil beispielsweise die Kompetenz „SQL“ genannt und in einem anderen die Kompetenz „Datenbanken“, so gibt es zwischen diesen beiden eine Beziehung, welche nicht direkt aus den Bezeichnern erkennbar ist. Diese Informationen werden häufig in Kompetenz-ontologien modelliert, so dass es sich anbietet, diese in den Profil-Vergleich mit einzubeziehen.

3.1 Verfahren

Ein Ansatz zur Einbeziehung von Kompetenzbeziehungen ist die Erweiterung des bekannten Vektorraummodells [GF02] um Term-/ Frageerweiterung, so dass nicht nur die Kompetenzen selbst, sondern auch ihre (auf die Hierarchie bezogenen) Eltern-, Kind- und Geschwister-Kompetenzen beim Profil-Vergleich mit berücksichtigt werden. Ein anderer Ansatz basiert auf der Berechnung der semantischen Ähnlichkeiten der Profile anhand der Ähnlichkeit ihrer Kompetenzen innerhalb des Katalogs. Dabei wird die Profil-Ähnlichkeit $sim(P_1, P_2)$ als gewichteter Durchschnitt der (maximalen) Kompetenz-ähnlichkeiten $sim(c_i, c_j)$ betrachtet:

$$sim(P_1, P_2) = \frac{\sum_{c_i \in P_1} [w(c_i) * \max_{c_j \in P_2} (sim(c_i, c_j))]}{\sum_{c_i \in P_1} w(c_i)}$$

² <http://www.pim.wiwi.uni-due.de/forschung/forschungsprojekte/kowien/>

P_1 und P_2 sind die zu vergleichenden Profilen mit den in ihnen enthaltenen Kompetenzen c_i und $w(c_i)$ die durch den Nutzer vergebenen Kompetenzgewichtungen.

Es gibt verschiedene Methoden um die Ähnlichkeit von Konzepten $sim(c_i, c_j)$ in semantischen Netzen zu bestimmen. Rada et al. [Ra89] zeigen, dass der kürzeste Pfad in einer Vererbungshierarchie ein gutes Maß für den konzeptuellen Abstand ist. Resnik [Re95] schlägt die Nutzung des Informationsgehalts vor, welcher durch die Wahrscheinlichkeit des Auftretens eines Konzepts im (Profil-) Korpus definiert wird. Die Ähnlichkeit zweier Konzepte ergibt sich bei diesem Ansatz als maximaler Informationsgehalt des Elternkonzepts, welches die Konzepte subsumiert. Li et al. [LBM03] hat verschiedene Ansätze mit der WordNet³-Ontologie getestet. Die besten Ergebnisse (bezogen auf die menschliche Erwartung) erzielte eine Kombination von kürzestem Pfad ($sp(c_1, c_2)$) und der Tiefe ($d(c_1, c_2)$) des ersten gemeinsamen Elternkonzepts innerhalb der Hierarchie:

$$sim(c_1, c_2) = e^{-\alpha * sp(c_1, c_2)} * \frac{e^{\beta * d(c_1, c_2)} - e^{-\beta * d(c_1, c_2)}}{e^{\beta * d(c_1, c_2)} + e^{-\beta * d(c_1, c_2)}}; \alpha \geq 0, \beta \geq 0$$

Die stärkste Korrelation mit der menschlichen Erwartung erzielten Li et al. dabei mit $\alpha=0,2$ und $\beta=0,6$.

Zusätzlich zu den Maßen betrachten wir noch den längeren der beiden kürzesten Pfade der Konzepte zum ersten gemeinsamen Elternknoten („Höhe“). Abbildung 1 zeigt ein Beispiel für die Maße kürzester Pfad, Höhe und Tiefe. Dabei werden die Konzepte 7 und 8 verglichen, Konzept 4 ist das erste gemeinsame Elternkonzept und die Kanten, welche die Maße beeinflussen, sind hervorgehoben.

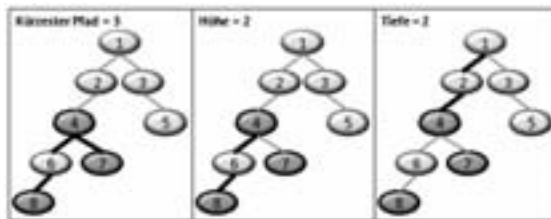


Abbildung 1: Vergleich der Maße kürzester Pfad, Höhe und Tiefe.

³ <http://wordnet.princeton.edu/>

3.2 Experimenteller Vergleich verschiedener Verfahren

Um ein möglichst gutes Ähnlichkeitsmaß zum Vergleich von Kompetenz-Profilen zu finden, wurden verschiedene Kombinationen der genannten Verfahren miteinander verglichen. In einem ersten Schritt wurden die Kombinationsmaße auf die Daten einer vorherigen Studie [Sc09] angewandt. Dabei wurde die menschliche Einschätzung der Attraktivität von Stellenanzeigen (Soll-Profilen) für bestimmte Kompetenz-Profile mit dem Ranking der Maße verglichen. Obwohl die Korrelation (Kendall- τ) allgemein gering war (Maximum 0,145), haben einige Verfahren deutlich bessere Ergebnisse gezeigt, als andere. Basierend auf diesen Ergebnissen wurden drei Verfahren für ein Experiment ausgewählt:

1. Eine Kombination aus kürzestem Pfad ($sp(c_1, c_2)$) und Höhe ($h(c_1, c_2)$):

$$sim_{SPH}(c_1, c_2) = \frac{1}{\alpha * sp(c_1, c_2) + (1 - \alpha) * h(c_1, c_2)} \quad \text{mit } \alpha = 0.6$$

2. Eine Kombination aus kürzestem Pfad, Höhe und Tiefe ($d(c_1, c_2)$):

$$sim_{SPHD}(c_1, c_2) = sim_{SPH}(c_1, c_2) * d(c_1, c_2)$$

3. Ein Vektorraummodell mit Term-Expansion auf die Kinder der Konzepte.

Während des Experiments profilierten sich 21 Probanden (Studenten und Berufseinsteiger aus dem Bereich Informatik) mit Hilfe eines Kompetenzkatalogs selbst und bewerteten im Anschluss daran 15 Vorschläge für passende Stellenanzeigen hinsichtlich ihrer Relevanz auf einer fünfstufigen Likert-Skala (mit eins als „nicht relevant“ und fünf als „sehr relevant“). Dabei stammten jeweils fünf der Vorschläge von einem der drei Maße, wobei für jedes Maß die drei am besten und die zwei am schlechtesten bewerteten Anzeigen (als Vergleichsgruppe) gewählt wurden. Die Probanden wurden während des Experiments von dem Versuchsleiter beobachtet und nach dem Experiment interviewt. Als Datenbasis für die Stellenanzeigen dienten ca. 11.000 Stellenanzeigen aus dem Bereich IT einer deutschen Online-Stellenbörse. Der Kompetenz-Katalog wurde auf Basis von 152 Stellenanzeigen der Jobbörse des BVDWs⁴ modelliert und enthielt 225 Kompetenzen, welche hierarchisch und nach Kompetenzarten sortiert angeordnet wurden.

Alle drei Maße zeigen eine mittlere bis starke Korrelation zwischen Benutzer- und Systemeinschätzung sowie eine durchgehend hohe Precision der Systemeinschätzung (siehe Tabelle 1). Während Algorithmus 1 und 3 vergleichbar gute Ergebnisse erzielen, ist die Übereinstimmung von Algorithmus 2 mit der Benutzererwartung signifikant höher.

⁴ <http://www.bvdw.org/>

Algorithmus	Korrelation	Precision
1	0,375	0,77
2	0,570	0,89
3	0,393	0,76

Tabelle 1: Korrelation zwischen Benutzer- und Systemeinschätzung sowie die Precision der Systembewertung

Algorithmus 2 bezieht verglichen mit Algorithmus 1 zusätzlich die Tiefe des ersten gemeinsamen Elternknotens ein. Auf Grund der detaillierten Modellierung der Fachkompetenzen (vgl. Abschnitt 2), liegen diese durchschnittlich „tiefer“ in der Hierarchie und fallen somit bei Einsatz des Tiefen-Maßes stärker ins Gewicht. Einerseits zeigen Beobachtungen während der Phase der Selbstprofilierung, dass insgesamt viele „Softskills“ ausgewählt werden, da die Probanden diese als wichtig empfinden. Andererseits neigen die Probanden bei der Bewertung der Relevanz der Stellenanzeigen für ihr Kompetenzprofil dazu, vor allem fachliche und methodische Kompetenzen zu berücksichtigen und andere Kompetenzarten eher zu vernachlässigen. Da das Tiefen-Maß zusammen mit den Struktureigenschaften der modellierten Ontologie fachliche und methodische Kompetenzen stärker berücksichtigt, erfüllt es die Erwartungen der Probanden besser, als die anderen Maße.

4 Semi-automatische Extraktion von Berufsprofilen

Berufs- und Stellenprofile in der IT-Branche befinden sich auf Grund technischer und organisatorischer Entwicklungen in einem ständigen Wandel. Da für neue Technologien häufig neue Fähigkeiten benötigt werden, spiegeln sich diese Entwicklungen auch in den Stellenausschreibungen der Firmen wieder. Die meisten Stellen in der IT-Branche werden online entweder auf den firmeneigenen Webseiten oder in Stellenbörsen ausgeschrieben [Di07], so dass diese eine einfach zugreifbare Ressource für Stellenausschreibungen in digitaler Form darstellen. Aus diesen Anzeigen können mit Methoden des Data Minings Informationen über Berufsprofile extrahiert werden.

4.1 Vorverarbeitung der Stellenanzeigen

Vor der Analyse der Stellenanzeigen, müssen diese zunächst so vorverarbeitet werden, dass sie als Eingabeformat geeignet sind. Die Schritte der Vorverarbeitung sind in Abbildung 2 dargestellt.



Abbildung 2: Vorverarbeitungsschritte

Die aus Online-Jobbörsen erhobenen Stellenanzeigen liegen zunächst in Form von HTML-Dateien vor, aus denen der eigentliche Text mit Hilfe eines HTML-Parsers extrahiert wird; Bilder und Stilinformationen werden nicht weiter betrachtet. Die Texte werden dann in einzelne Terme zerlegt (engl. „Tokenization“), so dass jedes Dokument durch eine Menge von Termen (engl. „Bag of Words“) beschrieben wird. Worte können jedoch durch Flexion, Derivation oder Komposition in verschiedenen morphologischen Varianten auftreten, so dass es beispielsweise in der deutschen Sprache zu jedem Wort durchschnittlich zehn Flexionsformen gibt [Ha98]. Die Verschmelzung der verschiedenen Wortformen zu einem Repräsentanten kann die Qualität von Information Retrieval daher signifikant erhöhen [Fr84] und wird auch im Data Mining häufig verwendet. Zur Verschmelzung von Wortformen gibt es zwei Ansätze: Linguistische Verfahren führen Wortformen auf ihre Grundform (Lexem) zurück (Grundformreduktion, Lemmatisierung), während nicht-linguistische Verfahren Wortformen über Regeln auf ihren Wortstamm reduzieren (Stammformreduktion, engl. „Stemming“) [GMS05]. Da die Verwendung einfacher Stemming-Verfahren zu mit komplexeren linguistischen Ansätzen vergleichbaren Ergebnissen führt [BR04], nutzen wir eine deutsche Variante⁵ des Porter-Stemmers [Po80] zur Stammformreduktion. Viele der in Stellenanzeigen vorkommenden Worte beschreiben keine Kompetenzen, so dass diese gefiltert werden müssen. Durch Anwendung genereller, sprachabhängiger Stoppwortlisten⁶ können inhaltsarme bis -freie Worte (wie z. B. Artikel, Konjunktionen, Präpositionen und Partikel) gefiltert werden. Dies reicht aber nicht aus, da viele Informationen in den Stellenanzeigen (wie z. B. Namen und Orte) zwar inhaltstragend, für die Kompetenzprofile aber irrelevant sind. Daher werden externe Wissensquellen mit einbezogen. Wort-Kategorisierungen durch das „Wortschatz“⁷-Wörterbuch (per Web Service-Anfrage) können dazu genutzt werden um nach Kategorien (z. B. Stadt, Ort, Land, Vorname, Nachname) zu filtern. Die Ontologie deutscher Worte GermaNet⁸ ordnet Worte in semantische Felder, welche auch zur Filterung genutzt werden können (es wurden beispielsweise die semantischen Felder Form, Körper, Menge, Nahrung, Ort, Pflanze, Substanz, Tier und Zeit als domänenabhängige Stoppworte behandelt). Die Beziehungen zwischen Dokumenten und Termen werden in einer Term-Dokument-Matrix festgehalten, welche das Gewicht jedes Terms für jedes Dokument beinhaltet. Die Term-Gewichte ergeben sich durch Methoden der Textstatistik, wie der relativen Termhäufigkeit (engl. „term frequency“, „tf“) [Sa68], welche häufig vorkommende Terme höher gewichtet oder der inversen Dokumentenhäufigkeit (engl. „inverse document frequency“, „idf“), welche berücksichtigt, dass Terme, die in vielen Dokumenten auftreten, diese weniger gut diskriminieren, als Terme, welche nur in wenigen Dokumenten auftreten. Häufig wird eine Kombination dieser Maße („tf-idf“) verwendet [WF05].

Unser Datensatz umfasst ca. 3000 Stellenanzeigen eines deutschen Jobportals aus dem Bereich der IT-Branche.

⁵ <http://snowball.tartarus.org/texts/germanic.html>

⁶ z.B.: <http://www.ranks.nl/stopwords/german.html>

⁷ <http://wortschatz.uni-leipzig.de/>

⁸ <http://www.sfs.uni-tuebingen.de/GermaNet/>

4.2 Identifikation von Berufsprofilen durch Clustern von Stellenanzeigen

Cluster-Verfahren werden im Data Mining dazu eingesetzt, um Datensätze (z. B. Stellenanzeigen) in natürliche Gruppen ähnlicher Beispiele zu zerlegen. Die Zentroide von Stellenanzeigen-Clustern können als Vorlagen für Stellen- bzw. Berufsprofile betrachtet werden, da sie im Mittel die Eigenschaften der Profile im Cluster wiedergeben. Da das Clustering von großen Datenmengen mit höherem Rechenaufwand verbunden ist werden nur Auszüge des gesamten Datensatzes, sog. Sample geclustert. Für diese Arbeit wurden verschiedene Sample bestehend aus je 200 Stellenanzeigen mit verschiedenen Algorithmen (k-Means, k-Medoids, FarthestFirst [Ho85]) und verschiedenen Parametern (k=5 und k=6) geclustert. Es entstanden 310 Cluster. Um aus diesen die „besten“ Cluster zu bestimmen, wurden die 310 Cluster als neuer Eingabe-Datensatz betrachtet und erneut geclustert. Zu diesem Zweck wurden die Zentroide der jeweiligen Cluster aus den Durchschnittswerten der TF-IDF-Gewichte der Terme im Cluster gebildet. Das Clustering der Zentroiden erfolgte anschließend mit dem X-Means-Algorithmus, welcher einen erweiterten k-Means-Algorithmus darstellt, der unter anderem die beste Anzahl von Clustern in einem gegebenen Bereich bestimmen kann [PM00]. Das Clustern von Clustern ist ein verbreiteter Ansatz für Mustererkennungs-Probleme [CC92]. Durch das Clustern der Ergebnisse verschiedener Clustering-Algorithmen kann die Robustheit und Qualität des endgültigen Clusterings signifikant verbessert werden [GMT07].

Das Ergebnis (vgl. Tabelle 2) zeigt vier große Cluster, welche sich bestimmten Stellenprofilen zuordnen lassen: Stellenanzeigen in Cluster 2 (SAP-Berater) sind durch Anforderungen im Bereich SAP-Beratung und Management, so wie Erfahrung mit (internationalen) Projekten geprägt, während Stellenanzeigen in Cluster 3 (Entwickler) Kompetenzen wie Java, Engineering, Design und Datenbanken fordern. Cluster 0 bündelt Kompetenzen für IT-Manager und -Berater, wie Erfahrung mit Projekten, Kunden und Prozessen, so wie spezielle Kenntnisse im Bereich von Entwicklung, Systemen und Datenbanken. In Cluster 1 zeigen sich vor allem Kompetenzen im kaufmännischen und im Support-Bereich.

Abbildung 3 zeigt die vier Stellenprofile bestehend aus den Kompetenzen in den Cluster-Zentroiden. Dabei zeigen sich einige Überschneidungen in den Kompetenzen, welche darauf hindeuten, dass es sich bei diesen um Kompetenzen handelt, die offensichtlich Schnittstellen zu anderen Profilen darstellen.

Cluster 0	Cluster 1	Cluster 2	Cluster 3
50 Anzeigen	82 Anzeigen	42 Anzeigen	136 Anzeigen
senior: 0.29 it: 0.17 management: 0.17 security: 0.14 berat: 0.12 consultant: 0.08 betrieb: 0.07 projekt: 0.07 tätigkeitsfeld: 0.06 servic: 0.06 kund: 0.06 prozess: 0.06 system: 0.05 datenbank: 0.05 entwickl: 0.05 selbststand: 0.05 business: 0.05 mehrjahr_berufserfahr : 0.05 berufserfahr: 0.04 fachlich: 0.04	kaufmann: 0.12 it: 0.1 servic: 0.1 dienstleist: 0.08 support: 0.07 administration: 0.07 belastbar: 0.07 abgeschloss_ausbild: 0.07 office: 0.07 technisch: 0.07 personalvermittl: 0.07 installation: 0.06 teamfah: 0.06 personal: 0.06 projektarbeit: 0.06 hardwar: 0.06 ausbild: 0.05 berufserfahr: 0.05 serv: 0.05 windows: 0.05	sap: 0.55 sd: 0.12 logist: 0.09 consultant: 0.08 manag: 0.07 berat: 0.07 international: 0.06 management: 0.06 projekt: 0.05 system: 0.05 bw: 0.05 kaufmann: 0.05 fachlich: 0.05 engagement: 0.05 dienstleist: 0.04 kund: 0.04 kontinui: 0.04 it: 0.04 weiterentwickl: 0.04 sich: 0.04	entwickl: 0.12 java: 0.1 softwar: 0.1 softwareentwickl: 0.08 engineering: 0.07 business: 0.06 reporting: 0.06 it: 0.05 technologi: 0.05 management: 0.05 informat: 0.05 design: 0.04 onlin: 0.04 kund: 0.04 oracl: 0.04 team: 0.04 serv: 0.04 anwendungsentwickl: 0.04 unix: 0.04 projektleit: 0.04
IT-Manager	IT-Kaufmann	SAP-Berater	Entwickler

Tabelle 2: Ergebnisse der Cluster-Clustering mit X-Means. Die Terme geben die Stems der Terme wieder und die Zahlen geben den durchschnittlichen TF-IDF-Wert der Terme im Cluster an.

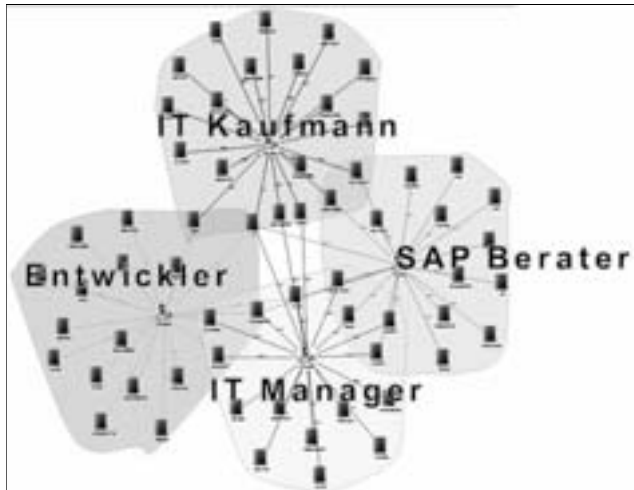


Abbildung 3: Stellenprofile

5 Diskussion

Die in Abschnitt 4 beschriebenen Ergebnisse zeigen, dass die maschinelle Auswertung von Stellenanzeigen ein geeigneter Ansatz zur Bestimmung von Berufsprofilen in der digitalen Wirtschaft ist. Die in Abbildung 3 gezeigten Bezeichnungen für Cluster zeigen zwar noch keine sehr spezialisierten Berufsprofile, jedoch ist es möglich diese noch weiter aufzuschlüsseln. In jedem Fall wird deutlich, dass die Verfahren nutzbar sind gerade weil die ermittelten Profile sofort mit Namen belegt werden konnten.

Von besonderem Interesse sind die Überlappungsbereiche. Entlang dieser Bereiche lassen sich fallweise neue Berufsprofile entwickeln, da in Interviews mit Branchenvertretern geäußert wurde, dass im Moment zwar keine passgenau qualifizierten Mitarbeiter gefunden werden könnten, jedoch gerne Personen mit Qualifikationen in benachbarten Bereichen eingestellt würden, in der Hoffnung, dass sie sich die benötigten Kompetenzen aus anderen Bereichen aneignen könnten. Dies gilt insbesondere bei Profilen, für die bevorzugt Akademiker oder Studienabbrecher eingestellt werden. Die Überlappungsbereiche sind also Kandidaten bei der Identifikation neuer Berufsprofile, da hier offensichtlich getrennte Profile zusammenwachsen. Die gewonnenen Erkenntnisse können dann z. B. von Verbänden wie dem BVDW aufgegriffen werden, um mittel- bis langfristig neuartige Ausbildungsprozesse zu initiieren und kurzfristig Weiterbildungsangebote für seine Mitglieder anzubieten.

		Anzahl Kompetenzen, die Voraussetzung sind:	
		Niedrig	Hoch
Anzahl der Kompetenzen für die die gewählte Kompetenz Voraussetzung ist	hoch	<i>Enabler-Kompetenz</i> Ist Bedingung für viele andere Kompetenzen, benötigt selbst jedoch wenig andere Kompetenzen (auch Vorwissen). Eröffnet neue Geschäftsbereiche	<i>Generalistenkompetenz</i> Benötigt viele verschiedene Kenntnisbereiche; kann sich in viele weitere Bereiche weiterentwickeln.
	niedrig	<i>Inselkompetenz</i> Ist Bedingung für wenig andere Kompetenzen und benötigt wenig Vorwissen. Kurzfristige, projektbezogene Qualifizierung	<i>Qualifizierungskompetenz</i> Braucht viele andere Kompetenzen als Voraussetzung; wird nicht von vielen anderen benötigt. Expertenkompetenz

Abbildung 4: Kompetenzbewertung nach Angebot und Nachfrage [Gü07]

Für die eingangs erwähnte Gruppe der Wiedereinsteiger und Weiterbildungsinteressierten sind diese Informationen ebenfalls von hohem Interesse, weil sie sich entscheiden müssen, welche Kompetenzen sie erwerben sollten. Die Informationen aus der Kompetenz-Ontologie lassen eine Bewertung nach dem in Abbildung 4 gezeigten Schema zu.

Zusammen mit der in Abschnitt 3 beschriebenen Methode zum Kompetenzprofilvergleich kann so eine informierte Entscheidung für den nächsten Weiterbildungsschritt unterstützt werden. Dabei werden die Berufsprofile als Sollprofile aufgefasst und das momentane persönliche Kompetenzprofil als Vergleich herangezogen. Es können dann in Fallsimulationen Kompetenzen zum aktuellen Profil hinzugefügt werden, unter Beachtung der in Abbildung 4 vorgeschlagenen Bewertung, um den Abstand zum gewünschten Berufsprofil zu verringern.

6 Zusammenfassung und Ausblick

Ausgehend von der Situation, dass Produkt- und damit verbundene Technologie-Innovationen der Digitalen Wirtschaft Probleme bereiten, qualifiziertes Personal zu finden, werden in diesem Papier verschiedene Unterstützungsverfahren für die Kompetenzentwicklung präsentiert. Dabei wird sowohl versucht Berufsprofile zeitnah herauszuarbeiten, um die Branche bei der Umstellung ihrer Qualifizierungsangebote sowohl kurz- als auch langfristig zu unterstützen, als auch Informationen für die individuelle Kompetenzentwicklungsplanung von vorqualifiziertem Fachpersonal zur Verfügung zu stellen. Weitere Arbeiten werden speziell an diesen Punkt anknüpfen. Einerseits sollen Was-Wenn-Simulationen zur Kompetenzentwicklung möglichst komfortabel möglich sein und andererseits wird versucht werden, die im wachsenden befindlichen Berufsprofile möglichst zielsicher und früh zu detektieren, um der Branche Gelegenheit zu geben ihre Qualifizierungsangebote rechtzeitig anzupassen, damit qualifiziertes Personal zur Verfügung steht und Innovationsbarrieren durch dessen Fehlen verringert werden.

Literaturverzeichnis

- [ABD04] Apke, S.; Bremer, A.; Dittmann, L.: Konstruktion einer Kompetenz-Ontologie, dargestellt am Beispiel der Deutschen Montan Technologie GmbH (DMT), KOWIEN-Projektbericht 06-2004, Institut für Produktion und Industrielles Informationsmanagement, 2004.
- [BR04] Braschler, M.; Ripplinger, B.: How Effective is Stemming and Decompounding for German Text Retrieval? *Information Retrieval*, 7, 2004; pp. 291-316.
- [CC92] Chan, K. P.; Cheung, Y. S.: Clustering of Clusters. *Pattern Recognition*, 25(2), 1992; pp. 211-217.
- [Di07] Dincher, R.: Personalmarketing und Personalbeschaffung. Forschungsstelle für Betriebsführung und Personalmanagement e. V., 2007.
- [ER03] Erpenbeck J.; von Rosenstiel, L.: Handbuch Kompetenzmessung, Schäffer Poeschel Verlag, Stuttgart, 2003.
- [Fr84] Frakes, W. B.: Term conflation for information retrieval. *Proceedings of the 7th annual international ACM SIGIT conference on research and development in information retrieval*, 1984; pp. 383-389.
- [GMS05] Galvez, G.; de Moya-Anegon, F.; Solana, V. H.: Term conflation methods in information retrieval: Non linguistic and linguistic approaches. *Journal of Documentation*, 61(4), 2005; pp. 520-547.

- [GMT07] Gionis, A.; Mannila, H.; Tsaparas P.: Clustering Aggregation. *ACM Transactions on Knowledge Discovery from Data*. 1(1), 2007.
- [GF02] Grossmann, D. A.; Frieder, O.: *Information Retrieval – Algorithms and Heuristics*, Kluwer Academic Publishers, 2002.
- [Gü07] Günther, A. et al.: Vermittlung und Entwicklung von Kompetenzen. *Jenseits von Virtualität - Arbeiten und Lernen in Projektnetzwerken*, Lohmar, 2007; S. 115-166.
- [Ha98] Hausser, R.: Drei prinzipielle Methoden der automatischen Wortformerkennung. *Sprache und Datenverarbeitung*, 22, 1998; S. 38-57.
- [Ho85] Hochbaum, S.: A best possible heuristic for the k-center problem. In *Mathematics of Operations Research*, 10(2), 1985; pp. 180-184.
- [Ho09] Hofert, S.: *Stellensuche und Bewerbung im Internet*, Humboldt, 2009.
- [Ke07] Kew, C.: *The TENCompetence Personal Competence Manager*, Proceedings of the EC-TEL 07, Crete, Greece, 2007.
- [KS06] Kunzmann, C.; Schmidt, A.: *Ontology-based Competence Management for Healthcare Training Planning – A Case Study*, Proceedings of I-KNOW 06, Graz, Austria, 2006.
- [LN01] Lehmann, G.; Nieke, W.: *Zum Kompetenz-Modell*, <http://www.bildungsserver-mv.de/download/material/text-lehmann-nieke.pdf>, 2001 (letzter Zugriff am 07.05.10)
- [LBM03] Li, Y.; Bandar, Z. A.; McLean, D.: An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 2003.
- [PAS09] PAS 1093: *Personalentwicklung unter besonderer Berücksichtigung von Aus- und Weiterbildung – Kompetenzmodellierung in der Personalentwicklung*, Beuth Verlag GmbH, Berlin, 2009.
- [PM00] Pelleg, D.; Moore, A. W.: X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In: *Seventeenth International Conference on Machine Learning*, 2000; pp. 727-734.
- [Po80] Porter, M. F.: An algorithm for suffix stripping. *Program*, 14(3), 1980; pp.130-137.
- [Ra89] Rada, R. et al.: Development and Application of a Metric on Semantic Nets, *IEEE Transactions on Systems, Man and Cybernetics*, 19(1), 1989.
- [Re95] Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy, *Proceedings of the 14th International Joint Conference for Artificial Intelligence*, 1995.
- [Ro71] Roth, H.: *Pädagogische Anthropologie, Bd.2: Entwicklung und Erziehung. Grundlagen einer Entwicklungspädagogik*, Schroedel, Hannover, 1971.
- [Sa68] Salton, G.: *Automatic Information Organization and Retrieval*. Mc Graw-Hill, 1968.
- [Sc09] Schröder, S. et al.: *Self-Profiling of Competences for the Digital Media Industry: An Exploratory Study*. Proceedings of EC-TEL 09, Nice, France, 2009
- [WF05] Witten, I. H.; Frank, E.: *Data mining – practical machine learning tools and techniques*. The Morgan Kaufman series in data management systems, 2005.
- [ZMH09] Ziebarth, S.; Malzahn, N.; Hoppe, H. U.: *Using Data Mining Techniques to Support the Creation of Competence Ontologies*, Proceedings of AIED 09, Brighton, England, 2009.