



BEUTH HOCHSCHULE FÜR TECHNIK BERLIN
University of Applied Sciences

Beuth Hochschule für Technik

Fachbereich 6

Bachelorarbeit

im Studiengang Medieninformatik - Schwerpunkt Medieninformatik

zur Erlangung des akademischen Grades
Bachelor of Science

Thema: Entwicklung eines Empfehlungssystems für Kompetenzen

Autor: Valentin Risch s55698@beuth-hochschule.de
MatNr. 798906

Version vom: 10. August 2017

1. Betreuer: Prof. Dr. Johnanes Konert

2. Gutachter: Prof. Dr. Heike Ripphausen-Lipa

Inhaltsverzeichnis

Abbildungsverzeichnis	4
Listingverzeichnis	5
Abkürzungsverzeichnis	6
1 Einleitung	8
1.1 Abstrakt	8
1.2 Inhalte dieser Arbeit	9
2 Grundlagen und Begrifflichkeiten	10
2.1 Kompetenzen	10
2.2 Kompetenzframeworks	10
2.3 RDF, Semantic Web Linked Data	11
2.3.1 Semantic Web	11
2.3.2 Linked Data	11
2.3.3 RDF	12
2.3.4 Ontologie/Vokabular	12
2.4 Graphen	12
2.4.1 Graphentheorie	12
2.4.2 Graphendatenbanken	13
2.4.3 Neo4J	13
2.5 Empfehlungssysteme	14
3 Verwandte Themen und Arbeiten	15
3.1 Openbadge	15
3.2 Europäischer Qualifikationsrahmen	15
3.2.1 Deutscher Qualifikationsrahmen	16
3.3 European e-Competence Framework	16
3.4 InLoc	16
3.5 ESCO	17
3.5.1 EURES	17
3.6 RDF Import Neo4J	18
3.7 Recommender Systeme	18
3.8 Movielense	18
3.8.1 RecSys Challenge	19
3.8.2 Job-Matching: Xing	19
4 Analyse und Aufgabenstellung	20
4.1 Aufgabenstellung	20
4.2 Anforderungsanalyse	20
4.2.1 Kompetenzen	20
4.2.2 Anforderungen an die Engine	21
4.2.3 Anforderungen an die API	21
4.2.4 Anforderungen an die Architektur	22
4.3 Algorithmen und Metriken	22

5 Entwurf	24
5.1 Format/Kompetenzmodell	24
5.1.1 ESCO als RDF	24
5.1.2 inLOC	25
5.1.3 Fazit: Auswahl eines Modells	26
5.1.4 Datenbank	26
5.2 Datenmodell	27
5.2.1 API	27
5.3 Anfragen Formulieren	28
6 Implementierung	29
6.1 Architektur	29
6.2 EC2	29
6.3 Neo4j	30
6.3.1 RDF Import	30
6.4 API Gateway	31
6.5 Serverless	31
6.5.1 AWS Lambda	31
6.5.2 Serverless Framework	31
6.6 Architektur im Überblick	32
6.7 Deployment mit dem Serverless Framework	34
6.8 Algorithmen und Metriken	34
6.8.1 Shortest Path	34
6.8.2 Jaccard Index	35
6.8.3 Simrank	36
6.8.4 Levenstein oder Edit-Distanz	37
7 Bewertung	38
8 Ausblick	40
9 Fazit	41
Literaturverzeichnis	43
Anhang	45
9.0.1 API Endpunkte	45

Abbildungsverzeichnis

1	Beschreibung	18
2	Beschreibung	19
3	Beschreibung	22
4	Beschreibung	25
5	Beschreibung	27
6	Beschreibung	28
7	Beschreibung	30
8	Beschreibung	32
9	Beschreibung	39

Listingverzeichnis

1	Das Listing zeigt einen Funktionsaufruf über die Neo4j	18
2	importRDF Prozeduraufruf	30
3	Lambda Handler Funktion	33
4	Serverless Konfiguration	34
5	Response eines GET Requests an competence-service-prd-getJaccardOfCompetence	38

Abkürzungsverzeichnis

LOD Linked Open Data

1 Einleitung

1.1 Abstrakt

Mit dem Abschluss einer jeden Berufsausbildung oder eines Studiums beginnt auch der Einstieg in das Berufsleben. Für Viele Bewerber bedeutet das, in Frage kommende Jobangebote zu sichten, und vergleichen. Es folgt das schreiben unzähliger individueller Bewerbungen, von denen im schlimmsten Fall nur wenige zu einem Vorstellungsgespräch führen. Doch oftmals kommen schon bereits nach den ersten Wochen des Hochgefühls endlich eine Stelle zu haben, erste Zweifel, ob der Job der richtige für einen ist, und die eigenen Wünsche und Anforderungen erfüllen kann.

Auch für Personaler stellt die Einstellung von neuem Personal eine schwierige Aufgabe dar. So müssen oft aus einer Reihe von Bewerbern die tatsächlich unqualifizierten aussortiert werden und eine Auswahl getroffen werden, welche Bewerber zu einem Gespräch eingeladen werden oder in die nächste Bewerbungsrunde gelangen.

Es besteht also Bedarf den Prozess von Arbeitsvermittlungen zu optimieren und zu vereinfachen. Das spart sowohl Bewerbern als auch Arbeitgebern viel Frust und Zeit. Ein automatischer Abgleich von Kompetenzen kann diesen Prozess beschleunigen, setzt aber einen einheitlichen Rahmen für die Beschreibung von Kompetenzen voraus.

Das Kernproblem bei der Arbeitsvermittlung besteht darin, eine vakante Stelle mit einem geeigneten Bewerber zu besetzen. Auf der einen Seite steht also der Bewerber der Kompetenzen anbietet, auf der anderen Seite der Arbeitnehmer, welcher gewisse Fertigkeiten und Qualifikationen fordert. Das technische Problem welches hier umgesetzt werden muss ist, beide Seiten zusammen zu bringen. Der naive Ansatz wäre einfach die Liste der Kompetenzen des Stellenangebotes mit der Liste aus dem Lebenslauf des Bewerbers zu vergleichen und die Passgenauigkeit zu berechnen. Das führt aber unter Umständen dazu, dass interessante Bewerber aussortiert werden weil eventuell andere Bezeichnungen für ihre Fertigkeiten angegeben haben.

Diese Problem wirft die Frage auf welche Standards und Richtlinien gibt es, um Kompetenzen abzubilden und zu modellieren und ob Quellen existieren um Ein Computer basiertes System zu entwickeln, welches dies Kompetenzen mit einander vergleicht und Ähnlichkeiten findet.

Für diese Vergleiche soll ein semantischer Ansatz gefunden werden, der nicht auf Textvergleiche zurückgreift, sondern vielmehr versucht Kompetenzen als Graph abzubilden. Diese Arbeit soll sich mit Graphen Basierten Technologien und Ansätzen auseinander setzen, um Kompetenzen zu modellieren und Ähnlichkeiten zu finden.

1.2 Inhalte dieser Arbeit

Der Begriff der *Kompetenz* wird in den Domänen Beruf und Bildung häufig verwendet, oft aber mit unterschiedlicher Definition und Bedeutung. So taucht der Begriff auch häufig in Verbindung mit anderen Substantiven auf. Beispiele sind: Fachkompetenz, Kernkompetenz, Handlungskompetenz, Teilkompetenz, Selbst- und Sozialkompetenz. Um eine Abgrenzung zu schaffen soll zunächst der Rahmen für den Kompetenzbegriff in dieser Arbeit gespannt werden.

Ein großes wissenschaftliches Interesse besteht im Bereich sogenannter *Recommender Systeme* oder auch Empfehlungsdienste. Die *Association for Computing Machinery*[acm], ist eine wissenschaftliche Gesellschaft, welche sich mit der Forschung im Bereich Informatik und Computer beschäftigt. Sie hält jährlich eine eigene Konferenz[rec] zum Thema Empfehlungsdienste ab auf der, neuste wissenschaftliche Ergebnisse diskutiert werden. Ein solcher Empfehlungsdienst soll für die Domäne von Berufs Kompetenzen im Rahmen dieser Arbeit entworfen werden.

Die Arbeit soll die Möglichkeit evaluieren, ob Kompetenzen als Graph modelliert werden können, es also entweder eine hierarchische Baumstruktur gibt, oder die Daten miteinander verknüpft sind. Darum soll eine Technologie ausgewählt um Kompetenzen in einer Graphen Struktur zu speichern, und mit Hilfe dieses Graphen Methoden oder Algorithmen gefunden werden, die für diesen Anwendungsfall geeignet sind, um ein Empfehlungssystem zu implementieren.

2 Grundlagen und Begrifflichkeiten

2.1 Kompetenzen

Im Allgemeinen spricht man bei der Verbindung von Wissen und Können um geforderte Handlungen zu bewältigen von Kompetenzen. Dabei liegt der Fokus auf Fähigkeiten und Wissen, welche dazu beitragen Probleme mit nicht standardmäßigen Handeln zu lösen, und auf verschiedene Situationen zu übertragen.¹Oftmals werden die beiden Begriffe Kompetenz und Qualifikation als Synonyme verwendet, jedoch gibt es signifikante Unterschiede in deren Bedeutung.

Kompetenz²(lateinisch: competere, zu etwas fähig sein) meint Lernerfolg im Hinblick auf den Lernenden selbst und seine Befähigung zu selbstverantwortlichem Handeln im privaten, beruflichen und gesellschaftlichen Bereich. Sie bezeichnet die subjektive Leistungsfähigkeit einer Person, welche nicht überprüfbar und objektiv Bewertbar ist.

Qualifikationen³ (lateinisch: qualis facere, Beschaffenheit herstellen) sind hingegen prüfbar und Zertifizierbar. Sie sind die äußere Seite der Leistungsanforderung und auf die Erfüllung vorgegebener Zwecke gerichtet.

2.2 Kompetenzframeworks

Kompetenzen abzubilden gestaltet sich schwierig, da diese einen dynamischen Charakter hat um vom Kontext, ihrer Domäne abhängig ist. Durch die Unterteilung in Branchen und Sektoren ist eine Klassifizierung von Kompetenzen dann möglich. Eine weitere Dimension zur Klassifikation ist das Niveau, es beschreibt das Maß oder auch den *Schwierigkeitsgrad* einer Handlung die mit der gegebenen Kompetenz zu bewältigen ist.

Da Abschlüsse keine Kompetenz bescheinigen, werden Qualifikationsrahmen benötigt, mit deren Hilfe die in einem Bildungsweg erworbenen Kenntnisse und Fähigkeiten zu Kompetenzentwicklung beitragen können. Mit diesen Lernergebnissen die das *Können* im Sinne von *in der Lage sein, etwas zu tun* beschreiben, ergeben sich dann Vergleichsmöglichkeiten.

Um Qualifikationen und Kompetenzen aus unterschiedlichen sektoralen oder nationalen Rahmenwerken zu vergleichen benötigt es einen Meta-Rahmen.

¹vgl. Bundesinstitut für Berufsbildung[bib17]

²vgl. Bauer 2015[BE15]

³vgl. Bauer 2015[BE15]

2.3 RDF, Semantic Web Linked Data

2.3.1 Semantic Web

Das Semantic Web stellt Daten, im World Wide Web in einem für Maschinen verarbeitbaren Art und Weise zur Verfügung mit dem Ziel der Interoperabilität, also der Möglichkeit Informationen zwischen Anwendungen und Plattformen auszutauschen und diese in Beziehung zu setzen. Kernaspekte sind das Auffinden relevanter Informationen, die Integration von Informationen aus verschiedenen Quellen und automatische Schlussfolgerung. Die Ziele des Semantic Web lassen sich grob zusammen fassen als Absicht Methoden und Wege zu finden um "Informationen so zu repräsentieren, dass Maschinen damit in einer Art und Weise umgehen können, die aus menschlicher Sicht nützlich und sinnvoll erscheint."⁴

Im folgenden sollen einige Konzepte und Technologien erläutert werden welche die Anforderungen des Semantic Webs umsetzen.

2.3.2 Linked Data

Die Beziehungen der Daten im Semantic Web müssen nicht auf ein Datensatz beschränkt bleiben. So können Daten auch Querverweise auf andere Datensätze haben. Ein Beispiel sind die Daten von Wikipedia die durch das Linked Dataset DBPedia zugänglich sind. Einzelne Einträge beinhalten dabei Verweise auf das Geonames Dataset. Einige dieser Datensätze sind öffentlich zugänglich (Linked Open Data). So stellt das Amt für Veröffentlichungen der Europäischen Union mit ihrem offenen Datenportal Daten ihrer Institutionen und anderer Einrichtungen zur Verfügung.

Die Vorgehensweise bei der Erstellung von Linked Data wird von Tim Berners-Lee⁵ in folgenden 4 Schritten beschrieben:

1. Dinge und Objekte werden durch URIs identifiziert
2. Die URIs sind über das HTTP Protokoll aufrufbar
3. Beim Aufruf werden relevante Informationen in standardisierten Formaten geliefert
4. Die gelieferten Daten enthalten Referenzen auf andere URIs

⁴vgl. Hitzler 2007[HKRS07, 12]

⁵vgl. Bernes-Lee[ld]

2.3.3 RDF

Um die Anforderung des Semantic Web's, Daten im Web auszutauschen, zu erfüllen, wurde das Resource Description Framework RDF als eine formale Sprache für die Beschreibung strukturierter Informationen geschaffen. Dabei soll die ursprüngliche Bedeutung erhalten bleiben und Kombinationen und Weiterbearbeitung der enthaltenen Informationen ermöglicht werden. Eine Resource kann generell jedes Objekt mit einer eindeutigen Identität sein, wie zb. Bücher, Orte, Menschen oder abstrakte Konzepte. Um Mehrdeutigkeiten zu verhindern werden URIs als Bezeichner verwendet. Diese RDF-Beschreibungen können auch durch Zeichenketten syntaktisch dargestellt werden, müssen vorher jedoch in Bestandteile zerlegt und serialisiert werden. Die dabei entstehenden Dokumente sind gerichtete Graphen, wobei Knoten und Kanten mit eindeutigen Bezeichnern beschriftet sind.

Triple RDF-Graphen lassen sich vollständig durch ihre Kanten beschreiben. Eine solche Kante hat einen Anfangspunkt, eine Beschriftung und einen Endpunkt. Dieses Triple wird bestimmt durch *Subjekt-Prädikat-Objekt*.

2.3.4 Ontologie/Vokabular

Als Ontologie oder Vokabular wird eine Sammlung von Begriffen bezeichnet, die innerhalb einer Domäne *Wissen* abbilden und klassifiziert werden.⁶ Mit Hilfe eines Vokabulars werden Regeln über die Semantik von Klassen, Attributen und Beziehungen von Daten definiert. Ontologien können auch in einer spezifischeren Domäne erweitert werden. In der *Simple Knowledge Organization System* Ontologie wird durch die *Concept* Klasse eine Idee oder ein Sachverhalt abgebildet und mit anderen verknüpft. Auch Vererbung und hierarchische Strukturen sind möglich.

2.4 Graphen

Ein Graph ist definiert als eine Menge von Knoten (Vertices) und deren Beziehungen, welche über Kanten (Edges) dargestellt werden. Mit Graphen kann man vernetzte Strukturen wie zb. Straßennetze, Computernetzwerke oder Datenstrukturen modellieren. So finden sich in vielen modernen Technologien wie Routenplaner oder Social Media Anwendungen graphentheoretische Konzepte wieder.

2.4.1 Graphentheorie

Die Graphentheorie ist ein Teilgebiet der Mathematik, in dem Graphen und ihre Beziehungen zueinander untersucht. Das älteste dokumentierte Problem der Graphentheorie

⁶What is a Vocabulary? [w3c]

ist das Königsberger Brückenproblem. Dabei wurde ein Rundweg durch Königsberg gesucht, der alle Brücken jedoch nur einmal überquerte. Leonhard Euler erkannte 1736, dass man die einzelnen Ufer als Punkte und Brücken als Kanten abstrahieren konnte. Euler zeigte, dass ein solcher Weg nicht existierte, da jeder Knoten mit einer ungeraden Anzahl von ungerichteten Kanten verbunden sein muss.

Da dieses Themengebiet sehr umfassend ist sollen hier nur einige für diese Arbeit relevante Konzepte erwähnt werden.

Gerichteter Graph:

Ein gerichteter Graph ist definiert als $G = (V, R, \alpha, \omega)$ mit V als nicht leere Menge von Knoten, R als Kantenmenge. α und ω sind jeweils Abbildungen für die gilt $r(\alpha, \omega)$, wobei r eine Kante ist die ihren Ursprung im Anfangsknoten α hat und im Endknoten ω endet. [KN12, 7]

Adjazenzmatrix:

Die Adjazenzmatrix oder auch Nachbarschaftsmatrix ist eine $n \times n$ Matrix mit $n = |V|$. Sie gibt welche Knoten im Graphen miteinander verbunden sind. Für die Adjazenzmatrix $A(G)$ gilt $a_{ij} = |\{ r \in R : \alpha(r) = v_i \text{ und } \omega(r) = v_j \}|$. [KN12, 17ff]

Gewichtete Kanten:

Eine Kante kann bewertet werden. Sie wird mit einer Zahl notiert. Diese Zahl kann dann als Parameter verwendet werden und gibt die Kosten an, die anfallen wenn die Kante von einem Algorithmus oder einer Funktion passiert wird. [KN12, 18]

2.4.2 Graphendatenbanken

Im Gegensatz zu relationalen Datenbanken werden Daten nicht in Tupeln als Tabellen gespeichert sondern als Knoten in einem Graphenmodell. Der Vorteil darin liegt zunächst in der Daten Modellierung, da das Graphen Modell sehr intuitive ist und sich einfach auf Papier modellieren lässt. Ein weiterer wichtiger Aspekt ist der Performance Vorteil bei Abfragen von stark verknüpften Daten. Abfragen in relationalen Datenbanken werden langsamer, je größer der Datenbestand. Im Graphen wird immer nur der Teilgraph durchlaufen welcher, die die Abfrage erfüllt. [RWE15b, 8]

2.4.3 Neo4J

Die Open Source Graphdatenbank Neo4j ist durch ein *Property Graph Model* implementiert. Knoten im Graph können durch *Label (Beschriftungen)* in Kategorien eingeordnet werden und speichern Informationen in Schlüssel-Wert Paaren als *Properties (Eigenschaften)*. Neben Knoten können auch Kanten Eigenschaften erhalten, was eine

Gewichtung der Beziehungen zwischen Knoten ermöglicht, indem man z.B. eine Zahl als Eigenschaft für eine Kante speichert. Eine Beziehung besteht immer aus einem Start- und Endknoten, sowie einer Richtung.[RWE15a, 26] Anfragen an die Neo4j Datenbank werden mittels der deklarativen Sprache *Cypher* ausgeführt. Mit dieser Sprache können Daten gefunden werden, welche einem Muster entsprechen. Diese Patterns setzen sich aus Knoten und Beziehungen zusammen.

Mit der *MATCH* Klausel, werden alle Pfade im Graphen gezeigt, welche dem Pattern entsprechen. Mit *WHERE* kann die Sammlung gefiltert werden.

2.5 Empfehlungssysteme

Ein Empfehlungssystem ist eine Software, welche Nutzern Vorschläge zu Artikeln, Music, Filmen, Büchern oder anderen Objekten macht. [RRSK10] Ausgehend von einem Objekt werden dem Benutzer andere Objekte mit Ähnlichkeiten zu dem Ausgangsobjekt aufgelistet. Für die Ergebnisse werden im wesentlichen zwei Ansätze verfolgt. Das Inhaltsbasierte Filtern und Kollaborative Filtern. Beim Inhaltsbasierenden Filtern werden Objekte mit einander verglichen. Ein einfacher Anwendungsfall sind zb. Produkt Empfehlungen basierend auf Eigenschaften der Produkte. Wählt ein Benutzer ein Produkt aus, so werden ihm Produkte mit ähnlichen Eigenschaften aufgelistet. Mit dem Kollaborativen Filtern wird das Verhalten von Nutzern verglichen. Ein Empfehlungssystem für Online Shops, kann dahingehend erweitert werden, dass nach Produkten gesucht wird die andere Benutzer ebenfalls gekauft haben.

3 Verwandte Themen und Arbeiten

3.1 Openbadge

Auf unserem Bildungsweg werden das Erlangen von Fertigkeiten und Kenntnissen mit Zeugnissen und Abschlüssen belegt. Doch oftmals genügt die formale Ausbildung nicht oder hat aufgrund der sich schnell verändernden Technologien oder Kompetenzen nur eine begrenzte Gültigkeit. Die Europäische Union fordert eine stärkere Anerkennung von informalem Lernen, damit auch Fertigkeiten und Kenntnisse die ohne ein formales Abschlusszertifikat erworben wurden Anerkennung finden.[Dor14]

Doch wie können Personen alle Ihre Fertigkeiten, welche an einer Hochschule, in einer staatlich Anerkannten Ausbildung oder in einem Online Seminar, in einem Workshop etc. erworben wurden präsentieren, damit auch Arbeitgeber und Bildungsinstitute in der Lage sind, sicherzustellen, dass Bewerber die nötigen Fertigkeiten mitbringen.[all]

In der digitalen Welt können sogenannte Badges ein Lösungsansatz sein. Ein digitaler Badge ist ein digitales Zertifikat für eine erbrachte Leistung oder eine Fähigkeit. Die Mozilla Foundation hat in Zusammenarbeit mit der MacArthur Foundation den Open Badge (OB) Standard entwickelt. Er stellt sicher, dass Alle Badges Informationen über Kriterien und Nachweise erhalten. Die Informationen in einem Badge können auch auf ein Kompetenzframework verweisen, und validiert werden.[all, 4]

Badges können von Institutionen, Schulen und Arbeitgebern verliehen werden. Sie definieren ein Set von Kompetenzen oder einen Lehrplan und eine Bewertung um festzustellen ob ein Empfänger die notwendigen Anforderungen erfüllt hat. Darüber hinaus können Badges von ihrem "issuer" mit einer verschlüsselten Zusicherung versehen werden, welche bestätigt, dass der "earner" des Badges die geforderte Leistung auch erbracht hat. Die Zusicherung kann dann in den Quellcode eines SVG oder PNG Bildes geschrieben werden, sodass dritte später eine elektronisch Überprüfung beim Herausgeber beantragen können. Über das Alignment-Attribut kann ein Badge auch auf eine Quelle verweisen, welche die Kompetenz oder Fähigkeit beschreibt.

3.2 Europäischer Qualifikationsrahmen

Der EQR ist ein Meta-Rahmen, welcher die Vergleichbarkeit von beruflichen Qualifikationen und Kompetenzen aus verschiedenen nationalen oder sektoralen Kompetenzrahmen ermöglichen soll. Er dient als Übersetzungswerkzeug zwischen den einzelnen nationalen Qualifikationsrahmen und ist in 8 Referenzniveaus aufgeteilt. Jedes Niveau wird durch *Wissen*, *Fertigkeiten* und *Kompetenz* definiert. Im EQR bedeutet Kompe-

tenz, verantwortungsvolles Handeln im Kontext. Darüber hinaus kann jedes Niveau auch durch Lernergebnisse, also das resultierende Wissen und das was der Lernende nach Abschluss einer Qualifikation in der Lage ist zu tun, beschrieben werden.

3.2.1 Deutscher Qualifikationsrahmen

Auf nationaler Ebene hat der DQR die selben 8 Niveaus, teilt aber in der Horizontalen die Kategorie Kompetenz, in *Sozialkompetenz* und *Selbständigkeit*. Damit wird ein engerer Bezug zum deutschen Bildungssystem hergestellt, in welchem das Kompetenzverständnis stärker auf Lernergebnissen und Abschlüssen fokussiert ist.

3.3 European e-Competence Framework

Der europäische Kompetenzrahmen für Fach- und Führungskräfte der Informations- und Kommunikationstechnologiebranche, ist eine sektor-spezifische Umsetzung des Europäischen Qualifikationsrahmens EQR. Er unterteilt Kompetenzen in 5 Feldern auf 5 Niveaus.

- Planen
- Erstellen
- Betreiben
- Ermöglichen
- Steuern

Die 5 Niveaus entsprechen den EQF Niveaus 3-8.

3.4 InLoc

Das Europäische Projekt InLOC (Integrating Learning Outcomes and Competences) soll Kompetenzen und Lernergebnisse verschiedener Kompetenzrahmen in einem einheitlichen semantischen Format abbilden. Dazu wird ein Datenmodell ⁷ vorgeschlagen, welches Lernergebnisse und Kompetenzen in 2 Klassen, *Definitionen* und *Struktur* einteilt. Definitionen können alle Entitäten sein, welche eine Kompetenz oder ein Lernergebnis beschreiben. Kompetenzen und Lernergebnisse können gruppiert oder hierarchisch angeordnet sein. Dieser Sachverhalt lässt sich durch die Struktur Klasse und ihre Beziehungen abbilden.

⁷inLoc[inl]

3.5 ESCO

In der EU gibt es nach einer Aktuellen Eurostat Statistik ca. 19 Millionen Menschen ohne Beschäftigung. Jedoch haben einige Branchen in Deutschland Probleme Stellen mit qualifiziertem Personal zu besetzen. So blieben im Jahr 2016 In der IT und Telekommunikationsbranche 375.034 Stellen unbesetzt.[Ade]

Die Europäische Kommission hat dieses Problem erkannt und mit ESCO eine mehrsprachige Klassifizierung für europäische Fähigkeiten, Kompetenzen, Qualifikationen und Berufe entwickelt, deren Zusammenhang durch Berufsprofile verdeutlicht wird. Eine der Aufgaben von ESCO ist es, die Lücken zwischen dem Arbeitsmarkt und den verschiedenen Bildungssystemen der einzelnen Mitgliedstaaten zu schließen. So unterscheiden sich Qualifikationen, welche Menschen in ihren Heimatländern erhalten, nicht nur voneinander, sondern können auch oftmals nicht mit aktuellen Entwicklungen des Arbeitsmarktes und dessen Anforderungen mithalten.

Die ESCO Daten werden gemäß den Praktiken für Linked Open Data veröffentlicht. Dies soll Entwicklern den Zugriff erleichtern und Anwendungen für Stellenausgleich, Berufsberatung und Selbsteinschätzungen ermöglichen.

Die Europäische Kommission hat mit ESCO eine Schnittstelle geschaffen, welche Informationen zwischen den nationalen Klassifizierung Systemen übersetzen soll und somit eine höhere semantische Interoperabilität schaffen wird. Ein Hauptinteresse von ESCO ist der kompetenzbasierte Job Abgleich. Arbeitnehmer sollen ihre eigenen Fähigkeiten, Kompetenzen und Qualifikationen mit freien Stellen vergleichen können, und so eventuelle Kompetenzlücken identifizieren zu können. Auf der anderen Seite, muss es Arbeitgebern möglich sein, Stellenausschreibungen durch Fähigkeiten, Kompetenzen und Qualifikationen zu beschreiben, und Bewerber mit den geforderten Kompetenzen abzugleichen. Diese Anforderungen müssen nun von IT-System erfüllt werden.

3.5.1 EURES

Das EURES Job-Portal bietet Arbeitssuchenden in der EU die Möglichkeit Lebensläufe und Profile mit Hilfe der ESCO Taxonomie zu erstellen. Diese werden dann mittels Job-Matching mit offenen Stellen europaweit abgeglichen.

Neo4J bietet über die Sandbox eine interaktive Möglichkeit auf einer temporär generierten Instanz im Browser zu arbeiten. Mit Schritt-für-Schritt Anleitungen werden Themen wie "Netzwerk Management" oder die "Panama Papers" in Neo4j näher gebracht.

Ein für diese Arbeit relevantes Themenfeld bietet die Sandbox "Recommendations".[neob] Als Datenquelle für die Instanz stehen die "Open Movie Database"[omd], und das MovieLens Projekt[gro16] zur Verfügung. Neben einer Erläuterung zum Property Graph Model und einer Einführung in die Cypher Query Language gibt es Beispiele zu verschiedenen Methoden und Metriken für das Filtern von Ergebnissen. Dabei werden verschiedene Ansätze zu den Methoden Inhaltsbasierte Filterung und Kollaborative Filterung erläutert.

3.6 RDF Import Neo4J

Jesús Barrasa ist *Senior Graph Solutions Consultant at Neo Technology* bei Neo4j, und hat ein Plugin für Neo4j entworfen, mit dessen Hilfe sich RDF Dokumente in Neo4j importieren lassen. So lässt sich mit dem Befehl :

```
1 CALL semantics.importRDF("file:///.../esco_skos.rdf", "RDF/XML",  
2 { languageFilter: 'de', commitSize: 5000 , nodeCacheSize: 250000})
```

Listing 1: Das Listing zeigt einen Funktionsaufruf über die Neo4j

Der komplette RDF Graph des ESCO Kataloges mit allen Beziehungen laden und Abfragen.

3.7 Recommender Systeme

3.8 Movielense

Das *Movielense*⁸ Projekt der GroupLens Forschungsgruppe an der *University of Minnesota* befasst sich mit Technologien für das kollaborative Filtern und Empfehlungen von Filmen. Die Daten⁹ von bis zu 45,000 Filmen und deren Bewertungen durch die Nutzer stehen zu Forschungszwecken als Download zur Verfügung.

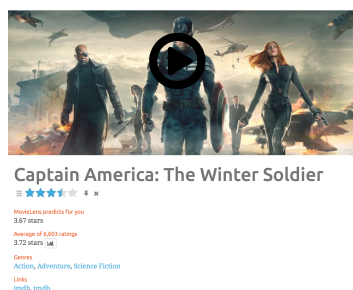


Abbildung 1: Movielense: Vorhersage für eine Filmempfehlung[mova]

⁸MovieLens [movb]

⁹Movielens Daten[gro16]

3.8.1 RecSys Challenge

Die *ACM RecSys Conference* befasst sich jedes Jahr mit den aktuellen Untersuchungsergebnissen und Techniken im Bereich der Empfehlungsdienste. Darüber hinaus veranstaltet sie auch die RecSys Challenge, einen mit 3000 € Siegerprämie dotierten Wettbewerb, bei dem Teilnehmer ein Empfehlungsdienst, in einer bestimmten Domäne entwickeln sollen. In den beiden vergangenen Jahren beschäftigte sich der Wettbewerb mit Job Empfehlungen der Karriereplattform Xing. Die Aufgabe war, anhand eines neuen Job Angebotes, all jene Nutzer zu finden, die interessiert sein könnten eine Benachrichtigung über das neue Angebot zu erhalten, aber auch Nutzer die ebenfalls für den Job geeignet sein könnten. Die Ergebnisse werden auf der RecSys Conference in Como, Italien am 27.08.2017 vorgestellt.

3.8.2 Job-Matching: Xing

Das Karrierenetzwerk bietet Arbeitssuchenden die Möglichkeit Job-Empfehlungen auf Basis des eigenen Profils zu erhalten. Dabei wird nach den Kriterien Entfernung zum Arbeitsort, Karrierestufe, Skills und Aktivitäten gefiltert. Aus diesen 4 Kriterien wird ein Relevanz-Indikator ermittelt, welcher die Treffgenauigkeit des Inserats angibt. Umgekehrt wird auch ein Inserat mit dem eigenen Profil gematcht, wobei auch die geforderten Skills den eigenen gegenübergestellt werden.[Hö16]

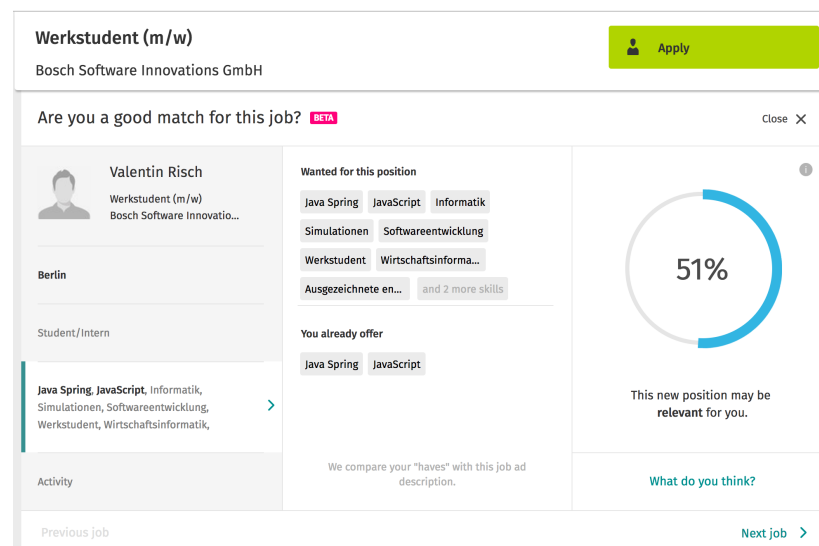


Abbildung 2: Jobmatching Xing-Profil

Dieser Ansatz berücksichtigt jedoch nicht die Persönlichkeit der Bewerber, und bietet auch nur eine teilweise höhere Bewerberpassgenauigkeit aus Sicht der Arbeitgeber. Wie qualifiziert ein Bewerber wirklich ist bleibt weiter unklar, da eine Liste der angegebenen Skills keine Auskunft darüber gibt wie kompetent der Bewerber auf einem Gebiet wirklich ist. So kann es passieren, dass ein Bewerber der trotz Eignung durch das Raster fällt, weil er andere Schlagworte als die geforderten angegeben hat.

4 Analyse und Aufgabenstellung

Eine der großen Herausforderung in der Personalentwicklung ist das Job-Matching, eine vakante Stelle mit einem passenden Bewerber oder Angestellten zu besetzen. Dabei greifen Personaler vermehrt auf Technologien und Matching-Algorithmen zurück, die diesen Prozess automatisieren sollen. Das Prinzip ähnelt dem, der Partnervermittlung. Person A, Der Arbeitgeber wird mit Person B, dem Arbeitssuchenden oder Angestellten zusammen gebracht. Unter Berücksichtigung von Kompetenzen die eine Stelle erfordert, kann dieses Matching Verfahren verfeinert werden. Im Kontext der Personalentwicklung, Jobvermittlung oder Berufsbildung ist es aber auch wichtig Karrierepfade aufzuzeigen, also heraus zu finden welche Kompetenzen sich ähnlich sind und welche Kompetenzen fehlen, um die Anforderungen für eine bestimmte Tätigkeit zu erfüllen. Dieser Prozess kann durch Algorithmen und Metriken implementiert werden.

4.1 Aufgabenstellung

Ein wichtiger Bestandteil von Stellenbeschreibungen sind auch eine Liste von geforderten Kompetenzen. Kompetenzen lassen sich als Graph modellieren, denn oftmals bestehen zwischen ihnen hierarchische Strukturen und Gemeinsamkeiten. Das Ziel dieser Arbeit soll es sein, Die Beziehungen dieser Kompetenzen in einem Graphenmodell zu analysieren, und Methoden und Metriken zu finden mit deren Hilfe es möglich sein wird, Ähnlichkeiten von einzelnen Kompetenzen oder ganzen Kompetenzsets zu ermitteln. Konkret soll ein Dienst entwickelt werden, der von anderen Programmen genutzt, oder in bestehende Systeme integriert werden kann.

4.2 Anforderungsanalyse

4.2.1 Kompetenzen

Kompetenzen sind die Entitäten welche von dem Empfehlungssystem verarbeitet werden sollen. Sie können ihren Ursprung außerhalb des Systems haben, werden aber innerhalb in einer Datenschicht persistiert. Ursprung von Kompetenzen können ein *Kompetenzrepository* wie es Herr Lopez in seiner Arbeit *Entwicklung und Evaluation eines e-Kompetenz-Verzeichnisses mit REST-API und eines automatisierten Crawlers zur Datensammlung* [Lop17] entwickelt hat, oder ein Katalog mit Kompetenzen wie *ESCO* von der Europäischen Kommission. Auch Kompetenzen aus Kompetenzrahmen wären als Quelle denkbar, sofern diese in einem maschinenlesbaren Format zur Verfügung stehen. Eine weitere denkbare Möglichkeit wäre das Auslesen des *criteria* oder *alignment* Feldes von Open Badges, wie es in der Veröffentlichung *roposal on Competency Alignment and Directory* [Kon16] des Open Badge Networks vorgeschlagen wird.

4.2.2 Anforderungen an die Engine

Ein solches Empfehlungssystem soll als eigener Dienst fungieren, und über eine API erreichbar sein. Die Eingabe für das Empfehlungssystem soll eine Kompetenz oder eine ein Set von Kompetenzen sein. Kompetenzen werden in einer Datenschicht persistiert, die die gängigen CRUD Datenbankoperationen ermöglicht. Die API leitet die Kompetenzen dann an eine Methode weiter, welche die nötigen Entitäten und deren Beziehungen zurück liefert. Auf resultierenden Datensätzen können weitere Operationen, wie Aggregation oder Sortierung durchgeführt werden und falls weitere Berechnungen oder Operationen nötig sind die mit der Abfragensprache nicht geleistet werden, mit einer gängigen Programmiersprache weitere algorithmische Schritte implementieren. Am Ende sollen 0, 1 oder mehrere Kompetenzen in sortierter Reihenfolge an den Aufrufer zurückgeliefert werden.

Folgende Anforderungen können an dieses System gestellt werden.

- Übereinstimmung von Kompetenzsets ermitteln
- Ähnlichkeit von Kompetenzsets ermitteln
- Ähnliche Kompetenzen finden
- Inkludierende Kompetenzen ermitteln
- Fehlende Kompetenzen ermitteln

Optionale Anforderungen

- Vergleich von Kompetenzen aus verschiedenen Frameworks

4.2.3 Anforderungen an die API

Das Empfehlungssystem soll über eine API für andere Softwaresysteme verfügbar sein. Beispiele für solche Systeme können Jobportale oder Human Resource Management Systeme sein. Sie können über HTTP Requests Daten mit Kompetenzsets an die Engine senden und erhalten verarbeitete Daten als Antwort. Mit der Übereinstimmung von Kompetenzsets z.B. das einer Stellenbeschreibung und das eines Lebenslaufes, kann ein Matching vorgenommen werden, welches Computersystemen ermöglicht passende Stellenangebote oder Bewerber zu finden. Über die Vernetzung der Kompetenzen im Graphen können Wege zur weiteren Bildung ermittelt werden.

4.2.4 Anforderungen an die Architektur

Der Dienst soll nach dem Microservice Muster implementiert werden, also bei Bedarf komplett austauschbar sein und wenige Abhängigkeiten enthalten. Nutzer sollen über das HTTP Protokoll mit dem Dienst kommunizieren können. Die Daten werden in einem Graphen persistiert, in welchem Kompetenzen als Knoten und Beziehungen der Kompetenzen als Kanten gespeichert werden. Für die Kommunikation soll ein gängiges Format zur Ein- und Ausgabe gewählt werden.

- Erreichbarkeit über eine Web-API
- Persistenz der Daten in einem Graphen
- Daten Ein und Ausgabe in einem standardisierten Format
- Skalierbarkeit: soll viele Anfragen gleichzeitig bearbeiten können

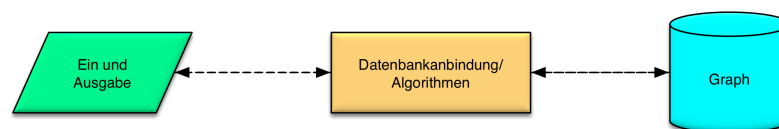


Abbildung 3: Entwurf

4.3 Algorithmen und Metriken

Das Kernproblem des Empfehlungsdienstes ist die Berechnung von Ähnlichkeiten von Knoten in einem Graphen. Es gibt bereits einige Algorithmen und Metriken die sich mit dieser Problematik auseinander setzen. Welche Metriken und Algorithmen angewandt werden können hängt aber auch vom Anwendungsfall ab. Einige Methoden zur Berechnung von Ähnlichkeiten sollen im folgenden aufgelistet, und später bei der Implementierung evaluiert werden. Ein grundsätzlicher Ansatz kann lautet, *je näher zwei Knoten im Graphen einander sind, desto ähnlicher sind sie sich*. Der folgende Graph soll dies veranschaulichen.

Das Problem des *kürzesten Pfades* ist ein häufig behandeltes Problem in der Graphentheorie. Ein kürzester Pfad von einem Knoten s zu t in einem Graphen G ist definiert durch eine Funktion $c: R \rightarrow \mathbb{R}$. Es gibt aber auch mehr Eigenschaften des Graphen, die sich ein Algorithmus zu nutzen machen kann. So können mit Hilfe von gewichteten Kanten Vektoren entlang der Pfade von einem Knoten zu einem anderen erstellt werden. Die *Kosinus-Ähnlichkeit* ist eine Metrik, die den eingeschlossenen Winkel zwischen zwei solcher Vektoren ermittelt.

Mit dem **Jaccard Koeffizienten** lassen sich zwei Mengen auf ihre Ähnlichkeit prüfen. Der Wert des Koeffizienten ist genau dann 1, wenn die beiden Mengen identisch sind. Je weniger Gemeinsame Elemente zwei Mengen besitzen, desto kleiner wird der Koeffizient und wird 0, wenn keine Gemeinsamkeiten vorliegen.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

In einem Graphen kann die Metrik für 2 Knoten angewandt werden, indem man die beiden Mengen aus Knoten zusammen setzt, die eine Beziehung zu den 2 Knoten haben. Kategorien oder Taxonomien sind dabei geeignete Kandidaten für gemeinsame Beziehungen. Am Beispiel des Movie Graphen, kann hier eine Ähnlichkeit von 2 Filmen berechnet werden, die sich in gemeinsamen Genres befinden. Je mehr gemeinsame Genres, desto höher die Ähnlichkeit. Tauchen beide Filme in denselben Genres auf, wie z.B. in Horror, Action, Comedy, so haben sie eine Ähnlichkeit von 1, haben sie kein gemeinsames Genre ist die Ähnlichkeit 0.

Ein weiterer Ansatz lautet *"Zwei Objekte sind Ähnlich, wenn sie mit ähnlichen Objekten verknüpft sind"*. Der **SimRank** [JW02] Algorithmus versucht Ähnlichkeiten von Knoten zu finden zwischen denen ein struktureller Zusammenhang besteht.

Ontologien im Semantic Web haben eine Baumstruktur. Um semantische Nähe zwischen Konzepten zu berechnen müssen neben der Distanz der Knoten auch die Tiefe im Baum berücksichtigt werden. **Wu und Palmer** [WP94] schlagen folgende Berechnung für 2 verwandte Konzepte C_1 und C_2 vor:

$$Sim_{wp} = \frac{2 \times N_3}{N_1 + N_2 + 2 \times N_3}$$

Wobei N_3 die Länge des Pfades vom Wurzelknoten bis zum nächsten gemeinsamen Knoten C_3 , N_2 die Länge des Pfades $p(C_3, C_2)$ und N_1 die Länge des Pfades $p'(C_1, C_3)$ ist.

Mit der **Edit-Distanz** [KN12] wird mit Hilfe von Edit-Operationen, also dem löschen und hinzufügen von Kanten eine Graphentransformation durchgeführt, und für jede Edit-Operation η Transformationskosten $c(\eta)$ veranschlagt. Die Summe der minimalen Kosten der Transformationsfolge D kann dann als Maß für eine Ähnlichkeit von zwei Graphen G_1 und G_2 verwendet werden.

5 Entwurf

Im folgenden sollen die aus der Analysephase ausgearbeiteten Anforderungen zu einem Entwurf zusammen gefasst werden, in dem die notwendigen Komponenten und deren Schnittstellen genauer beschrieben werden. Zunächst soll aber das Datenmodell für die Kompetenzen und dessen Format gewählt werden.

5.1 Format/Kompetenzmodell

Wie in Kapitel 1 beschrieben, gibt es unterschiedliche Ansätze und widersprüchliche Meinungen um ein Modell für Kompetenzen zu beschreiben. Für den Entwurf des Empfehlungssystems sollen zunächst einige in Frage kommenden Modelle evaluiert werden.

5.1.1 ESCO als RDF

ESCO wird als RDF Graph im SKOS Format zum Download angeboten. Das Vokabular[esc] für dieses Schema beinhaltet unter anderem folgende relevante Klassen: ESCO concepts

Relationship

IRI: <http://data.europa.eu/esco/modelRelationship>

Beziehung zwischen zwei ESCO Säulen z.B. Occupation und Qualification.

Qualification

IRI: <http://data.europa.eu/esco/modelQualification>

Ein offiziell, oder formelles Zertifikat eines, oder mehrerer Kompetenzen oder Fertigkeiten. Kann verknüpft mit Skills oder Berufen sein.

Skill

IRI: <http://data.europa.eu/esco/modelSkill>

In dieser Klasse werden Konzepte für Fertigkeiten und Kompetenzen und Wissen zusammengefasst.

ESCO wendet hier dieselben Definitionen für diese Begriffe an, wie sie der EQF vorgibt.[?] ESCO unterscheidet in seinem Modell nicht zwischen Fertigkeiten und Kompetenzen sondern nur zwischen Fertigkeiten und Wissen. Dieser Unterschied wird durch hierarchische Strukturen implementiert.

Occupation

IRI: <http://data.europa.eu/esco/modelOccupation>

In dieser Klasse werden Berufsgruppen zusammengefasst.

Kritik für das ESCO Modell kommt hingegen vom Bundes Ministerium für Berufliche Bildung. Es sieht mögliche negative Rückwirkungen auf duale Berufsbildungssysteme, beim Einsatz von europaweiten Teilkompetenzen und fordert eine genauere Überprüfung der Kompatibilität von ESCO mit Standards wie EQR und NQR hinsichtlich der Qualifikationsniveaus der Beschreibungen.

5.1.2 inLOC

Das Projekt InLOC ermöglicht das Abbilden von mehreren Kompetenzrahmen in einem einheitlichen maschinenlesbaren Format(RDF,XML,JSON-LD). Auf der Projekt Website wird anhand einer exemplarischen Kompetenz des e-CF in Version 2.0 die Struktur des inLOC Informations Modells beschrieben.

Das folgende Digramm zeigt die Struktur für Kompetenzen aus dem e-CF. Deutlich zu erkennen sind die 4 Dimensionen *Kompetenzfelder*, *Kompetenzen*, *Kompetenzniveaus*, *Beispiele für Wissen und Fähigkeiten*

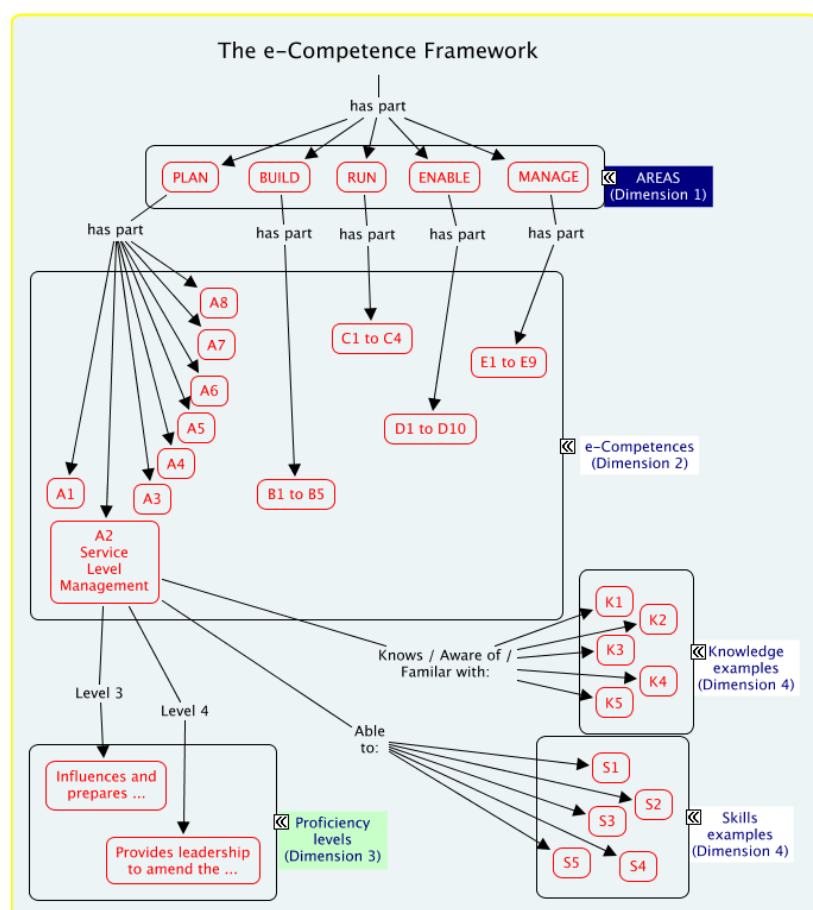


Abbildung 4: e-Kompetenz in inLOC

5.1.3 Fazit: Auswahl eines Models

Das inLOC Model findet aktuell keine Anwendung, und man müsste eine erst eine Reihe von Beispiel Daten erstellen, z.B. durch manuelles eintragen der Kompetenzen aus dem e-CF.

Für die Implementierung soll zunächst nur das Modell, welches durch den ESCO Katalog angeboten wird, verwendet werden. Zum einen Steht dieser Katalog in der aktuellen Version mit ca. 13.500 Einträgen zu Wissens und Fertigkeiten/Kompetenz Konzepten zum Download zur Verfügung, und ist darüber hinaus EU Länder übergreifend, was ihn für einen breiteres Spektrum an Anwendungen interessant macht. Der Aspekt der Kompatibilität mit einzelnen nationalen Bildungssystemen kann vernachlässigt werden, da es nicht um die Findung einer Lösung geht, welche den Arbeitsmarkt und die Berufsbildung in gleichem Maße berücksichtigt, sondern den rein technischen Ansatz, für ein Verfahren Kompetenzen aufgrund ihrer strukturellen Modellierung zu vergleichen.

5.1.4 Datenbank

Nachdem nun das Datenmodell für die Speicherung der Kompetenzen gewählt wurde, muss eine geeignete Datenbank gewählt werden. Das SKOS Schema des ESCO Katalogs liegt zwar bereits als gerichteter Graph, nämlich im RDF Format vor, soll aber aus Gründen der Flexibilität und Möglichkeiten in ein Datenbankmanagement System überführt werden. Alternativ könnte mit Hilfe von SPARQL, einer graphen basierten Abfragesprache, direkt auf dem RDF Graphen gearbeitet werden ,hier sollen jedoch von den Vorteilen eines *Label Property Graphen* Gebrauch gemacht werden. Möchte man eine Kompetenz in RDF modellieren, so muss zunächst ein Knoten für die Kompetenz selbst angelegt werden. Soll die Kompetenz neben dem *Identifier* auch noch ein Namens- oder Sprachattribut hinzufügen, müssen zunächst Knoten und Beziehungen für jedes Attribut angelegt werden. Daraus resultiert ein stärkere Vernetzung des Graphen. Im *LPG* können die Attribute einfach dem Kompetenzknoten hinzugefügt werden. Emil Eifrem, CEO der Open-Source Graphendatenbank [?] übt Kritik an der RDF Community. So bemängelt er, dass zu wenig Rücksicht auf Entwickler und die Integration in Software Systeme genommen wird. [Ana17]

"There is a pool of super smart people in the Semantic Web community, but their approach is typically extremely academic. " Warum keine RDF Technologie?

5.2 Datenmodell

Da der ESCO Katalog als Ontologie implementiert und im Linked Data Format als SKOS veröffentlicht wird, liegen die Konzepte von Kompetenzen als hierarchische Struktur vor. Dabei gibt es ausgehend vom Root Knoten einen Unterknoten für die *Kompetenzen/Fertigkeiten* Säule, von dem ausgehend sich der Baum aufspaltet in Job-spezifische und Transversale Skills. Hierarchien in SKOS werden über *broader-narrower* Beziehungen hergestellt. Kompetenzen sind Teil der Konzept Klasse *Skill* und werden mit der *LeafGroup* und *Member* Konzept Klasse strukturiert. Beispielsweise ist die Kompetenz *Windows XP* Teil der Gruppe *Computing* die wiederum Teil der Gruppe *Job-specific skills/competences* Gruppe ist.

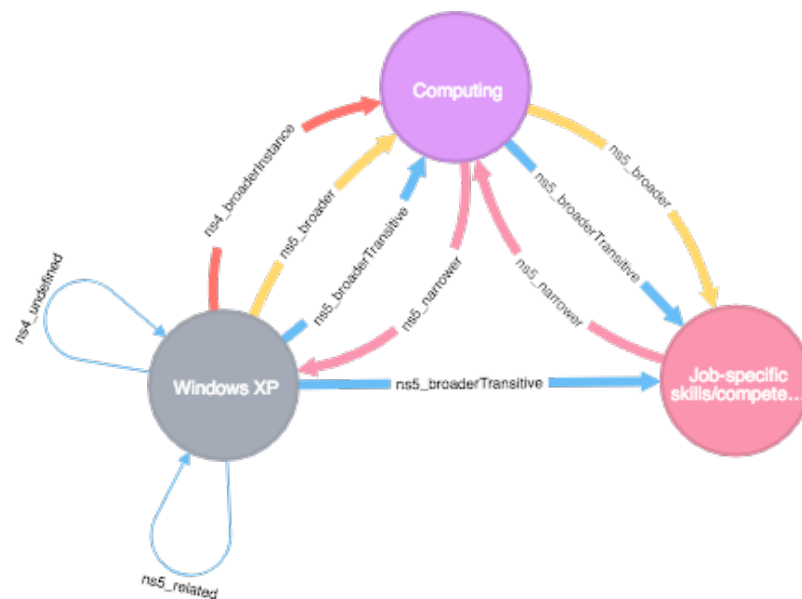


Abbildung 5: Hierarchische Struktur von Kompetenzen

5.2.1 API

Die Funktionsweise der Recommendation Engine soll sich lediglich auf folgende Prozesse beschränken:

Eingabe und Ausgabe einer Kompetenz Berechnung von Ähnlichkeiten im Graphen durch Graph Traversierung und Abfragen an die Datenbank. Über das HTTP Protokoll sollen Anwendungen mit dem Empfehlungsdienst kommunizieren können. Dabei werden die CRUD Methoden POST und GET auf entsprechenden Endpunkten aufgerufen.

Die Zweite Komponente muss eine Datenbankanbindung für die Datenbank gewährleisten, und Daten aus der Eingabe in Anfragen an die Datenbank verarbeiten, sowie Ergebnisse der Anfragen zurückliefern. Falls die Anfragensprache der Datenbank nicht

alle Möglichkeiten ausschöpfen kann die Ergebnisse wie gewünscht zu filtern, können in der zweiten Komponente weitere Algorithmen implementiert werden.

Die Anwendung ist also wenig Komplex. Ein viel höherer Implementierungsaufwand ist den Vergleichsalgorithmen, bzw Graph Queries zuzurechnen.

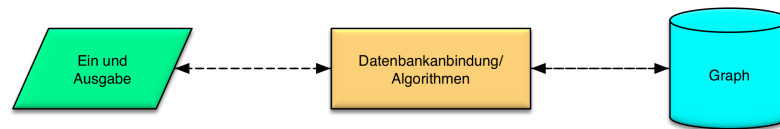


Abbildung 6: Entwurf

5.3 Anfragen Formulieren

Bevor Abfragen in einer Abfragensprache für die Datenbank erstellt werden, sollen diese zunächst in natürlicher Sprache formuliert werden, und später in eine entsprechende Abfragensprache transformiert werden. Dabei soll das zuvor erstellte Datenmodell helfen. Anhand der Knoten und Kanten können die Entitäten und deren Beziehungen identifiziert werden, aus diesen lassen sich dann Abfrage Sätze formulieren. In der Formulierung wird zum Ausdruck gebracht, welche die gewünschten Knoten sind und welcher Bedingung das Ergebnis unterliegt. Die Einfachste Abfrage würde demnach lauten: *Zeige Alle Knoten*. Das Ergebnis wäre der Komplette Graph. Eine spezifischere Ergebnis erhält man durch das hinzufügen von Bedingungen.

Zeige Alle Knoten, mit dem Attribut name = {Softwareentwicklung}.

6 Implementierung

In diesem Kapitel sollen die einzelnen Implementierungsschritte für den Empfehlungsdienst erläutert werden. Alle Komponenten können in der *Amazon Webservices Cloud* implementiert werden. Dabei sollen nur die eingesetzten Dienste beschrieben werden, jedoch auf grundlegende Einstellungen wie die Benutzer Rollen und Berechtigungen in der AWS Konsole über welche sich sämtliche Dienste steuern und konfigurieren lassen, nicht weiter eingegangen werden. Im einzelnen kommen folgende Dienste zum Einsatz:

6.1 Architektur

EC2: Mit Elastic Compute Cloud können virtuelle Rechner Umgebungen erstellt werden auf welchen skalierbare Applikationen entwickelt werden können. Rechen und Speicherkapazitäten können den persönlichen Bedürfnissen angepasst werden.

AWS API Gateway: Mit einem API Gateway können Anwendungen über das Internet auf die Cloud Dienste von Amazon zugreifen. Dazu werden HTTP-Endpunkte erstellt.

AWS Lambda Code in Lambda Funktionen wird nur ausgeführt wenn entsprechende Events die Lambda Funktion auslösen.

6.2 EC2

Für die Datenhaltung wird das Open Source Graphendatenbank Management System Neo4j verwendet. Dieses stellt jedoch einige Vorraussetzungen an die Umgebung[neoa]. Da Neo4j eine Java VM muss z.B. für Ubuntu das OpenJDK 8 installiert sein. Die Neo4j Instanz wird auf einer Ubuntu Umgebung installiert. Dafür müssen jedoch erst einige Abhängigkeiten und Pakete installiert werden. Um den Aufwand gering zuhalten, wird eine Amazon Web Services EC2 Instanz hochgefahren mit einem speziellen Cloudformation Template, welches alle System Vorraussetzungen erfüllt. Cloudformation erleichtert das Erstellen von Umgebungen mit wichtigen Laufzeitparametern. Das folgende Schaubild veranschaulicht alle wichtigen Parameter:

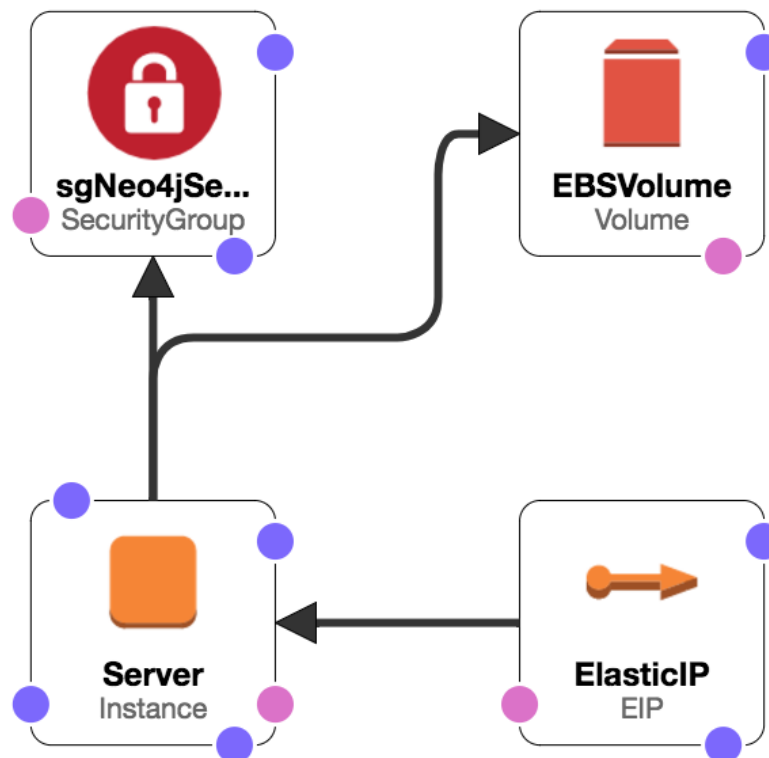


Abbildung 7: EC2 Cloudformation Design Template

6.3 Neo4j

Über den freigegebenen Port 7474 bietet Neo4j eine Weboberfläche zur Abfrage der Daten und Visualisierung der Ergebnisse an. Die Datenbank soll nun initial mit realen Daten befüllt werden.

6.3.1 RDF Import

Neo4j lässt sich über Plugins erweitern. Diese so genannten *Prozeduren* werden in Java geschrieben und können mit der Cypher Query Language aufgerufen werden. Mit dem Plugin *neosemantics* können Daten aus einem Triplestore in Neo4j geladen werden. Über die Weboberfläche kann mit einer Cypher Abfrage die Methode zum importieren der Daten aufgerufen werden.

```
1 CALL semantics.importRDF("file:///.../esco_skos.rdf", "RDF", {
    shortenUrls: true, typesToLabels: true, commitSize: 9000,
    languageFilter: 'de'})
```

Listing 2: importRDF Prozeduraufruf

6.4 API Gateway

Das API Gateway wird hier als Proxy verwendet, mit dessen Hilfe HTTP Requests an Lambda Funktionen weiter gereicht werden. Die Lambda Funktionen wiederum rufen innerhalb eines virtuellen Netzwerks eine Datenbankabfrage für die EC2 Neo4j Instanz auf. Die URL für diese Instanz ist nur innerhalb des Virtuellen Netzwerks erreichbar, und ist somit für die Anwendung vor dem Proxy verborgen. Innerhalb der Lambda Funktion kann dann mit den berechneten Daten ein Response Objekt erstellt werden und an den Aufrufer der API zurückgesendet werden.

6.5 Serverless

Mit den Amazon Web Diensten können serverlose Anwendungen geschrieben werden, die es dem Entwickler ermöglichen sich auf den Kern der Applikation, den Quellcode zu konzentrieren ohne sich Gedanken über Ressourcen und Kapazitäten zu machen. Bereits in der Entwurfsphase konnte fest gestellt werden, dass die Komplexität der Anwendung weniger im Programmieraufwand, sondern mehr im finden und implementieren von Algorithmen liegt. Dennoch ist das Bereitstellen von einigen Ressourcen wie einer Laufzeitumgebung und einer API notwendig, um anderen Anwendungen über das Internet das Ausführen des Codes mit Parametern zu ermöglichen. Dabei spielt es keine Rolle wieviele Anwendungen gleichzeitig auf die API zugreifen, die Rechenkapazität werden automatisch angepasst. Draus ergibt sich ein weiterer Vorteil für Entwickler und Betreiber. Code, der nicht ausgeführt wird, verbraucht keine Ressourcen, und wird von Ressourcen Dienstleister auch nicht in Rechnung gestellt. Der geringe Overhead, die geringen Betriebskosten und die geringe Komplexität der zu entwickelnden Anwendungen laden daher dazu ein, den serverlosen Ansatz zu wählen.

6.5.1 AWS Lambda

Lambda bietet die Möglichkeit Code auszuführen ohne die nötigen Ressourcen managen zu müssen. Lambda ist Ereignis gesteuert, was bedeutet der Code wird als Reaktion auf ein Ereignis ausgeführt.

6.5.2 Serverless Framework

Das Serverless Framework ist ein Kommandozeilen Tool, und wurde zur Unterstützung des Entwicklungs und Deployment Prozess von serverlosen Anwendungen entwickelt. Der Quellcode zur Datenverarbeitung in den Lambda Funktionen kann mit diesem Framework nicht nur lokal getestet werden, sondern kann auch direkt mit einem API Gateway verknüpft werden. Alle wichtigen Parameter wie Umgebungsvariablen zb. Zugangsdaten, URLs und Endpunkte, werden in einer Konfigurationsdatei eingetragen. Über die Kommandozeile kann die Lambda Funktion aufgerufen werden, sowohl lokal

als auch auf dem Stage System. Ein Problem, was nun an den Tag tritt, ist die lokale Nichtverfügbarkeit von Diensten die nur in der Cloud existieren. Nimmt man zb. Änderungen im Code der Lambda Funktion vor und möchte diese testen, muss der Code zunächst neu deployt werden bevor neue Anfragen an die API gesendet werden können. Dieser Prozess braucht Zeit und verlängert den Entwicklungsprozess. Besser wäre es eine lokale simulierte API zur Verfügung zu haben, mit welcher sich die Änderungen im Quell Code der Lambda Funktion direkt testen lassen. Dieses Feature wird über Plugins des *Serverless Frameworks* zur Verfügung gestellt.

6.6 Architektur im Überblick

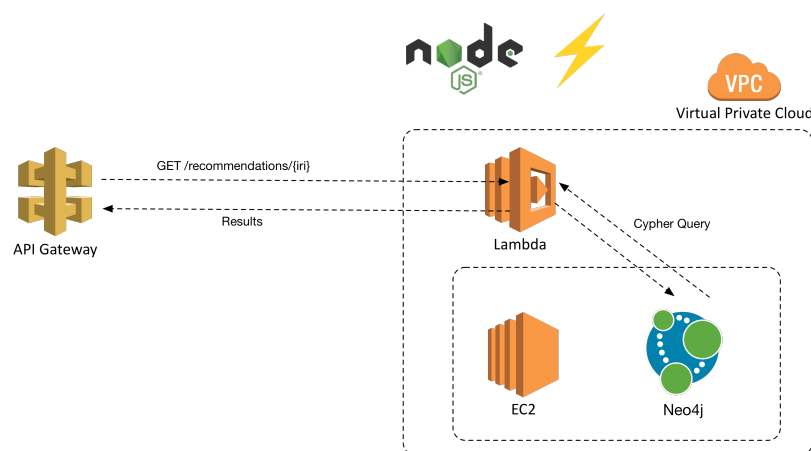


Abbildung 8: Architektur

Nachdem nun alle Komponenten eingerichtet sind erfolgt das Implementieren der Algorithmen und Datenbankabfragen. Als Laufzeitumgebung für die Lambda Funktionen wird Node.js in der Version 6.10 gewählt. Für die Anbindung an die Neo4j wird die *neo4j-driver* Bibliothek geladen. Damit lässt sich zunächst für den gesamten Scope der Lambda Funktion der Datenbanktreiber der *neo4j* Klasse instanzieren. Wird nun beim ausführen des Codes durch Lambda die *Handler* Funktion aufgerufen, kann für den jeweiligen Handler eine neue Session geöffnet werden und ein String für die Datenbankabfrage, inklusive Parameter geformt werden. Ist die Abfrage beendet wird die Callback Funktion aufgerufen und ein Response Objekt erstellt, welches das Ergebnis der Datenbankabfrage im JSON String Format enthält.

Die Im Entwurf beschriebene Kern Komponente für die Datenbankanbindung, die Ein- und Ausgabe von Daten und die Implementierung von Algorithmen, soll nun Lambda Funktionen in einer Node.js Laufzeitumgebung implementiert werden. Als erstes wird eine Instanz des Datenbanktreibers erzeugt welche von mehreren Handlern benutzt werden kann. Für jede Anforderung der Engine, soll ein eigener Handler implementiert werden, der genau dann ausgeführt wird wenn ein HTTP Request auf dem

zu dem Handler gehörenden Pfad aufgerufen wird. Parameter aus dem Query String, oder Daten die mittels POST übermittelt werden können aus dem *event* Objekt gelesen werden.

Innerhalb des Handlers wird ein *session* Objekt erzeugt, welches die Verbindung zur Datenbank herstellt und über diese auch die Abfragen an die Datenbank erfolgen. Dazu wird die *run* Methode auf dem Objekt aufgerufen und der Querystring und Queryparameter übergeben. Es wird ein Promises erzeugt, auf welchem eine Callbackfunktion aufgerufen wird sobald die Abfrage alle Ergebnisse geliefert hat. Aus dem Resultset kann ein Objekt erzeugt werden welches mit der *callback* Funktion des Lambda Handlers an den Aufrufer, in diesem Fall die API zurück gegeben wird.

Mit dem folgenden Handler wird ein Node *n* über ein eindeutiges Attribut wie einer Id oder eines IRI in der Datenbank abgefragt und an die API zurückgeliefert.

```
1   'use strict';
2
3   var neo4j = require('neo4j-driver').v1;
4
5   var driver = neo4j.driver("bolt://" + process.env.NEO4J_URL, neo4j.auth.
      basic("user", "password"));
6   module.exports.getNode = function(event, context, callback) {
7     var session = driver.session();
8     session
9       .run('MATCH (n:{queryParam}) return n;', {queryParam: event.
      queryStringParameters.label})
10    .then(function (result) {
11      result.records.forEach(function (record) {
12        callback(null, {
13          statusCode: 200,
14          body: JSON.stringify({
15            message: record,
16            input: event,
17          }),
18        });
19      });
20      session.close();
21      driver.close();
22    })
23    .catch(function (error) {
24      console.log(error);
25    });
26  };
```

Listing 3: Lambda Handler Funktion

6.7 Deployment mit dem Serverless Framework

Beim Erzeugen eines neuen Projektes mit dem Serverless Framework, wird neben einem leeren Handler auch eine Konfigurationsdatei mit dem Namen *serverless.yml* erzeugt. In dieser werden neben der Route die die Funktion auslöst auch die Art der HTTP Methoden eingetragen.

```
1
2  functions:
3    getNode:
4      handler: handler.getNode
5      events:
6        - http:
7          path: node
8          method: get
```

Listing 4: Serverless Konfiguration

Die notwendigen Paketabhängigkeiten für die Datenbankanbindung an Neo4j werden mit *npm install* in den node modules Ordner geladen. Über das Serverless Kommandozeilen Programm, wird mit dem *deploy* Aufruf der gesamte Inhalt des Projektes, inklusive Abhängigkeiten gepackt, Cloudformation Templates erstellt und dann auf ein S3 Bucket in der AWS Cloud geladen.

6.8 Algorithmen und Metriken

Das technische Gerüst ist nun komplett. Kompetenzen können per GET Request an die API, und unter Angabe ihrer IRIs abgefragt werden. Jedoch sind die Anforderungen an das Empfehlungssystem mittels Metriken und Algorithmen Ähnlichkeiten zu bestimmen.

6.8.1 Shortest Path

Neo4j bringt eine Implementierung von kürzesten Pfaden direkt von Haus aus mit. Dafür müssen in der *MATCH* Klausel nur ein spezifischer Anfangsknoten *s* und ein Endknoten *t* abgefragt werden und diese beiden in der *shortestPath* Funktion als Parameter übergeben. Wahlweise kann der Pfad, oder die Länge des Pfades zurückgegeben werden.

Als grobe Metrik kann die *shortestPath* Funktion dazu verwendet werden um heraus zu finden wie weit eine Kompetenz von einer anderen entfernt ist. Je näher, desto ähnlicher.

6.8.2 Jaccard Index

Zunächst soll geprüft werden, ob es Knoten Typen und Beziehungen im Graphen gibt mit denen sich eine Schnittmenge für zwei Kompetenzen berechnen lässt. Dabei soll eine Kompetenz von außen über die API als Eingabe erfolgen und solche Kompetenzen die eine Schnittmenge mit der gegebenen haben in geordneter Reihenfolge zurückgegeben werden. Diese Metrik lässt sich mit einer Cypher Abfrage direkt implementieren ohne weiteren Code ausführen zu müssen. Dazu wird zunächst die Abfrage formuliert.

Welche Kompetenzen haben wieviele gemeinsame Beziehungen vom selben Typ zur gegebenen Kompetenz

Als erstes werden mit einer **MATCH** Klausel alle Knoten $c_{gesucht}$ gesucht, die neben dem gegebenen Knoten c_{input} eine eingehende Beziehung zu einem Knoten o vom Typ `ns4_Occupation` haben und die Schnittmenge bilden.

```
1 MATCH (c:ns4_Skill {ns5_prefLabel:{queryParam}})-[:ns5_related]->(o:
    ns4_Occupation)<-[:ns5_related]-(other:ns4_Skill)
```

Die Kompetenz aus dem Input wird mittels Parameter der Query übergeben und das Resultat des Patterns in den Variablen c , o , $other$ mit der **WITH** Klausel in den nächsten Query Teil weiter gereicht. Mit der Aggregatsfunktion **COUNT** wird dann der Betrag der Schnittmenge von Knoten vom Typ `ns4_Occupation`, die eine eingehende Beziehung vom Typ `ns5_related` von Knoten c_{input} und Knoten $c_{gesucht}$ vom Typ `ns4_Skill` haben.

```
1 MATCH (c:ns4_Skill {ns5_prefLabel:{queryParam}})-[:ns5_related]->(o:
    ns4_Occupation)<-[:ns5_related]-(other:ns4_Skill)
2
3 WITH c, other, COUNT(o) AS intersection, COLLECT(o.ns5_prefLabel) AS i
```

Als nächstes müssen die beiden Mengen in separaten Variablen $s1_A$ und $s2_B$ gespeichert werden,

```
1 MATCH (c:ns4_Skill {ns5_prefLabel:{queryParam}})-[:ns5_related]->(o:
    ns4_Occupation)<-[:ns5_related]-(other:ns4_Skill)
2
3 WITH c, other, COUNT(o) AS intersection, COLLECT(o.ns5_prefLabel) AS i
4 MATCH (c)-[:ns5_related]->(c_other:ns4_Occupation)
5
6 WITH c, other, intersection, i, COLLECT(c_other.ns5_prefLabel) AS s1
```

```

7 MATCH (other)-[:ns5_related]->(oc:ns4_Occupation)
8
9 WITH c,other,intersection,i, s1, COLLECT(oc.ns5_prefLabel) AS s2

```

um dann mittels Listenfunktion **FILTER** die Vereinigungsmenge zu berechnen.

```

1 MATCH (c:ns4_Skill {ns5_prefLabel:{queryParam}})-[:ns5_related]->(o:
   ns4_Occupation)<-[:ns5_related]-(other:ns4_Skill)
2 WITH c, other, COUNT(o) AS intersection, COLLECT(o.ns5_prefLabel) AS i
3
4 MATCH (c)-[:ns5_related]->(c_other:ns4_Occupation)
5 WITH c,other,intersection, i ,COLLECT(c_other.ns5_prefLabel) AS s1
6
7 MATCH (other)-[:ns5_related]->(oc:ns4_Occupation)
8 WITH c,other,intersection,i, s1, COLLECT(oc.ns5_prefLabel) AS s2
9
10 WITH c,other,intersection,s1+filter(x IN s2 WHERE NOT x IN s1) AS
   union, s1, s2
11
12 RETURN c.ns5_prefLabel, other.ns5_prefLabel, s1,s2,intersection,((1.0*
   intersection)/SIZE(union))
13
14 AS jaccard ORDER BY jaccard DESC LIMIT 10

```

Der Quotient aus dem Betrag der Schnittmenge, und dem Betrag der Vereinigungsmenge ergibt dann den Jaccard Index und wird gemeinsam mit den Namens Attributen für die Kompetenzen c_{input} und $c_{gesucht}$ in, in sortierter Reihenfolge in der **RETURN** Klausel in das Ergebnis geschrieben.

6.8.3 Simrank

In der Arbeit *SimRank: A Measure of Structural-Context Similarity* wird gemäß dem Ansatz *Zwei Objekte sind ähnlich, wenn sie mit ähnlichen Objekten verknüpft sind* ein Algorithmus vorgestellt, welche Die Ähnlichkeit von Objekten berechnet die in einem strukturellen Zusammenhang stehen. Für einen Knoten v in einem gerichteten Graphen mit $G = (V, E)$ wird eine Liste $I(v)$ mit Knoten die eine eingehende Beziehung haben, und eine Liste $O(v)$ mit ausgehenden Beziehungen erstellt. Ausgehend von der Annahme, dass ein Objekt eine maximale Ähnlichkeit von 1 zu sich selbst hat, soll nun die Ähnlichkeit von jedem Objekt über die Beziehungen zum nächsten propagiert werden,

6.8.4 Levensthein oder Edit-Distanz

Die Edit-Distanz oder auch

7 Bewertung

Die API ist über die URL `https://lndkrk7h5i.execute-api.eu-central-1.amazonaws.com/prd/` erreichbar. Die Pfade der Lambda Funktionen werden mit diesem Basis Pfad konkateniert.

`/competence-service-prd-getJaccardOfCompetence`

Für die Ermittlung des Jaccard Indexes wird ein GET Request mit `curl` aufgerufen

```
1 curl -X GET 'https://lndkrk7h5i.execute-api.eu-central-1.amazonaws.com/prd/getJaccardOfCompetence?label=Windows%20XP'
```

Er liefert eine Antwort mit einem JSON Objekt, dass alle wichtigen Informationen zur Anfrage erhält. `c.ns5prefLabel` und `other.ns5prefLabel` sind die Beiden Kompetenzen die miteinander verglichen werden, `s1` und `s2` die beiden Mengen aus denen der Jaccard Index gebildet wird. `intersection` gibt den Betrag der Schnittmenge von `s1` und `s2` an. Die Variable `jaccard` ist der Berechnete Index er liegt zwischen 0 und 1.

```
1  "keys": [
2      "c.ns5_prefLabel",
3      "other.ns5_prefLabel",
4      "s1",
5      "s2",
6      "intersection",
7      "jaccard"
8  ],
9  "length": 6,
10  "_fields": [
11      "Windows XP",
12      "Linux",
13      [
14          "System engineer",
15          "IT architect",
16          "Computer instructor",
17          "Computer adviser",
18          "Database administrator",
19          ...,
20      ],
21  ],
22  [
23      "Computer engineer (computer programmers, computer
24      systems designers and analysts)",
25      "System operator",
26      "PC coordinator",
```

```

26         "Computer adviser",
27         "Database administrator",
28         ... ,
29
30     ],
31     {
32         "low": 65,
33         "high": 0
34     },
35     0.9420289855072463
36 ]

```

Listing 5: Response eines GET Requests an competence-service-prd-getJaccardOfCompetence

Das Listing zeigt einen Ausschnitt der Antwort auf einen GET Request an die API. Eine übersichtlichere Darstellung erhält man über die Web Oberfläche der Neo4j Instanz.

\$ MATCH (c:ns4_Skill {ns5_prefLabel:"Windows XP"})-[:ns5_related]->(o:ns4_Occupation)<-[:ns5_related]-(other:ns4_Skill) WITH c, other, COUNT(o) AS...

c.ns4_prefLabel	other.ns4_prefLabel	s1	s2	intersection	jaccard
Windows XP	Linux	[System engineer, IT architect, Computer instructor, Graduate engineer (system development), Interaction designer, IT support engineer, Network manager, System programmer, IT system engineer, Web supervisor, Network designer, Application developer, Dispatcher, IT project manager, Program developer, Systems manager, System tester, Usability expert, Software programmer, Software designer, Method analyst (computing), Operations technician (computing), System administrator, PC coordination manager, Network administrator, System operator, Web developer, PC coordinator, Study leader, Study instructor (other teaching associate professionals), Computer engineer (computer programmers, computer systems designers and analysts), IT consultant, System technician, Data processing operator, IT teacher, System analyst, Computer adviser, System director, Computer teacher, Media instructor, ...]	[Computer engineer (computer programmers, computer systems designers and analysts), System operator, PC coordinator, Computer adviser, Database administrator, Operations technician (data processing), Web programmer, System director, Data processing operator, Network designer, Database designer, System attendant, Web supervisor, IT consultant, System technician, System analyst, Computer teacher, Media instructor, System architect, IT architect, System designer, System technician (network), Communications technician, Software designer, Software programmer, Network administrator, Web developer, Study instructor (other teaching associate professionals), IT teacher, IT project manager, Program developer, Usability expert, System programmer, Operations technician (computing), Study leader, Supervisor (data processing), IT technician, System developer, IT support engineer, Interaction designer, Network manager, System tester, Computer trainer, Database developer, Network technician, Database engineer, System supervisor, System investigator, Usability designer, System manager, Programmer, Application engineer, Web manager, PC coordination manager, Method analyst (computing), Network operator, System administrator, Systems manager, System engineer, Graduate engineer (system development), IT system engineer, Data processing consultant, Computer operator, Application developer, Computer instructor, Webmaster]	65	0.9420289855072463

Started streaming 100 records after 1006 ms and completed after 1016 ms.

Abbildung 9: Cypher Result

8 Ausblick

Der Prototyp des Empfehlungsdienstes ist in seinen Grundzügen fertig implementiert, kann und soll erweitert und verbessert werden. So können weitere Daten mit Kompetenzen in die Datenbank geladen werden, oder neue Beziehungen eingeführt. Eine wichtige Eigenschaft des Neo4j *Label Property Graph* Models, die Gewichtung von Kanten, wird bisher noch nicht genutzt. Der Europäische Qualifikationsrahmen sieht aber Niveau Levels vor, mit diesen könnte sich die Kosinus Ähnlichkeit als weitere Metrik berechnen lassen. Die implementierten Methoden zur Ähnlichkeitsmessung sind lediglich ein Ansatz, für einen Vollfunktionsumfänglichen Empfehlungsdienst müssten die einzelnen Metriken evaluiert und gegebenenfalls angepasst werden. Im Idealfall werden diese nicht nur einzeln angewandt, sondern mit einander kombiniert.

Zum Zeitpunkt der Entscheidung die Daten von ESCO zu verwenden lagen standen diese nur in einer frühen Version 0.8 zur Verfügung. Seit dem 28.07.2017 sind die Daten nun in überarbeiteter Struktur verfügbar und es wurden z.B. für die Informations- und Technologiebranche relevante Sektor spezifische Kompetenzen hinzugefügt.

Die Europäische Kommission ist momentan die einzige Organisation, welche Kompetenzen auf Basis eines Standards(EQF) als maschinenlesbares Format veröffentlicht. Wenn andere Organisation in Zukunft diesem Beispiel folgen, stellt sich die Frage ob ein einheitliches Format notwendig ist damit ein Empfehlungssystem, wie das in dieser Arbeit entwickelte, damit umgehen kann. In diesem Fall könnte der Lösungsansatz des inLOC Projektes verwendet werden.

9 Fazit

Was wurde erreicht?

Das Ergebnis dieser Arbeit ist ein Prototyp für einen Empfehlungsdienst, der über eine Web API erreichbar und an den eine Graphendatenbank angebunden ist. Kompetenzen die von anderen Organisationen als *Linked Data* veröffentlicht werden, können in die Datenbank importiert werden. Momentan bietet lediglich die EU Kommission mit ESCO eine Ontologie mit Kompetenzen in einem Maschinenlesbaren Format zum Download an.

Eine wichtige Erkenntnis ist genau dieses Veröffentlichungsformat. Da Kompetenzen oder Kompetenzen aus Sektorspezifischen Frameworks, sich gut in Ontologien darstellen lassen, können Metriken und Algorithmen die sich mit Ontologien befassen untersucht werden, und geprüft werden ob sich diese auch auf Kompetenzontologien anwenden lassen.

Mit der Neo4j Datenbank wurde ein wichtiges Werkzeug gewählt welche nicht nur die Aufgabe des Datenspeichers übernimmt, sondern auch einige relevanten Graphen Operationen implementiert. Mit der mächtigen Abfragensprache *Cypher* können neben kürzesten Pfaden auch Adjazenzmatrizen gebildet werden und Beziehungen können genau wie Knoten mit Attributen versehen werden, was die Gewichtung von Kanten ermöglicht.

Eine entscheidende Komponente für das erfolgreiche importieren von Kompetenzen in die Neo4j Datenbank war das RDF Import Plugin von Jesús Barassa. Da sich für die Veröffentlichung von Linked Data Formate wie RDF und JSON-LD durchgesetzt haben, jedoch für die Traversierung dieser Graphen und die Implementierung von APIs diese Technologie weniger eignet, ist eine Überführung der Triples in eine Graphendatenbank wie Neo4j für diese Arbeit essentiell gewesen.

Wie kann dieser Dienst genutzt werden?

Karriereportale können mit Hilfe des Empfehlungssystems ihre Treffergenauigkeit bei vorgeschlagenen Stellenangeboten oder Job-Matching erhöhen. Die Möglichkeiten erstrecken sich im gesamten Personalentwicklungsbereich und Kompetenzmanagement, so wäre auch eine Nutzung des Dienstes durch Systeme denkbar die Karrierepfade zeigen sollen.

Das in dieser Arbeit verwendete Datenmodell basiert auf der Ontologie des ESCO Katalogs, welcher als Linked Data veröffentlicht wird. Der Focus von ESCO liegt auf der Interoperabilität am europäischen Arbeitsmarkt und an der Übersetzung von nationalen Standards. Bis Ende 2017 soll der Katalog stetig weiter entwickelt und aktualisiert werden. Mitgliederstaaten sollen auch dabei unterstützt werden ihre nationalen Standards in maschinenlesbaren Formaten zur Verfügung zu stellen.

Ob das ESCO Datenmodel der wissenschaftliche korrekte Ansatz ist, um Kompetenzen in diesem Kontext zu modellieren wurde in dieser Arbeit nicht erörtert. Der Fokus lag auf der Vergleichbarkeit von Kompetenzen und der Entwicklung eines Empfehlungssystems. Dennoch ist das Datenmodel eine wichtige Grundlage für die Implementierung von Algorithmen und Metriken. Mit einer Änderung der Struktur des ESCO Modells, geht eventuell auch eine Änderung des Empfehlungssystems einher. Diese Abhängigkeit ist eine Schwäche die in Zukunft weiter diskutiert werden kann.

Karriereplattformen bieten schon jetzt die Möglichkeit elektronische Lebensläufe oder Profile anzulegen, in welchen Benutzer dann ihre Abschlüsse, Qualifikationen und Fertigkeiten angeben. Diese Lebensläufe und Profile werden dann mit Anforderungen von Jobinseraten verglichen. Für die Repräsentation von eigenen Kompetenzen ist eine Verknüpfung mit anderen Technologien wie zb. dem Mozilla Open Badge Network denkbar. Damit würden erworbene Kompetenzen nicht nur visuell darstellbar, sondern es gäbe auch eine Referenz auf einen Nachweis für die Kompetenz. Openbadges in CVs für Jobportale.

Literaturverzeichnis

- [acm] *About the ACM Organization.* <http://www.acm.org/about-acm/about-the-acm-organization>
- [Ade] ADECCO: *Arbeitsmarkt: Stellenangebote in Deutschland 2016 nach Berufsgruppen | Statistik.* <https://de.statista.com/statistik/daten/studie/693192/umfrage/offene-stellenangebote-in-deutschland-nach-berufsgruppen/>
- [all] *Expanding Education and Workforce Opportunities Through Digital Badges.* <http://all4ed.org/reports-factsheets/expanding-education-and-workforce-opportunities-through-digital-badges/>
- [Ana17] ANADIOTIS, George: *Graph databases and RDF: It's a family affair.* <http://www.zdnet.com/article/graph-databases-and-rdf-its-a-family-affair/>. Version: May 2017
- [BE15] BAUER, Horst P. ; ENDERLE, Inga: *Karriere durch Fachschulen für Betriebswirtschaft - Langzeituntersuchung zur beruflichen Entwicklung der Absolventen.* Berlin Heidelberg New York : Springer-Verlag, 2015. – 39–40 S. – ISBN 978–3–658–09367–9
- [bib17] *Definition und Kontextualisierung des Kompetenzbegriffes.* <https://www.bibb.de/de/8570.php>. Version: Jun 2017
- [Dor14] DORN, Jürgen: Elektronische Abzeichen zur Unterstützung von informalem Lernen. In: *Informatik-Spektrum* 37 (2014), Nr. 4, 329–332. <http://dx.doi.org/10.1007/s00287-014-0796-2>. – DOI 10.1007/s00287-014-0796-2. – ISSN 1432–122X
- [esc] *ESCO Ontology.* <https://ec.europa.eu/esco/resources/data/static/model/html/model.xhtml>. – bezieht sich auf veraltete Version 1.0
- [gro16] *MovieLens.* <https://grouplens.org/datasets/movielens/>. Version: Oct 2016
- [HKRS07] HITZLER, Pascal ; KRÖTZSCH, Markus ; RUDOLPH, Sebastian ; SURE, York: *Semantic Web - Grundlagen.* 1. Aufl. Berlin Heidelberg New York : Springer-Verlag, 2007. – ISBN 978–3–540–33994–6
- [Hö16] HÖLSCHER, Denny: *Mit XING zum Traumjob: Wie gut ist das Job-Matching des Karrierenetzwerks?* <https://hr-blog.prosoft.net/mit-xing-zum-traumjob-wie-gut-ist-das-job-matching-des-karrierenetzwerks/>. Version: Dec 2016
- [inl] *A brief overview of the InLOC Model.* <http://www.cetis.org.uk/inloc/Overview%2band%2bOrientation>
- [JW02] JEH, Glen ; WIDOM, Jennifer: SimRank: A Measure of Structural-context Similarity. In: *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* New York, NY, USA : ACM, 2002 (KDD '02). – ISBN 1–58113–567–X, 538–543

- [KN12] KRUMKE, Sven O. ; NOLTEMEIER, Hartmut: *Graphentheoretische Konzepte und Algorithmen* -. 3. Aufl. Berlin Heidelberg New York : Springer-Verlag, 2012. – ISBN 978-3-834-82264-2
- [Kon16] KONERT, Johannes: Open Badge Network Proposal on Competency Alignment and Directory. (2016). http://www.openbadgenetwork.com/wp-content/uploads/2016/01/OBN03-A2__CompetencyAlignmentAndRepository__proposal_v1.1.pdf
- [ld] *Linked Data*. <https://www.w3.org/DesignIssues/LinkedData.html>
- [Lop17] LOPEZ, Benjamin: *Entwicklung und Evaluation eines e-Kompetenz-Verzeichnisses mit REST-API und eines automatisierten Crawlers zur Datensammlung*. 2017
- [mov a] *MovieLens*. <https://movielens.org/movies/110102>
- [mov b] *Non-commercial, personalized movie recommendations*. <https://movielens.org/>
- [nea] *2.1. System requirements - Chapter 2. Installation*. <https://neo4j.com/docs/operations-manual/current/installation/requirements/>
- [neob] *Neo4j Graph Database Sandbox - Get Started with Graphs*. <https://neo4j.com/sandbox-v2/>. – kein direkte Link zur Instanz. Diese wird erst nach einem Login nur temporär erzeugt. Die Url sieht dann etwa wie folgt aus: <https://xx-x-x-x-xxxxxx.neo4jsandbox.com/>
- [omd] *OMDb API*. <http://www.omdbapi.com/>
- [rec] *RecSys – ACM Recommender Systems*. <https://recsys.acm.org/>
- [RRSK10] RICCI, Francesco ; ROKACH, Lior ; SHAPIRA, Bracha ; KANTOR, Paul B.: *Recommender Systems Handbook* -. 2011. Aufl. Berlin Heidelberg : Springer Science und Business Media, 2010. – ISBN 978-0-387-85820-3
- [RWE15a] ROBINSON, Ian ; WEBBER, Jim ; EIFREM, Emil: *Graph Databases - New Opportunities for Connected Data*. Sebastopol : "O'Reilly Media, Inc.", 2015. – ISBN 978-1-491-93086-1
- [RWE15b] ROBINSON, Ian ; WEBBER, Jim ; EIFREM, Emil: *Graph Databases Ian Robinson; Jim Webber; Emil Eifrem*. O'Reilly and Associates, 2015
- [w3c] *VOCABULARIES*. <https://www.w3.org/standards/semanticweb/ontology#summary>
- [WP94] WU, Zhibiao ; PALMER, Martha: Verbs Semantics and Lexical Selection. In: *Proceedings of the 32Nd Annual Meeting on Association for Computational Linguistics*. Stroudsburg, PA, USA : Association for Computational Linguistics, 1994 (ACL '94), 133–138

Anhang

9.0.1 API Endpunkte

