

Logo 1

Logo2

Beuth Hochschule für Technik

Fachbereich 6

Bachelorarbeit

im Studiengang Medieninformatik - Schwerpunkt Medieninformatik

zur Erlangung des akademischen Grades
Bachelor of Science

Thema:	Empfehlungssystem für Kompetenzen
Autor:	Valentin Risch s55698@beuth-hochschule.de MatNr. 798906
Version vom:	26. Juni 2017
1. Betreuer:	Prof. Dr. Johnanes Konert
2. Gutachter:	Prof. Dr. Y

Zusammenfassung

Abstract

Inhaltsverzeichnis

Abbildungsverzeichnis	5
Tabellenverzeichnis	5
Listingverzeichnis	5
Abkürzungsverzeichnis	5
1 Einleitung	6
2 Grundlagen	7
2.1 Kompetenzen	7
2.2 Kompetenzframeworks	7
2.3 RDF, Semantic Web Linked Data	8
2.3.1 Semantic Web	8
2.3.2 Linked Data	8
2.3.3 RDF	8
2.4 Graphen	9
2.4.1 Graphentheorie	9
2.4.2 Graphendatenbanken	9
2.4.3 Neo4J	10
2.5 Empfehlungssysteme	10
3 Verwandte Themen und Arbeiten	10
3.1 Openbadge	10
3.2 European e-Competence Framework	11
3.3 InLoc	11
3.4 ESCO	11
3.4.1 GraphGist: Recommendation System Sandbox	12
3.5 RDF Import Neo4J	12
3.6 Recommender Systeme	13
3.6.1 RecSys Challenge	13
3.6.2 GraphGist Recommendation Engine Neo4j	13
4 Analyse und Aufgabenstellung	13
4.1 Aufgabenstellung	13
4.2 Anforderungsanalyse	13
4.2.1 Kompetenzmodell	13
4.2.2 Anforderungen an die Engine	13
4.2.3 Anforderungen an die Datenbank	14
4.3 Aufgabenstellung	14
5 Entwurf	14
5.1 Format/Kompetenzmodell	14
5.1.1 ESCO RDF Import	14
5.1.2 inLOC	15
5.1.3 Datenbank	15
5.1.4 API	15
5.2 Datenmodell	16

5.3	Anfragen Formulieren	16
5.4	Algorithmen	16
6	Ausblick	16
7	Fazit	16
	Literaturverzeichnis	18
	Anhang	19
	Eidesstattliche Erklärung	19

Abbildungsverzeichnis

1	Beschreibung	15
2	Beispiel einer Bildbeschreibung	16
3	Beschreibung	16

Tabellenverzeichnis

Listingverzeichnis

1	Das Listing zeigt einen Funktionsaufruf über die Neo4j	12
2	Das Listing zeigt Java Quellcode	16

Abkürzungsverzeichnis

LOD Linked Open Data

1 Einleitung

Mit dem Abschluss einer jeden Berufsausbildung oder eines Studiums beginnt auch der Einstieg in das Berufsleben. Für Viele Bewerber bedeutet das, vermeintlich 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. Er-

langte Fähigkeiten und Wissen werden im allgemeinen durch Abschlüsse, Zertifikate oder Urkunden bestätigt. Jeder Arbeitnehmer hat einen oder mehrere Abschlüsse die ihn für einen bestimmten Beruf qualifizieren. Leider unterscheiden sich die Bezeichnungen dieser Abschlüsse, beinhalten aber oft dieselben oder ähnliche Kompetenzen. So gibt es neben dem normalen Informatikstudium auch Angebote wie:

- Wirtschaftsinformatik
- Angewandte Informatik
- Medizininformatik
- Bioinformatik
- Medieninformatik
- Technische Informatik
- Navigation und Umweltrobotik

Diese Studiengänge haben zwar unterschiedliche Inhalte, bauen aber auf denselben Grundlagen wie zb. Mathematik, Algorithmik, Datenbanken etc. auf.

2 Grundlagen

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.[bib17]

Oftmals werden die beiden Begriffe Kompetenz und Qualifikation als Synonyme verwendet, jedoch gibt es signifikante Unterschiede in deren Bedeutung.

Kompetenz(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.

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

Fachkompetenzen Beispiele für Kompetenzen und Qualifikationen:

2.2 Kompetenzframeworks

Kompetenzen abzubilden gestaltet sich als 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 nicht 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, ergebenben sich dann Vergleichsmöglichkeiten.

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

Zusammenfassung:

"Kompetenzrahmen sind eine wichtige Brücke zwischen dynamischer Kompetenz und statischen Bildungs- und Zertifikatssystemen. Denn Kompetenzrahmen machen Kompetenz klassifizierbar und unterstützen ihre Anerkennung und Anrechnung."

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. “Finde Wege und Methoden, Informationen so zu repräsentieren, dass Maschinen damit in einer Art und Weise umgehen können, die aus menschlicher Sicht nützlich und sinnvoll erscheint.”[HKRS07, 12]

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 Europäische 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-Lee 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

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. Zb. Bücher, Orte, Menschen, abstrakte Konzepte usw. 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”.

Vokabu-
bzw.
Ontolo-
gien

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

In diesem Teilgebiet der Mathematik werden Graphen und ihre Beziehungen zueinander untersucht. Das älteste Dokumentierte Problem der Graphentheorie 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.

In der Informatik

graphentheoretische Konzepte zitieren Seite4

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 Model sehr intuitive ist und sich einfach auf Papier modellieren lässt. Ein weiterer wichtiger Aspekt ist der Performance Vorteil bei 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 *Labeled Property Graph Model* implementiert. Knoten im Graph können durch Labels in Kategorien eingeordnet werden und speichern Informationen als Schlüssel-Wert Paare in Properties. Neben Knoten können auch Kanten Properties erhalten, was eine Gewichtung der Beziehungen zwischen Knoten ermöglicht. 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 Inhalts basierte Filtern und Kollaborative Filtern.

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 European e-Competence Framework

Der europäische Kompetenzrahmen für Fach- und Führungskräfte der Informations- und Kommunikationstechnologiebranche, ist eine sektor-spezifische Umsetzung des Europäischen Qualifikationsrahmens EQR.

3.3 InLoc

Das Europäische Projekt InLOC(Integrating Learning Outcomes and Competences) erlaubt es Kompetenzen und Lernergebnisse verschiedener Kompetenzrahmen in einem einheitlichen semantischen Format abzubilden.

3.4 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.

Ein 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.

Wichtig
Mehr
Infor-
ma-
tionen
ausar-
beiten

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.

3.4.1 GraphGist: Recommendation System Sandbox

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".[neo] 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. Ein Beispiel für eine Metrik zum finden von Ähnlichkeiten ist der Jaccard Index. Er ist genau dann 1, wenn zwei Mengen identisch sind. Je weniger Gemeinsame Elemente zwei Mengen besitzen, desto kleiner wird der Jaccard Index und wird 0, wenn keine Gemeinsamkeiten vorliegen.

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

3.5 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.

3.6 Recommender Systeme

3.6.1 RecSys Challenge

3.6.2 GraphGist Recommendation Engine Neo4j

4 Analyse und Aufgabenstellung

Open Badges können über das criteria und alignment field mit Kompetenzen aus externen Kompetenzframeworks verlinkt werden. Daraus ergibt sich die Anforderung Badges mit ähnlichen Kompetenzen, Badges die sich ergänzen oder identisch sind zu ermitteln, aber auch fehlende Kompetenzen zu identifizieren die benötigt werden um einen gewünschten Badge zu erlangen. [Kon16]

Problematisch sind die verschiedenen Standards der einzelnen Kompetenzframeworks. So gibt es nicht nur nationale und sprachliche Unterschiede sondern auch Unterschiede in den veröffentlichten Formaten. Von den bisher veröffentlichten Kompetenzframeworks bietet momentan nur ESCO den Download der Daten in einem maschinenlesbaren Format(CSV, XML, RDF) an.

4.1 Aufgabenstellung

Diese Arbeit soll Methoden und Metriken evaluieren um Kompetenzen in einem Graphen zu modellieren und Ähnlichkeiten festzustellen. Die Kompetenzen können einem Kompetenzverzeichnis entnommen werden welches Kompetenzen aus verschiedenen Kompetenzrahmen verwaltet. Eine andere Herangehensweise wäre mit jedem neuen Eintrag im Kompetenzverzeichnis auch einen neuen Eintrag in den Graphen vorzunehmen.

Das Ergebnis dieser Arbeit ist ein Prototyp eines Empfehlungssystems, welches folgende Anforderungen erfüllen sollte.

4.2 Anforderungsanalyse

4.2.1 Kompetenzmodell

???

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

eindeutige ID für eine Kompetenz sein, als Ausgabe soll eine oder mehrere Kompetenzen in sortierter Reihenfolge geliefert werden.

- Ähnliche Kompetenzen finden
- Identische Kompetenzen finden
- Wichtigkeit von Kompetenzen anzeigen
- Inkludierende Kompetenzen ermitteln
- Lücken von Kompetenzen finden

Optionale Anforderungen

- Vergleich von Kompetenzen aus verschiedenen Frameworks

4.2.3 Anforderungen an die Datenbank

4.3 Aufgabenstellung

5 Entwurf

5.1 Format/Kompetenzmodell

Für die Modellierung der Kompetenzen gibt es mehrere Möglichkeiten.

5.1.1 ESCO RDF Import

ESCO steht als RDF Graph zur mit einem eigenen Vokabular zur Verfügung und kann mittels RDF Importerin in die Neo4j Instanz geladen werden. Dabei werden nach den ESCO Ontologien folgende interessante Knoten und Beziehungstypen angelegt.

ESCO concepts Relationship

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

Beziehung zwischen zwei ESCO Säulen tb. Occupation und Qualification.

Qualification

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

Ein offiziell oder formelles Zertifikat eines oder mehrerer Kompetenzen oder Fertigkeiten.

Kann verknüpft im Skills oder Berufen sein.

Skil

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

Jesus
Barrasa
zitieren

weitere
auf <https://ec.europa.eu/esco/>

5.1.2 inLOC

Das Projekt InLOC ermöglicht das Abbilden von allen Kompetenzrahmen in einem einheitlichen maschinenlesbaren Format(RDF,XML,JSON-LD).

5.1.3 Datenbank

Für die Persistenz der Kompetenz Daten wird eine Instanz des Neo4j Datenbank Management Systems eingesetzt. Die Vorteile

Alternative Datenbanken Warum keine RDF Technologie?

5.1.4 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

Die Anwendung ist also wenig komplex. Ein viel höherer Implementierungsaufwand ist den Vergleichsalgorithmen, bzw Graph Queries zuzurechnen. Um den Implementierungsaufwand der Applikation weiter zu reduzieren, soll das Serverless Framework für Amazon Web Services Lambda zum Einsatz kommen.

Für die erste Komponente wird eine Web Api benötigt, über welche mit einem Internetfähigen Protokoll wie zb. HTTP zugegriffen werden kann.

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.

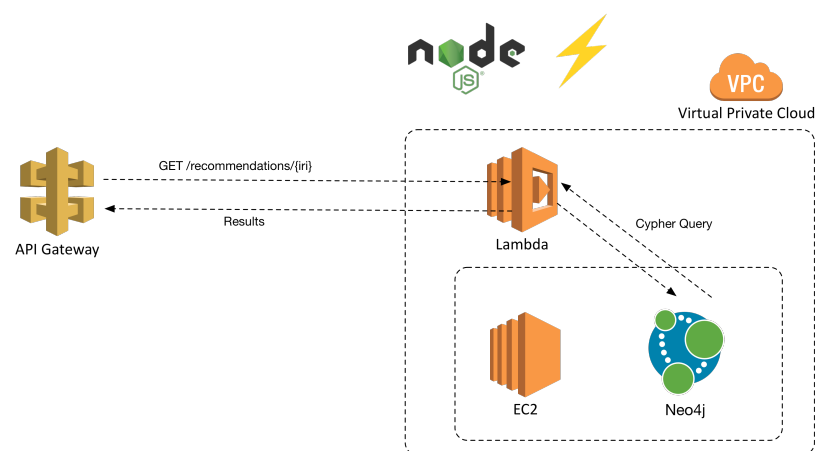


Abbildung 1: Beschreibung

5.2 Datenmodell

5.3 Anfragen Formulieren

5.4 Algorithmen

Shortest Path Dijkstra Algorithm Depth First Breadth first search

6 Ausblick

7 Fazit

Abbildung 2: Beispiel einer Bildbeschreibung¹



Abbildung 3: Beschreibung

Abbildung 3 [S.16]

Überschrift 1	Überschrift 2
Info 1	Info 2
Info 3	Info 4

```

1  /* generate TagCloud */
2  Cloud cloud = new Cloud();
3  cloud.setMaxWeight(_maxSizeOfText);
4  cloud.setMinWeight(_minSizeOfText);
5  cloud.setTagCase(Case.LOWER);
6
7  /* evaluate context and find additional stopwords */
8  String query = getContextQuery(_context);
9  List<String> contextStoplist = new ArrayList<String>();
10 contextStoplist = getStopwordsFromDB(query);
11
12 /* append context stoplist */
13 while(contextStoplist != null && !contextStoplist.isEmpty())
14     __stoplist.add(contextStoplist.remove(0));
15
16 /* add cloud filters */
17 if (__stoplist != null) {

```

¹Bildquelle: Beispielquelle


```

18 DictionaryFilter df = new DictionaryFilter(_stoplist);
19 cloud.addInputFilter(df);
20 }
21 /* remove empty tags */
22 NonNullFilter<Tag> nnf = new NonNullFilter<Tag>();
23 cloud.addInputFilter(nnf);
24
25 /* set minimum tag length */
26 MinLengthFilter mlf = new MinLengthFilter(_minTagLength);
27 cloud.addInputFilter(mlf);
28
29 /* add taglist to tagcloud */
30 cloud.addText(_taglist);
31
32 /* set number of shown tags */
33 cloud.setMaxTagsToDisplay(_tagsToDisplay);

```

Listing 2: Das Listing zeigt Java Quellcode

Die Zuordnung aller möglichen Werte, welche eine Zufallsvariable annehmen kann nennt man *Verteilungsfunktion* von X .

Die Funktion $F: \mathbb{R} \rightarrow [0,1]$ mit $F(t) = P(X \leq t)$ heißt Verteilungsfunktion von X .²

Für eine stetige ABC $X: \Omega \rightarrow \mathbb{R}$ heißt eine integrierbare, nichtnegative reelle Funktion $w: \mathbb{R} \rightarrow \mathbb{R}$ mit $F(x) = P(X \leq x) = \int_{-\infty}^x w(t)dt$ die *Dichte* oder *Wahrscheinlichkeitsdichte* der Zufallsvariablen X .³

²Konen, vgl. [?] [S.55]

³Konen, vgl. [?] [S.56]

Literaturverzeichnis

- [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/>
- [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
- [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
- [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
- [neo] *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-xxxxx.neo4jsandbox.com/>
- [omd] *OMDb API*. <http://www.omdbapi.com/>
- [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

Anhang

Eidesstattliche Erklärung

Eidesstattliche Erklärung zur <-Arbeit>

Ich versichere, die von mir vorgelegte Arbeit selbstständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.

Unterschrift :

Ort, Datum :

