

Beurteilung der Durchführbarkeit und Performance der Integration von Finanzdaten im Semantic Web

Bachelor Thesis
von
Alexander Büscher

an der Fakultät für Wirtschaftswissenschaften

In dem Studiengang
Informationswirtschaft

eingereicht am 07.11.2014 beim
Institut für Angewandte Informatik
und Formale Beschreibungsverfahren
des Karlsruher Instituts für Technologie

Referent: Prof. Dr. Rudi Studer
Betreuer: Dr. Andreas Harth

Eidesstattliche Erklärung

Ich versichere hiermit wahrheitsgemäß, die Arbeit und alle Teile daraus selbstständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderung entnommen wurde.

Karlsruhe, den 07.11.2014

Alexander Büscher

Abstract

Diese Arbeit betrachtet die Durchführbarkeit und Performance der Integration verschiedener Finanzdaten im Semantic Web. Zusätzlich wird der dadurch geschaffene Mehrwert evaluiert. Nach einer thematischen Einführung sowie der Ermittlung relevanter Kennzahlen erfolgt die Präsentation des Ansatzes in Form eines Anwendungsszenarios. Dieses präsentiert die verschiedenen Vorgehensweisen und Möglichkeiten, welche die Integration der Daten aus mehreren Quellsysteme mit sich bringt. Eine ausführliche Diskussion und Evaluation der besprochenen Gesichtspunkte rundet den Hauptteil der Arbeit ab. Als Ergebnis lässt sich eine generelle Bestätigung der verschiedenen Aspekte der aufgestellten Hypothese festhalten. Die generelle Durchführbarkeit wird vor allem durch den präsentierten Anwendungsfall bestätigt, die Performance konnte nach einigen Optimierungen während der Bearbeitung auf ein gutes Level gehoben werden und der erhöhte Nutzen für Unternehmen durch den präsentierten Ansatz wird durch die Teilnehmer der Evaluation bestätigt.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Fragestellung	2
1.3	Vorgehensweise	2
2	Thematische Einführung und Stand der Forschung	3
2.1	Einführung und Gebrauch von XBRL	4
2.1.1	Historie	4
2.1.2	Syntax und Aufbau von XBRL	5
2.2	Finanzdaten im Semantic Web	8
2.2.1	Grundlegende Ideen und Vorzüge von RDF	8
2.2.2	Mapping Ansatz von Kämpgen et al.	9
2.2.3	Mapping Ansatz von R. García und R. Gil	10
2.2.4	Mapping Ansatz von Bao et al.	12
2.2.5	Betrachtung der Ansätze	13
2.3	Ergebnisse	13
3	Entwicklung und Herleitung der Hypothese	15
4	Betrachtung der Kennzahlenauswahl	16
4.1	Grundlage zur Veröffentlichung von Daten	16
4.2	Das Dilemma der Veröffentlichungen	16
4.3	Auswahl von Kennzahlen für das Anwendungsszenario	17
4.4	Zusammenfassung und Klassifizierung	21
5	Anwendungsfall	23
5.1	Abgrenzung ISIC und SIC	23
5.2	Bestimmen von Unternehmen der gleichen Klassifizierung	24
5.2.1	Generelles Vorgehen zur Unternehmensbestimmung	24
5.2.2	Probleme bei der Bestimmung von Unternehmen der gleichen Branche	25
5.3	Anforderungen	26

5.4	Wahl und Beschreibung des Anwendungsfalls	27
5.4.1	Auswahl und Begründung	27
5.4.2	Datenintegration der gewählten Branche	28
6	Umsetzung	29
6.1	Assets (oder totalAssets) - Aktivposten der Bilanz	30
6.2	Unternehmenseffizienz – Gesamtumsatz pro Mitarbeiter	33
6.3	Unternehmenseffizienz – Pro-Kopf-Gewinn	38
6.4	Umsatz pro Geschäftsstelle	41
6.5	Effektivität der Produktion	45
6.6	Erzielte Gewinne / Verluste über mehrere Perioden	48
6.7	Präzision von Angaben verschiedener Quellen	50
6.8	Aufbau des Gesamtunternehmens	53
6.9	Entwicklung des Aktienkurses	57
7	Ansätze zur weiteren Forschung	62
7.1	Globale Anwendbarkeit	62
7.2	Veröffentlichung eigener XBRL-Konzepte / Nicht-einheitliche Veröffentli- chungen	63
7.3	Präzision der Daten vs. Automatisierbarkeit	64
7.4	Diskrepanz zwischen verschiedenen Publikationsformen der Finanzberichte	65
7.5	Visualisierbarkeit	66
8	Diskussion und Evaluation	67
8.1	Diskussion	67
8.2	Evaluation des Nutzens und Analyse der Performance	70
8.2.1	Beschreibung der Grundgesamtheit der Umfrage	70
8.2.2	Methodologie	70
8.2.3	Evaluation der Umfrageergebnisse	71
8.2.4	Performance verschiedener Abfragen	76
9	Fazit	78

A	Anhang	80
A.1	Finanzkonzepte Abbildung 2	80
A.2	80 SIC Gruppen innerhalb der Datenbank	81
A.3	Ergebnisse SPARQL–Abfragen	84
A.3.1	Abfrage Ergebnisse zu Abbildung 19	84
A.3.2	Abfrage Ergebnisse zu Abbildung 22	84
A.4	SPARQL–Abfragen	85
A.5	Umfrage	89
A.6	Auswertungen der Befragungen	92

Abkürzungsverzeichnis

10-K	Annual Finance Report
10-Q	Quarterly Finance Report
BWL	Betriebswirtschaftslehre
CIK	Central Index Key
CSV	Comma Separated Values
CURIE	compact URI
DEA	Data Envelopment Analysis
DUNS	Data Universal Numbering System
EDGAR	Electronic Data Gathering, Analysis, and Retrieval
FFIEC	Federal Financial Institutions Examination Council
GAAP-US	Generally Accepted Accounting Principles - (US)
GUI	Graphical User Interface
HGB	Handelsgesetzbuch
iCal	RDF Vokabular für iCalender Events
ISIC	International Standard Industrial Classification
MR	misresponse to reversed items
NEC	Not elsewhere classified
NAICS	North American Industry Classification System
OWL	Web Ontology Language
QB	RDF Vokabular für Datensatz Ausdrücke
RDF	Resource Description Framework
RDFS	RDF Schema
ROA	Return on Assets
ROE	Return on Earnings
SEC	United States Securities and Exchange Commission
SIC	Standard Industrial Classification
SPARQL	SPARQL Protocol and RDF Query Language
URI	Uniform resource identifier
XBRL	eXtensible Business Reporting Language
XSD	XML Schema Definition
XML	Extensible Markup Language

Abbildungsverzeichnis

1	Datenintegration aus verschiedenen Quellsystemen, Eigene Darstellung auf Grundlage von [KWO ⁺ 14]	3
2	Darstellung Einzeltabelle aus einem Finanzbericht ohne XBRL, Mastercard (Central Index Key (CIK): 0001141391) 10-Q Bericht, 2014-07-31	4
3	Codebeispiel (1) eines financial facts (Beispiel: Mastercard)	5
4	Codebeispiel (2) eines financial facts (Beispiel: Mastercard)	5
5	Codebeispiel zu genaueren Informationen eines contextRef Verweises	6
6	Codebeispiel eines eigenen Taxonomie Tags (Beispiel: Mastercard)	6
7	Codebeispiel eines eigenen Taxonomie facts (Beispiel: Mastercard)	7
8	Generelle Darstellung eines financial facts	7
9	Generelle Darstellung eines eigenen Taxonomie facts	8
10	Mapping Beispiel auf Grundlage von [KWO ⁺ 14]	10
11	Mapping Beispiel (Triple-Form) auf Grundlage von [KWO ⁺ 14]	10
12	Mapping Beispiel auf Grundlage von [GG10]	11
13	Mapping Beispiel eines Taxonomie-Dokuments auf Grundlage von [BRLD10]	12
14	Mapping Beispiel eines Instanz-Dokuments auf Grundlage von [BRLD10]	12
15	Benutzte Präfixe der zusammenfassenden Tabelle	22
16	Zusammenhang CIK und SIC in der SEC Datenbank – Visualisierte Darstellung der RDF-Triple	24
17	SPARQL-Abfrage der SIC bei gegebenem CIK	25
18	SPARQL-Abfrage aller Unternehmen in einer SIC-Gruppe	25
19	SPARQL-Abfrage für alle Unternehmen assoziiert zur SIC 6022	28
20	Regel zur Abfrage von Daten mit gleicher Bezeichnung aus mehreren GAAPs	29
21	rdfs:subClassOf-Beziehungen (allgemein)	29
22	SPARQL-Abfrage Assets und dazugehörige Banken	30
23	SPARQL-Abfrage Assets der Fifth Third Bancorp	31
24	Verlauf des Wertes der Assets der Fifth Third Bancorp	32
25	Verbinden der Datenquellen SEC und DBpedia (Beispiel: Fifth Third Bank)	33
26	Ermittlung des Jahresumsatzes pro Mitarbeiter am Beispiel der Fifth Third Bank 2009	34

27	Deklaration mehrerer properties zu gleichen Informationen zu einem Typ (Beispiel: Mitarbeiteranzahl)	35
28	Auswahl des deklarierten Typs in Abfragen (Beispiel: Mitarbeiteranzahl) .	35
29	Konkurrenzanalyse im Bereich des Gesamtumsatzes pro Mitarbeiter	36
30	Ermittlung des Jahresgewinns pro Mitarbeiter am Beispiel der Fifth Third Bank 2009	38
31	Konkurrenzanalyse im Bereich des Jahresgewinns pro Mitarbeiter	39
32	Abfrage Umsätze sowie Anzahl der Standorte (Beispiel: Fifth Third Bank)	41
33	Entwicklung der Kennzahl $\frac{UmsatzPeriode(t)}{AnzahlStandorte}$ 2009 - 2011 (Beispiel: Fifth Third Bank)	42
34	Entwicklung der Kennzahl $\frac{UmsatzPeriode(t)}{AnzahlStandorte}$ 2009 - 2011 (Beispiel: Fifth Third Bank)	43
35	Abfrage Personalaufwand und nicht zu verzinsendes Einkommen (Beispiel: Wilmington Trust Corp.)	45
36	Abfrage Personalaufwand und nicht zu verzinsendes Einkommen (Beispiel: Fifth Third Bancorp)	46
37	Entwicklung der Kennzahl $\frac{nonInterestIncome}{personnelCosts}$ 2009 - 2010 (Beispiel: Wilming- ton Trust Corp.)	47
38	Abfrage Gewinne (Verluste) einzeln (Beispiel: Fifth Third Bank)	48
39	Allgemeine Berechnung von Wertveränderungen über mehrere Perioden . .	49
40	Vergleich von Angaben zum netIncome von DBpedia und SEC im Jahres- bericht 2010 der Fifth Third Bancorp	51
41	Vergleich von Angaben zum netIncome von DBpedia und SEC im Jahres- bericht 2010 der Fifth Third Bancorp mithilfe des ASK Operators	52
42	Abfrage zu zugehörigen Unternehmen der Gruppe, zusammen mit deren Hauptsitz und Anzahl der Mitarbeiter	54
43	Abfrage zu zugehörigen Unternehmen der Gruppe, zusammen mit deren Hauptsitz und Anzahl der Mitarbeiter (Beispiel: Wilmington Trust)	55
44	Verbindung der Yahoo! Finance Daten in RDF mit dem datastore (Beispiel: Fifth Third Bancorp)	57
45	Abfrage von notierten Werten der Fifth Third Bancorp bei Börsenschluss über das Jahr 2009	58
46	Verlauf des Aktienkurses 2009 (Beispiel: Fifth Third Bancorp)	59

47	Abfrage zur Änderung des Aktienkurses im Jahre 2009 der Fifth Third Bancorp	60
48	Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich Semantic Web	72
49	Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich BWL Finanzwirtschaft & Rechnungswesen	72
50	Auswertung der Evaluationsergebnisse: Angaben zur Komplexität bzw. dem Verständnis bzgl. der Abfragen (1)	73
51	Auswertung der Evaluationsergebnisse: Angaben zur Komplexität bzw. dem Verständnis bzgl. der Abfragen (2)	73
52	Auswertung der Evaluationsergebnisse: Angaben zur Relevanz und dem geschätzten Mehrwert für Unternehmen	74
53	Auswertung der Evaluationsergebnisse: Angaben zur Individualisierbarkeit der Abfragen	75
54	Financial facts aus Abbildung 2 1/4 (Beispiel: Mastercard)	80
55	Financial facts aus Abbildung 2 2/4 (Beispiel: Mastercard)	80
56	Financial facts aus Abbildung 2 3/4 (Beispiel: Mastercard)	80
57	Financial facts aus Abbildung 2 4/4 (Beispiel: Mastercard)	80
58	Beide Abfragen zu Kennzahl 2 innerhalb einer Query	85
59	Beide Abfragen zu Kennzahl 3 innerhalb einer Query	86
60	Modifizierte Abfragen zu Kennzahl 4 (Jahr 2010)	87
61	Modifizierte Abfragen zu Kennzahl 4 (Jahr 2011)	88
62	Altersverteilung der Teilnehmer	92
63	Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich Informatik	92
64	Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich BWL allg.	93
65	Auswertung der Evaluationsergebnisse: Angaben zur Einschätzung, ob Unternehmen den präsentierten Ansatz benutzen	93
66	Auswertung der Evaluationsergebnisse: Angaben zur Einschätzung, ob durch den Use Case alle relevanten Vorgehensweisen abgedeckt wurden	94
67	Auswertung der Evaluationsergebnisse: Angaben zur Darstellung und Interpretationsmöglichkeit der Ergebnisse	94

1 Einleitung

Die Masse an verschiedensten Informationen im Bereich von Finanzen kann alle Beteiligten vor große Probleme stellen: Aktiengesellschaften müssen – in der Regel vom Gesetzgeber dazu verpflichtet – durch das Offenlegen von Quartals- und Jahreszahlen Transparenz für ihre Teilhaber schaffen. Hierbei genügt keine oberflächliche Darstellung der Entwicklung von zwei bis drei Kennzahlen, wie beispielsweise Gewinn und Umsatz, vielmehr wird eine extrem detaillierte Darstellung des kompletten Geschäftsjahres bzw. der betrachteten Periode verlangt.

Noch größere Probleme haben jedoch die Finanz-Analysten. Um ein Unternehmen auf “Herz und Nieren“ prüfen und hierfür eine vollständige Analyse anfertigen zu können, müssen Analysten auf Daten verschiedenster Quellsysteme zugreifen und diese verwerten können. Um eine Data Envelopment Analyse (DEA) ([Wei01], [GS14]), die zu den alltäglichen Werkzeugen eines Analysten zählt, durchführen zu können, werden wirtschaftliche Inputs und Outputs im Bezug auf das Unternehmen benötigt, die aus verschiedenen Quellsystemen stammen können – genau diese gilt es zu vereinheitlichen. Ferner reicht es heute nicht mehr, ausschließlich standardisierte Kennzahlen, wie beispielsweise “Return on Invest“ (ROI), zu berechnen und anhand dieser das Unternehmen zu beurteilen. Ebenfalls ist es wichtig, den Gesamtaufbau des Unternehmens mit Tochterunternehmen zu kennen, um grenzübergreifend, zügig und effektiv auf Ereignisse reagieren zu können. Dieses Problem illustrieren auch O’Riain et al. in ihrer Publikation aus dem Jahre 2012 ([OCH12]) gut.

1.1 Motivation

Das Beispiel des Bedarfs der Finanz-Analysten zeigt, wie wichtig es ist, verschiedenste Daten einheitlich darzustellen, um diese in vollem Umfang nutzen zu können. Erste Abhilfe hierzu schaffte die Einführung eines einheitlichen, maschinenlesbaren Formats, in welchem zukünftig Finanzdaten publiziert werden sollen – die eXtensible Business Reporting Language (XBRL) ([HS01]). Da dieses Format jedoch lediglich die Darstellung von Finanzberichten einführte, besteht weiterhin das Problem bezüglich Daten verschiedener Quellsysteme in unterschiedlichen Dateiformaten, die nichtsdestotrotz alle für eine Auswertung gebraucht werden (bspw. die klassische Kennzahl des Pro-Kopf-Umsatzes). Der große Sprung zur Vereinheitlichung von Daten aus allen Quellsystemen kann nun mithilfe des Resource Description Framework (RDF) bewerkstelligt werden ([Mil98]).

1.2 Fragestellung

Fraglich ist nun, ob Aufwand und Ertrag in einem angemessenen Verhältnis stehen. Das Umformen von Daten aus verschiedenen Quellen in eine kanonische Form ist eine aufwändige und extrem detaillierte Arbeit. Zusätzlich erschwert wird diese Arbeit durch die Möglichkeit, immer weitere Daten aus anderen, bisher nicht bedachten Systemen in die Analyse einfließen zu lassen. Insgesamt soll bei der Fragestellung zwar in erster Linie auf die generelle Funktionalität geachtet werden, aber auch festgestellt werden, ob eine derartige Analyse hinsichtlich wirtschaftlicher Aspekte wie beispielsweise Performance (vor allem im Hinblick auf die Dauer des Abfragevorgangs) überhaupt lohnenswert ist.

1.3 Vorgehensweise

Die vorliegende Arbeit betrachtet diese Fragestellung gründlich. Anfangs wird der Leser auf den aktuellen Stand der Forschung hingewiesen und an grundlegende Ideen der Thematik herangeführt. Es werden verschiedene Fachbegriffe und Konzepte, die später sowohl im Anwendungsfall als auch generelle Vorgehensweise wichtig sind, genannt und erläutert. Ferner wird die Auswahl der im Anwendungsfall betrachteten und als wichtig erachteten Kennzahlen genannt und erläutert.

Nach der Präzisierung der Fragestellung wird auch ein detailliert beschriebener Anwendungsfall vorgestellt. In diesem soll die Thematik, spezialisiert auf einen speziellen Sektor, getestet werden. Mithilfe von Abfragen in der SPARQL Protocol and RDF Query Language (SPARQL) werden branchenspezifische Auswertungen getätigt und betrachtet. Basierend auf diesen Abfragen wird eine generalisierte, branchen- und unternehmensunabhängige Vorgehensweise beschrieben. Aufgrund der Ergebnisse soll eruiert werden, ob die vereinheitlichte Darstellung von Daten aus allen Datenquellen in RDF in Kombination mit Auswertungen der Abfragesprache SPARQL eine Erleichterung zur Analyse von Unternehmen aufgrund öffentlich zugänglicher Daten im Bereich Finanzen bietet.

Basierend auf den Funden des Hauptteils findet die thematische Diskussion statt. Es werden Auswertungen des Anwendungsfalls und der generalisierten Form der Abfragen mit einbezogen, um anschließend eine Aussage darüber tätigen zu können, ob und in welchem Umfang die gestellte Hypothese bestätigt werden konnte. Zusätzlich werden ausführlich Vor- und Nachteile der angewandten Methode beleuchtet und ausgewertet. Es folgt die Verifizierung bzw. Falsifizierung der in Kapitel 3 hergeleiteten Hypothese. Diese wird auf der einen Seite durch eine Anwendungsbeurteilung in Form der Auswertung eines Fragebogens und auf der anderen Seite mit einer Analyse der Abfrageperformance durchgeführt. Die Arbeit wird durch einen Ausblick abgerundet, welcher verrät, wo zukünftig die nächsten Schritte zu setzen sind und wo noch Bedarf zur Nacharbeit besteht.

2 Thematische Einführung und Stand der Forschung

Bevor es Sinn macht, einen Anwendungsfall und eine generalisierte Form des aufgestellten Problems einzuführen, ist es unumgänglich eine angemessene Wissensgrundlage beim Leser zu schaffen. Ferner gilt es, die bereits geleistete Arbeit zusammenzufassen, um anschließend nahtlos von dieser ausgehend den Anwendungsfall und die eventuell entstehenden Problematiken aufzuzeigen. Abbildung 1 illustriert hierbei das Zusammenspiel der verschiedenen Quellsysteme auf dem Weg zu einer vollständigen Auswertung und fasst die beschriebenen Schwerpunkte des Kapitels zusammen.

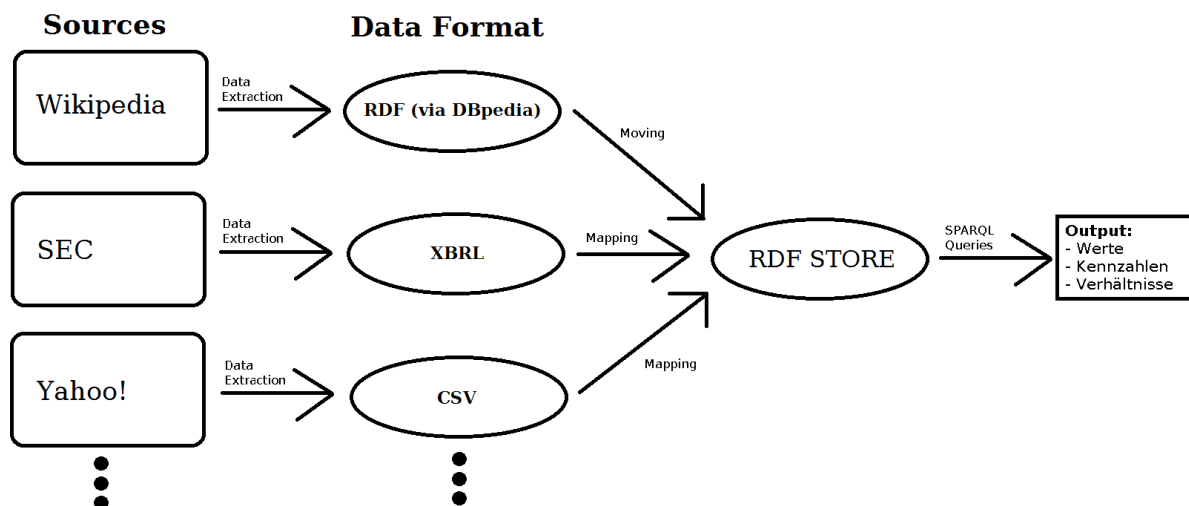


Abbildung 1: Datenintegration aus verschiedenen Quellsystemen, Eigene Darstellung auf Grundlage von [KWO⁺14]

Wie eingangs bereits beschrieben kann der Abbildung entnommen werden, dass mit verschiedenen Dateiformaten umgegangen bzw. gearbeitet werden muss. An dieser Stelle wurden zur exemplarischen Untermalung die folgenden Ressourcen für Daten gewählt:

- Wikipedia¹ (bzw. DBpedia²)
- United States Securities and Exchange Commission (SEC), deren Daten mithilfe des “Electronic Data Gathering, Analysis, and Retrieval“ (EDGAR³) Tools bestimmt werden
- Yahoo! Finance⁴

¹www.wikipedia.org Abruf: 09.10.2014

²<http://dbpedia.org/About> Abruf: 09.10.2014

³www.sec.gov/edgar.shtml Abruf: 09.10.2014

⁴<http://finance.yahoo.com/> Abruf: 09.10.2014

Während von der Informationsquelle DBpedia allgemeine Informationen über ein Unternehmen erhalten werden können, bietet die SEC eine Übersicht aller Finanzberichte von Privatpersonen und Firmen, die vom Gesetzgeber verpflichtet sind diese in regelmäßigen Abständen zu publizieren. Ferner können über Yahoo! Finance (Live-)Daten über den Stand aller an der Börse notierten Aktien erhalten werden. Problematisch hierbei ist, dass die Kombination der Informationen unmittelbar nach der Ermittlung der Daten aus den Quellsystemen aufgrund der verschiedenen Formate nahezu unmöglich ist.

Das Ziel dieses Kapitels ist also, beim Leser einen angemessenen Wissensstand zu schaffen, um die Ausführungen des Anwendungsfalls vollständig verstehen und interpretieren zu können.

2.1 Einführung und Gebrauch von XBRL

Um an das Kernproblem, nämlich die Vereinheitlichung der verschiedenen Datentypen zu gelangen, ist es relevant, zumindest die wichtigsten bereits bestehenden Formate genauer zu betrachten. Eines von diesen ist XBRL.

2.1.1 Historie

Die größte Menge an unterschiedlichen Informationen, bezüglich des finanziellen Stands eines Unternehmens, können den Jahres- und Quartalsberichten entnommen werden. Doch vor der Einführung von XBRL, welches auf Basis von Extensible Markup Language (XML) als einheitlichem Dateiformat für diese Veröffentlichungen, arbeitet, wurden diese in einer Vielzahl verschiedener, unzusammenhängender und nur schwer auswertbarer Tabellen veröffentlicht (siehe Abbildung 2) – wenn man Glück hatte, wurde immerhin ein Download in Form einer Excel Tabelle bereitgestellt.

Note 6. Accrued Expenses and Accrued Litigation

Accrued expenses consisted of the following:

	June 30, 2014		December 31, 2013
		(in millions)	
Customer and merchant incentives	\$ 1,258	\$	1,286
Personnel costs	265		413
Advertising	88		149
Income and other taxes	146		95
Other	146		158
Total accrued expenses	\$ 1,903	\$	2,101

Abbildung 2: Darstellung Einzeltabelle aus einem Finanzbericht ohne XBRL, Mastercard (Central Index Key (CIK): 0001141391) 10-Q Bericht, 2014-07-31

Möglicherweise war unter anderem dies ausschlaggebend, dass sich 1998 eine Gruppe

von Experten mit ihren Ideen zusammenschloss, wie eine strukturierte Darstellung der Daten von Unternehmen, speziell im Bereich Finanzen, auszusehen hätte. Nach stetiger Arbeit an diesem Projekt war es dann 2009 soweit, dass die erste Taxonomie in Einklang mit den Generally Accepted Accounting Principles der Vereinigten Staaten (US-GAAP) veröffentlicht werden konnte. Diese stellt weiterhin die Grundlage vieler veröffentlichter XBRL-Konzepte dar und wird andauernd nachhaltig verbessert, um die Datenqualität von XBRL Dokumenten stetig zu erhalten oder noch zu verbessern⁵. Bis Ende 2014 sollen alle offiziell gelisteten Unternehmen ihre Berichte in XBRL veröffentlichen ([ZW10]).

2.1.2 Syntax und Aufbau von XBRL

Die Frage, warum XBRL der hohen Individualität und daraus resultierenden Komplexität von veröffentlichten Finanzberichten gewachsen ist, lässt sich mit einem genaueren Blick in die Syntax erklären. Jedes interaktive (XBRL-)Dokument besteht zu einem Teil aus Instanzen und zum anderen Teil aus einer Taxonomie (manchmal auch als *dictionary* bezeichnet).

Um das Beispiel Mastercard (siehe Abbildung 2) aufzugreifen, soll anhand von diesem der Aufbau von Instanz Dokumenten aufgezeigt werden. In diesen werden in XBRL unterschiedliche financial facts in entsprechender Syntax dargestellt. Als financial fact wird in diesem Fall eine *einzelne* Aussage eines Jahresberichtes bezeichnet. Mehrere financial facts zusammen bilden also somit einzelne Bereiche einer Bilanz (vgl. Tabelle Abbildung 2). Die Ergebnisse der aufgezeigten Tabelle aus Abbildung 2 würden in XBRL dementsprechend wie folgt aussehen:

```
<us-gaap:AccruedLiabilitiesCurrent
  id="Fact-F33D1242C9A0CE10BD10EE7C4B67DDEE"
  contextRef="FI2014Q2"
  unitRef="usd"
  decimals="-6">1903000000</us-gaap:AccruedLiabilitiesCurrent>
```

Abbildung 3: Codebeispiel (1) eines financial facts (Beispiel: Mastercard)

```
<us-gaap:AccruedLiabilitiesCurrent
  id="Fact-4D7C81F71B1998AC784CEE7C4B639F27"
  contextRef="FI2013Q4"
  unitRef="usd"
  decimals="-6">2101000000</us-gaap:AccruedLiabilitiesCurrent>
```

Abbildung 4: Codebeispiel (2) eines financial facts (Beispiel: Mastercard)

⁵<http://xbrl.us/pages/about.aspx> Abruf: 09.10.2014

Man kann also erkennen, dass das Konzept, die Währung, die Anzahl gerundeter Stellen samt Wert, sowie eine Kontext Referenz und eine ID gegeben sind. Die dazugehörigen Informationen der **contextRef**, hier exemplarisch für den financial fact aus Abbildung 4, werden ebenfalls einem Teil des Instanz-Dokuments entnommen:

```
<xbrli:context id="FI2013Q4">
  <xbrli:entity>
    <xbrli:identifizier scheme="http://www.sec.gov/CIK">0001141391</xbrli:identifizier>
  </xbrli:entity>
  <xbrli:period>
    <xbrli:instant>2013-12-31</xbrli:instant>
  </xbrli:period>
</xbrli:context>
```

Abbildung 5: Codebeispiel zu genaueren Informationen eines contextRef Verweises

Auffällig ist, dass die restlichen Inhalte der Tabelle, also wie sich die Ergebnisse der beiden Quartale zusammensetzen, als eigene financial facts angesehen werden und somit in eigenen Konzepten angegeben sind (siehe Anhang A.1).

Ebenfalls gut erklären lässt sich der Gebrauch von Taxonomien in diesem Beispiel. Diese bieten allen Herausgebern von Finanzdaten in XBRL die Möglichkeit, ihre eigenen Konzepte zu entwerfen, zu beschreiben und im Endeffekt in ihren Instanz-Dokumenten zu benutzen. Die Quelle der Taxonomie, aus welcher die Konzepte stammen, kann jeweils vor dem Namen des Konzeptes abgelesen werden – im Regelfall ist dies die (US)–GAAP Taxonomie. Allerdings hat auch Mastercard in einem ihrer financial facts ein eigenes Konzept benutzt, was man am tag “ma“ erkennen kann:

```
<ma:AccruedCustomerPrograms
  [...]
>VALUE</ma:AccruedCustomerPrograms>
```

Abbildung 6: Codebeispiel eines eigenen Taxonomie Tags (Beispiel: Mastercard)

Die Taxonomie, in der dieses Konzepts genauer beschrieben wird, soll nun betrachtet werden. Die genaue Beschreibung kann in der durch Mastercard mitgelieferten Taxonomie wie folgt gefunden werden:

```

<xsd:element
  id="ma_AccruedCustomerPrograms"
  type="xbrli:monetaryItemType"
  xbrli:periodTime="instant"
  substitutionGroup="xbrli:item"
  nillable="true"
  name="AccruedCustomerPrograms"
  xbrli:balance="credit">

```

Abbildung 7: Codebeispiel eines eigenen Taxonomie facts (Beispiel: Mastercard)

Durch die Taxonomien können also beliebig viele, auf jedes Unternehmen individuell zugeschnittene, Finanzkonzepte eingeführt werden. Im vorliegenden Beispiel handelt es sich um das von Mastercard eingeführte Konzept “AccruedCustomerPrograms“. Da dieses nicht in der (US-)GAAP Taxonomie vorhanden ist, muss es durch die unternehmenseigene Taxonomie genauer definiert werden. In dieser können Informationen zum type, über den Zeitraum der Betrachtung, über die substitution group (tuple oder item), über den balance-type (debit oder credit) und außerdem darüber, ob der Wert “null“ angenommen werden darf, erhalten werden⁶ (siehe Abbildung 7).

Das Beispiel verdeutlicht, wie flexibel die Möglichkeiten zum Einsatz von XBRL sind. Den Unternehmen sind keine bis wenige Grenzen gesetzt, wie und welche Konzepte (und die damit verbundenen financial facts) sie veröffentlichen. Die 2003 veröffentlichten GAAPs ([DFO⁺09]) bieten außerdem bereits eine Menge vordefinierter Konzepte und nehmen somit Unternehmen die Arbeit ab, häufig benutzte Konzepte eigens beschreiben zu müssen.

Auf Basis dieses Beispiels kann nun eine generalisierte Form der Syntax in Instanz Dokumenten sowie Taxonomien hergeleitet werden.

```

<taxonomy:conceptName
  id="idReferenz"
  contextRef="contextReferenz"
  unitRef="unit"
  decimals="roundings">VALUE</taxonomy:conceptName>

```

Abbildung 8: Generelle Darstellung eines financial facts

Hierbei ist der Begriff “taxonomy“ als Variable zu verstehen, die angibt, in welcher Taxonomie die genaue Beschreibung des jeweiligen Konzepts zu finden ist.

⁶Laut XBRL Glossar gibt es keinen Grund, ein Element so zu beschreiben, dass die bool’sche Variable nillable “false“ annimmt. www.sec.gov/spotlight/xbrl/glossary.shtml Abruf: 09.10.2014

```

<xsd:element
  id="taxonomy_conceptName"
  type="xbrli:monetaryItemType"
  xbrli:periodTime="instant/duration"
  substitutionGroup="xbrli:item/xbrli:tuple"
  nillable="true/false"
  name="conceptName"
  xbrli:balance="credit/debit">

```

Abbildung 9: Generelle Darstellung eines eigenen Taxonomie facts

Die Vorzüge durch die Einführung von XBRL liegen auf der Hand: Der Spagat zwischen der Vereinheitlichung komplexer Daten in einer maschinenlesbaren Sprache und einer dabei angemessen hohen Flexibilität, was die Repräsentationsmöglichkeiten der Daten betrifft, gelingt. Aufgrund dessen sagte auch John Stantial bereits im Juni 2007 ([Sta07]) voraus, dass die vernünftige Anwendung von XBRL die nötigen Arbeitsstunden der zu veröffentlichenden Quartalsberichte um ca. 17% senken wird.

Die Einführung von XBRL war somit ein wichtiger Schritt in Richtung vereinheitlichter Abfragen aus mehreren Datenquellen. Da nicht alle relevanten Daten in XBRL präsentiert werden, gilt es nun also folgerichtig, durch das graphenbasierte Dateiformat RDF eine Darstellung zu finden, in welche alle benötigten Daten integriert werden können. Thema des nächsten Abschnitts wird die Suche des hierfür benötigten mappings von XBRL in RDF sein.

2.2 Finanzdaten im Semantic Web

Die Vereinheitlichung im Bereich Finanzdaten ist für die weiteren Schritte dieser Arbeit unumgänglich. Daher ist die nächste Aufgabe, ein simples, aber dennoch alle Tiefen von XBRL beschreibendes mapping in RDF zu finden. Hierzu werden mehrere Ansätze betrachtet, deren Vor- und Nachteile beleuchtet und schlussendlich ein Ansatz zur weiteren Vorgehensweise ausgewählt. Zu betonen ist an dieser Stelle, dass es hierbei (noch) nicht *das* richtige oder standardmäßige Vorgehen gibt ([KWO⁺14]).

2.2.1 Grundlegende Ideen und Vorzüge von RDF

Eine kanonische Darstellung **aller** nötigen Daten in RDF ist das langfristige Ziel. Ist dieses erst erreicht, können nahezu beliebig modifizierbare und automatisierbare SPARQL Abfragen über diese Daten durchgeführt werden. Ebenfalls unterstützt RDF die gewünschten, individuellen Erweiterungsmöglichkeiten durch die mögliche Integration von Daten aus verschiedenen Quellsystemen und Applikationen([BRLD10], [GG10]). Ein weiterer

Vorteil von RDF ist die gute Visualisierbarkeit, welche auch ohne große Vorkenntnisse im Bereich des Semantic Web intuitiv ist und sich somit gut dazu eignet, Leuten aus anderen Fachgebieten die präsentierte Vorgehensweise erklären zu können. Diese Visualisierbarkeit wird vor allem auch durch die möglich Notation als Graphen (vgl. beispielsweise Abbildungen 10 und 12).

2.2.2 Mapping Ansatz von Kämpgen et al.

Der erste betrachtete mapping Ansatz macht sich das RDF Vokabular Data Cube (QB)⁷ zu Nutze. In ihrem Paper berichten Kämpgen et al. ([KWO⁺14]), wie sie das mapping bewerkstelligen:

Da jedes XBRL Dokument, sei es eine Instanz oder eine Taxonomie, als mehrdimensionaler Datensatz betrachtet wird, liegt es nahe, sich für die Modellierung an das bereits bestehende RDF Data Cube Vokabular (qb:) zu halten. Für jede XBRL Instanz samt Taxonomie wird nun besagtes Datenset (qb:DataSet) mit einer “data structure definition” (qb:DataStructureDefinition) angelegt. Für jeden einzelnen Financial Fact (vgl. Abbildung 3 und Abbildung 4) wird nun eine Observations-Instanz (qb:Observation) angelegt und mit den gegebenen Informationen verlinkt. In diesem Fall könnten dies beispielsweise Start- und Endzeitpunkt einer Beobachtung, die Einheit des betrachteten Wertes, das Konzept oder der Betroffene in Zusammenhang mit den Daten sein. In den folgenden Teilen dieser Arbeit wird für alle Informationen, deren Quelle die SEC Datenbank ist, unter Berücksichtigung der Richtlinien von compact URI (CURIE)⁸ das Präfix “edgar:” benutzt. Weitere Präfixes sind “ical:”⁹ und “sdmx-measure:”¹⁰. Ersteres tätigt Aussagen über den Zeitraum des betrachteten financial facts, während das Zweite den jeweils betrachteten Wert zurückgibt. Abbildung 10 fasst die Vorgehensweise – bezogen auf den oberen Anwendungsfall – anschaulich zusammen:

⁷<http://www.w3.org/TR/vocab-dta-cube/> Abruf: 09.10.2014

⁸<http://www.w3.org/TR/curie/> Abruf: 09.10.2014

⁹<http://www.w3.org/2002/12/cal/> Abruf: 09.10.2014

¹⁰<http://publishing-statistical-data.googlecode.com/svn/trunk/specs/src/main/vocab/sdmx-measure.ttl> Abruf: 09.10.2014

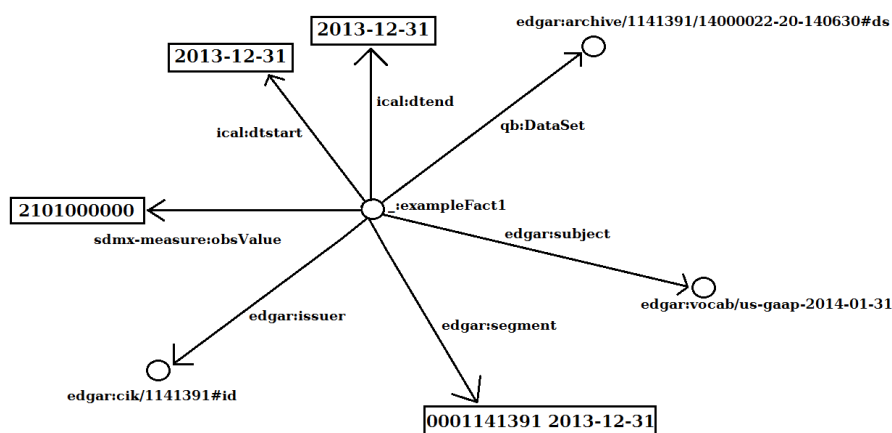


Abbildung 10: Mapping Beispiel auf Grundlage von [KWO+14]

Die zur zuvor gegebenen Abbildung gehörende Schreibweise dieses mappings in sog. "Tripleform" (nach CURIE Regeln):

```
_:exampleFact1 edgar:segment "0001141391 2013-12-31" .
_:exampleFact1 edgar:subject edgar:vocab/us-gaap-2014-01-31 .
_:exampleFact1 qb:DataSet edgar:archive/1141391/1400022-20-140630#ds .
_:exampleFact1 ical:dtstart "2013-12-31" .
_:exampleFact1 ical dtend "2013-12-31" .
_:exampleFact1 sdmx-measure:obsValue "2101000000" .
_:exampleFact1 edgar:issuer edgar:cik/1141391#cik .
```

Abbildung 11: Mapping Beispiel (Triple-Form) auf Grundlage von [KWO+14]

Diese beiden Darstellungen lassen sich nun mit allen potentiellen anderen Informationen im RDF Format beliebig ergänzen. Beispielsweise wäre es möglich zusätzliche Informationen über das Unternehmen Mastercard aus der DBpedia Datenbank, die bereits in RDF vorliegt, zu ergänzen¹¹.

2.2.3 Mapping Ansatz von R. García und R. Gil

Einen etwas anderen Ansatz verfolgten Roberto García und Rosa Gil in ihren Überlegungen zu einem geeigneten mapping von XBRL ([GG10]). Diese stützten ihre Vorgehensweise auf eine Teilung des Gesamtproblems in zwei Hälften:

1. Die Übersetzung von XML Datenstrukturen von Instanz-Dokumenten in RDF (XML2RDF)

¹¹<http://de.dbpedia.org/page/MasterCard> Abruf: 09.10.2014

2. Die Übersetzung der XML Schemata von XBRL in Web Ontology Language (OWL)–Ontologien (XSD2OWL)

Die Betrachtung von Teil 1 des Problems stützen die beiden Autoren auf die Veröffentlichung von M. Klein aus dem Jahre 2002 ([Kle02]). Anhand dessen Arbeit wurde ein Übersetzungsalgorithmus (vgl. Table 1. [GG10]) entwickelt, der die XBRL–Teile der Instanz–Dokumente in RDF/XML überführt und somit das gewünschte Ergebnis erzielt. Hier wird so vorgegangen, dass der betrachtete financial fact als Instanziierung einer Klasse gehandhabt wird. Dieser hat insgesamt vier ausgehende Kanten: Zwei davon zu Literalen für die Dezimalstellen sowie den Wert des facts und die anderen beiden zu weiteren Knoten, die Informationen über die Einheit des Wertes sowie den durch “contextRef” referenzierten Informationen. Letzterer Knoten hat wiederum zwei (drei) ausgehende Kanten, die Informationen über die Entität, den Zeitraum (Period) bzw. zusätzliche Szenarien bieten – sofern diese gegeben sind. Abbildung 12 stellt, wieder in Bezug auf Abbildung 3, eine Form dieses mappings bereit.

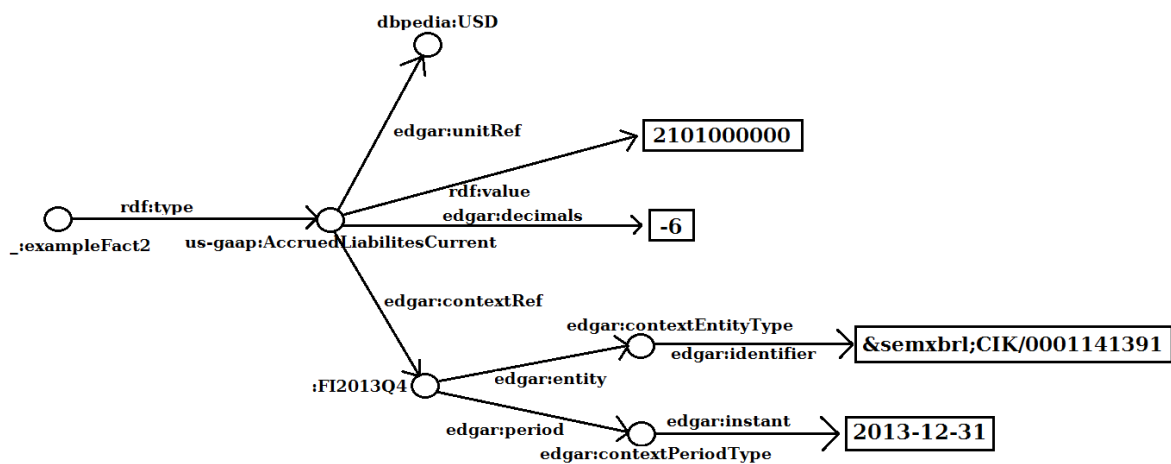


Abbildung 12: Mapping Beispiel auf Grundlage von [GG10]

Für den zweiten Teil des Problems, nämlich das mapping von XML Schema Daten aus Taxonomie–Dokumenten, haben die Autoren eine Tabelle angefertigt, in welcher die Form in XML Schema und die dazugehörige Form in OWL stehen. Ferner wird zu jedem mapping erläutert, warum man welchen Typ wie mapped (vgl. Table 2. [GG10]).

2.2.4 Mapping Ansatz von Bao et al.

Der dritte betrachtete Ansatz wurde von Bao et al. im Jahre 2010 durchgeführt ([BRLD10]). Diese führten einen Ansatz von Übersetzungen aller XBRL-Dokumente in OWL 2 DL. Hierbei erklären sie in ihrem Ansatz, dass drei Probleme zu betrachten seien. Benötigt werden:

1. Eine XBRL Ontologie, die alle XML-Schema Spezifikationen und Implikationen bezüglich Aufbau und Struktur der Dokumente gibt
2. Eine Taxonomie Ontologie, welche sich im OWL Bereich analog zu den Taxonomie-Dokumenten des XBRL Bereichs verhält
3. Eine Instanz Ontologie, welche sich im OWL Bereich analog zu den Instanz-Dokumenten des XBRL Bereichs verhält

Das Paper stellt übersichtlich dar, nach welchen Regeln das mapping der oben genannten drei Punkte vonstatten gehen soll. Erwähnt wird, dass die generelle Ontologie (Punkt 1) xbrlo: als Standard URI hat. Wie beispielsweise die Übersetzung des eigenen Taxonomie-Facts (vgl. Abbildung 7) in OWL aussähe, stellt die folgende Abbildung dar:

```
Declaration(Class(ex:AccruedCustomerPrograms))
SubClassOf(ex:AccruedCustomerPrograms xbrlo:monetaryType)
SubClassOf(ex:AccruedCustomerPrograms xbrlo:credit)
```

Abbildung 13: Mapping Beispiel eines Taxonomie-Dokuments auf Grundlage von [BRLD10]

Die Autoren merken an, dass die Punkte “id” (aufgrund von Redundanz zum Attribute “name”) sowie “substitutionGroup” (da man die bereitgestellten Informationen bereits aus dem “type” Attribut ableiten kann) in der Transformation nicht beachtet werden. Weiterhin erklären die Autoren, wie die Übersetzung eines financial facts (vgl. Abbildung 3, Abbildung 5) aussieht:

```
ClassAssertion(ex:AccruedLiabilitiesCurrent _:c1)
ObjectPropertyAssertion(xbrlo:hasContext _:c1 ex:co1)
DataPropertyAssertion(xbrlo:value _:c1 "2101000000"^^xsd:double)
DataPropertyAssertion(xbrlo:precision _:c1 "6"^^xsd:integer)

ClassAssertion(xbrlo:numericContext ex:co1)
ObjectPropertyAssertion(xbrlo:period ex:co1 _:time)
DataPropertyAssertion(time:inXSDDateTime _:time "2013-12-31"^^xsd:dateTime)
```

Abbildung 14: Mapping Beispiel eines Instanz-Dokuments auf Grundlage von [BRLD10]

2.2.5 Betrachtung der Ansätze

Die drei gewählten Ansätze sind nur ein Auszug aller bereits angedachten mapping Verfahren. Dies zeigt, dass die Forschung vor allem im Bereich “Umwandlung von XBRL in RDF“ für diese Arbeit bereits erhebliche Vorarbeit tätigen konnte. Alle drei Ansätze haben verschiedene Vor- und Nachteile und sind auf ihre eigene Art und Weise intuitiv.

Die Ansätze 1 und 2 haben den Vorteil, dass entweder eine exemplarische Darstellung in Form eines Graphen gegeben war (Ansatz 1), oder man anhand einer genauen Anleitung des Papers ohne große Probleme den aus der dieser entstehenden Graphen nachvollziehen und entwerfen konnte (vgl. Abbildung 12). Bei Ansatz 3 ist dies logischerweise nicht direkt umzusetzen, da sich das Autorenteam auf die alleinige Übermittlung in OWL gekümmert hatte. Ein weiterer zu beachtender Punkt in Ansatz 3 ist das Wegfallen von manchen optionalen Attributen. Dies wird mit zu verhindernder Redundanz begründet.

Es werden also drei Ansätze betrachtet, von denen der erste sich auf eine Übersetzung in RDF spezialisiert hat, der zweite Ansatz den ersten Teil der XBRL-Dokumente in RDF und den zweiten Teil in OWL übersetzt und den letzten Ansatz, der lediglich OWL als Zielformat auswählt. Die Arbeit von Kämpgen et al. ([KWO⁺14]) hat noch einen weiteren Vorteil, der bisher noch keine Erwähnung in dieser Arbeit fand: Nicht nur wurde eine Übersetzung von XBRL in RDF visualisiert, sondern direkt noch das Zusammenspiel von DBpedia¹² sowie Yahoo! Finance¹³ graphisch dargestellt (vgl. Fig.2 [KWO⁺14]). Zusätzlich scheint die Implementierung der Datensätze über das RDF Vokabular Data Cube sinnvoll und gelungen umgesetzt.

Aufgrund dieser Vorteile wird also die Darstellung von Kämpgen et al. aus dem Jahre 2014 die Art des mappings von XBRL-Dokumenten in RDF sein, die für die weiteren Schritte dieser Arbeit zu Grunde gelegt wird.

2.3 Ergebnisse

Zum Abschluss der thematischen Einführung und des Standes der Forschung sollen die in diesem Kapitel besprochenen Inhalte zusammengefasst werden.

Zu Beginn wurde die allgemeine Problematik der Arbeit illustriert. Nach einer kurzen Einführung zum Aufbau von XBRL-Dokumenten und einer Beschreibung der generellen Syntax von diesen, wurden die Vorzüge von XBRL gegenüber der bisherigen Darstellung von Finanzberichten genannt, sowie eine kurze Historie zur Einführung dieser Sprache gegeben. Als Folge einer kurzen Begründung, warum eine Umwandlung von XBRL-Dokumenten in RDF für die weitere Arbeit unumgänglich ist, wurde als wichtigster Schritt auf argumentativer Grundlage ein passendes mapping gewählt, mit dem nahezu beliebige

¹²<http://dbpedia.org/About> Abruf: 09.10.2014

¹³<http://finance.yahoo.com/> Abruf: 09.10.2014

XBRL Datensätze auf die benötigte Syntax überführt werden können. Durch diese Überlegungen hat die Forschung bereits immense Vorarbeit für weitere Ausführungen dieser Arbeit geleistet.

Auf Grundlage dieses mappings gilt es nun, weitere relevante Quellsysteme zu identifizieren, die potenziell wichtige Daten für die Analyse des in Kapitel 5 präsentierten Anwendungsfalls zur Verfügung stellen können. Zuvor müssen für die Analysen relevante Kennzahlen identifiziert werden, denn erst wenn dies geschehen ist, kann geprüft werden, welche Daten für die entsprechenden Abfragen zusätzlich zu denen aus den Jahresberichten noch benötigt werden. Bevor dies geschieht, werden im nächsten Kapitel die Ziele der Arbeit auf Basis der nun vorherrschenden Grundlagen präzisiert.

3 Entwicklung und Herleitung der Hypothese

Aufgrund der zuvor bereitgestellten Informationen kann nun bei allen Lesern ein ähnliches und vor allem angemessenes Vorwissen zur Thematik vorausgesetzt werden, sodass nun die Herleitung sowie Entwicklung der genauen Hypothese verständlicher erscheint.

Wie insbesondere das vorangehende Kapitel gezeigt hat, gibt es bereits viele Ansätze, die sich mit der Vereinheitlichung von verschiedenen Finanzdaten beschäftigt haben (vgl. [BRLD10], [KWO⁺14], [GG10]). Der nächste Schritt, nämlich diese Daten automatisiert und unter speziellen Vorgaben in Analysen oder Berichte umzuformen, wurde jedoch bisher – wenn überhaupt – nur sporadisch verfolgt. Dies könnte den Grund haben, dass neben Wissen im Bereich der Informatik (bzw. im Bereich Linked Data), ebenfalls Kenntnisse aus der Betriebswirtschaftslehre (BWL), genauer gesagt der Finanzwirtschaft und des Rechnungswesens nötig sind, um dieses Ziel effektiv angehen zu können.

Exakt aus diesem Grund wird sich der nächste Teil der Arbeit auf einen Anwendungsfall in der Praxis der Betriebswirtschaftslehre erstrecken, für welchen im letzten Kapitel die in diesem Zusammenhang aus der Informatik benötigten Grundlagen erläutert wurden. Im Laufe der nächsten Kapitel soll nun geprüft werden, ob eben diese Auswertung auch generell ein profitables Feld für Finanzanalysten werden kann.

Hypothese: SPARQL–Abfragen über ein integriertes Datenset, bestehend aus Finanzdaten verschiedener Quellsysteme in RDF, sind:

1. in einem angemessenen Rahmen durchführbar
2. performant ausführbar und
3. können die Arbeit im Bereich der Analyse von Finanzdaten erleichtern

Diese Hypothese stützt sich auf die Funde, die in den Kapiteln 4 und 5 präsentiert werden. Ferner werden nach Klärung der generellen Durchführbarkeit, Anregungen zu weiteren Forschungsgebieten gegeben (vgl. Kapitel 7), eine Diskussion aus Autorensicht geführt (vgl. Kapitel 8.1) sowie eine Prüfung der Ergebnisse, die durch die Begutachtung von Umfragewerten (vgl. Kapitel 8.2.2) sowie in Form einer Performance–Analyse (vgl. Kapitel 8.2.3) durchgeführt wurde, präsentiert. All dies soll dabei helfen beurteilen zu können, ob auch bei größeren Datensätzen und komplexeren Abfragen die Datenintegration in einem angemessen kurzen Zeitraum möglich ist. Ferner wird überprüft, ob auch wirklich Daten geliefert werden, mit denen in realistischen Szenarien gearbeitet werden kann.

4 Betrachtung der Kennzahlenauswahl

Dieses Kapitel beleuchtet neben der Komplexität, die das geplante Verfahren aufgrund des hohen Grades an durch Einstellungsmöglichkeiten bedingter Individualität mit sich bringt, auch eine Auswahl an Kennzahlen, die interessant für Analysen einzelner Unternehmen sein können. Zuvor wird kurz das Dilemma des Interessenskonfliktes zwischen Analysten und Herausgebern von Finanzberichten erläutert, wohingegen am Ende des Abschnitts auf die Problematiken, die der Versuch einer Auswahl “allgemein wichtiger“ Kennzahlen mit sich bringt, eingegangen wird.

4.1 Grundlage zur Veröffentlichung von Daten

Fraglich ist, warum Unternehmen überhaupt ihre Finanzberichte öffentlich zugänglich machen müssen. Staaten verabschieden Gesetze, in denen die Art und Weise der gewünschten Publikationen genau geklärt sind. Da sich die Ausführungen dieser Arbeit zu einem großen Teil auf Daten der U.S. Securities and Exchange Commission (SEC) beziehen, gilt in diesem Fall das amerikanische Recht¹⁴. Dieses gibt darüber Auskunft, wann welche Form der Finanzberichte vorgelegt werden müssen und in welcher Frist ein an der Börse gelistetes Unternehmen dies zu erledigen hat. Ferner gibt es Vorlagen der SEC, die vorgeben, welchen Inhalt der jeweilige Bericht auf alle Fälle wiedergeben muss (hier am Beispiel der Form 10-K¹⁵). Das deutsche Pendant zu diesem Gesetz ist nach [Bur11] bei den Vorschriften zur Buchführung¹⁶ in Verbindung mit § 325¹⁷ im Handelsgesetzbuch (HGB) zu finden.

4.2 Das Dilemma der Veröffentlichungen

Trotz der vielen Vorschriften bezüglich der Finanzberichte, gibt es bei der Veröffentlichung ein anhaltendes Dilemma zwischen Publizierenden und Analysten. Während Unternehmen möglichst nur die Dinge preisgeben wollen, die ihnen vom Gesetzgeber vorgeschrieben werden, sind vor allem für Analysten die Daten darüber hinaus interessant.

Dieser Problematik wurde auch mit der Einführung von XBRL-Dokumenten entgegengetreten: Diese lassen in der Veröffentlichung von Daten bei Weitem weniger Spielraum, als die vom Unternehmen selbst zu wählenden Formulierungen (ähnlich wie bei dem Fall von durchweg “positiven“ Aussagen in Arbeitszeugnissen). Zusätzlich wird durch die Möglichkeit zur Einführung eigener Konzepte die Aussicht auf die Veröffentlichung “zusätzlicher“ Informationen bis zu einem **selbst** wählbaren Punkt gestärkt.

¹⁴<http://www.law.cornell.edu/uscode/text/15/78m> Abruf: 09.10.2014

¹⁵<https://www.sec.gov/about/forms/form10-k.pdf> Abruf: 09.10.2014

¹⁶<http://dejure.org/gesetze/HGB/238.html> Abruf: 09.10.2014

¹⁷<http://dejure.org/gesetze/HGB/325.html> Abruf: 09.10.2014

4.3 Auswahl von Kennzahlen für das Anwendungsszenario

Die Wahl der Kennzahlen ist ein zugleich für das Szenario entscheidender, aber auch komplexer und nur schwer zu vereinheitlichender Schritt, nicht zuletzt auch wegen vieler verschiedener Ansätze, um die Performance von Banken zu analysieren ([KPDZ06]). Schon Hanley und Suter (1997) sowie Bhagwat und Sharma (2007) erkannten in ihren Studien, dass die Betrachtung standardisierter Controlling Zahlen wie “return on earnings“(ROE) oder “return on assets“(ROA) nicht zwangsläufig eine verlässliche Aussage darüber geben, wessen Performance nun gut oder schlecht ist([HS97], [BS07]).

Außerdem müssen auf der einen Seite Kennzahlen gewählt werden, die auf den Anwendungsfall angepasst sind, andererseits werden aber auch allgemeingültige Kennzahlen gebraucht, um die Anwendbarkeit in anderen Branchen bzw. die Übertragbarkeit des Szenarios auf diese gewährleisten zu können. Ferner wird der Versuch unternommen, die einzelnen Kennzahlen in verschiedene Bereiche (beispielsweise Verhältnis, Zeitreihe,...) einzuordnen.

1. Assets (oder totalAssets) – Aktivposten der Bilanz

Zur Einleitung in die Wahl der Kennzahlen werden die (total-)Assets gewählt. Diese bedürfen keiner zusätzlicher Datenquellen außerhalb der SEC Datenbank und können als Zeitreihe dargestellt werden, um die Entwicklung eines Unternehmens in diesem speziellen Kontext zu unterstreichen. Die Signifikanz dieser Kennzahl spiegelt sich auch in den Ausführungen von Zhu und Wu aus dem Jahre 2010 ([ZW10]) wieder, in welchen sie die Häufigkeit der benutzten Elemente in XBRL-Dokumenten untersucht hatten.

Die Aktivposten befinden sich auf der “linken“ Seite (Soll) der Bilanz. Sie werden in kleinere Unterpunkte gegliedert, wie beispielsweise das Anlagevermögen oder das Umlaufvermögen, um “die Verwendung der eingesetzten Mittel deutlich widerzuspiegeln¹⁸“. Hierbei sind einzelne Punkte wie Barvermögen, Bankvermögen oder der Wert der aktuell fertigen Erzeugnisse und Ressourcen enthalten.

2. Unternehmenseffizienz – Gesamtumsatz pro Mitarbeiter

Die Betrachtung der Effizienz macht sich zu großen Teilen den Data Envelopment Analysis (DEA) Ansatz in der Betriebswirtschaftslehre zu Nutze. Simpel ausgedrückt geben diese das Verhältnis zwischen einem Input und einem Output an. Hierbei ist die Wahl der Input- sowie Outputvariablen bis auf zwei Einschränkungen frei gestaltbar (vgl. [Wei01], [GS14]):

¹⁸<http://www.wirtschaftslexikon24.com/d/aktiva/aktiva.htm> Abruf: 09.09.2014

1. alle Variablen sind messbare Werte
2. alle Variablen sind positiv

Zusätzlich unterstützt wird die Signifikanz von DEA Werten im Kontext der Performance von Banken durch die Arbeiten von Chen et al. sowie Giokas aus dem Jahre 2008, die eine Data Envelopment Analyse in zwei unabhängigen Anwendungsszenarien im Bankenbereich durchführten ([CCP08], [Gio08]).

Im vorliegenden Beispiel werden zur Messung der Effizienz eines Unternehmens die Variablen “Umsatz der betrachteten Periode t ¹⁹“ als Outputvariable und die Anzahl der Mitarbeiter des Unternehmens als Inputvariable gewählt. Das Maß der Effizienz errechnet sich dann durch

$$\frac{Revenues(t)}{numberOfEmployees} \quad (4.1)$$

Diese Kennziffer wurde bereits 1999 in den Ausführungen zur Analyse der Top 55 US commercial banks von Seiford und Zhu im Kontext mit Unternehmensperformance erwähnt ([SZ99]) und beinhaltet nun Daten mehrerer Quellsysteme. Der Gesamtumsatz der jeweiligen Periode kann also in der Regel als “totalRevenue“ aus den XBRL-Dokumenten der SEC Datenbank erhalten werden. Die Anzahl der Mitarbeiter wird unter anderem durch DBpedia gegeben.

3. Unternehmenseffizienz – Pro-Kopf-Gewinn

Ebenfalls stellten Seiford und Zhu in ihren Studien aus dem Jahre 1999 ([SZ99]) eine weitere, mit der Anzahl der Mitarbeiter zusammenhängende Kennzahl, basierend auf einer DEA auf. Sie behaupteten, dass in der zuvor erwähnten Kennzahl neben dem Umsatz auch Profite oder Verluste als Output gewählt werden können. Als weitere Kennziffer ergibt sich somit:

$$\frac{Profits(t)}{numberOfEmployees} \quad (4.2)$$

Die im nachfolgenden Kapitel angestellten Untersuchungen werden zusammen mit der Auswertung des Anwendungsfalls zeigen, welche der beiden Kennzahlen performanter ist und als signifikanter für die Analyse der Leistungsfähigkeit wahrgenommen werden sollte.

4. Umsatz pro Geschäftsstelle

Dieser Ansatz betrachtet die Umsätze eines Unternehmens im Verhältnis zur Anzahl der Standorte, die es besitzt. Hierbei wird, ähnlich wie zu den vorhergehenden Kennzahlen,

¹⁹ t kann in diesem Fall – je nach Bericht – jährlich oder vierteljährlich sein

wieder die Methode der DEA genutzt. Als Input wird die Anzahl der Geschäftsstellen gewählt, welche durch die DBpedia Datenbank abzufragen ist. Ferner wird als Output der Jahresumsatz festgelegt, welcher wie gewohnt durch SEC Dokumente für die Abfragen bereitgestellt wird.

$$\frac{Revenues(t)}{numberOfLocations} \quad (4.3)$$

Durch diese Kennzahl kann eine Aussage über die Durchdringung des Einzugsgebietes getan werden. Auch Devlin und Gerrard erkannten in ihrer Studie von 2005 zur Bankenauswahl die Signifikanz der Entfernung zum nächsten Standort und die damit verbundene Wichtigkeit einer guten Durchdringung des potentiellen Zielmarktes mit Außenstellen ([DG05]).

5. Effektivität der Produktion

Diese Kennzahl stützt sich auf die Erkenntnisse der Studie von Giokas aus dem Jahre 2008 ([Gio08]). Dieser untersuchte mithilfe der DEA und verschiedenen In- sowie Outputs die Effektivität von Banken in den Bereichen "Production efficiency", "Transaction efficiency" sowie "Intermediation efficiency". Letztere bezeichnet "[...]die Vermittlerrolle von Finanzinstitutionen zwischen Angebot und Nachfrage nach Krediten, also ganz allgemein Handel und Vermittlung von Finanzdienstleistungen.²⁰".

Eine Übersicht über mögliche In- und Outputs zur Messung dieser Kennzahl auf Basis der Funde von Gioas gibt Wu in seinem Artikel des Jahres 2012 (vgl. Table 1, [Wu12]). Für diesen Fall wird auf Basis der Auswahlmöglichkeiten zur "Production efficiency" von Wu folgende Formel gewählt:

$$\frac{Non - interestIncome}{PersonnelCosts} \quad (4.4)$$

Beide benötigten Werte können den XBRL-Dokumenten der SEC entnommen werden.

6. Erzielte Gewinne / Verluste über mehrere Perioden

Diese Kennzahl geht den Schritt des Outputs von Nummer 2 weiter. Betrachtet wird in diesem Fall nicht mehr der Gesamtumsatz einer Periode, sondern der Jahresüberschuss. Begründung, warum dieser nicht direkt in der Effizienzbetrachtung berücksichtigt wird, ist die Voraussetzung (2) für die Anwendung der DEA. Der Jahresüberschuss kann natürlich auch negativ sein (man spricht dann von Verlust) und dürfte somit nicht als Input-Wert gewählt werden.

²⁰http://www.finanz-lexikon.de/intermediation_2495.html Abruf: 15.09.2014

Auch wenn die Möglichkeit besteht, den Gewinn/Verlust einer einzelnen Periode zu betrachten, kann eine Aussage über eine Tendenz erst durch die Analyse eines längeren Zeitraums getroffen werden. Diese Variable kann als zeitlich betrachtete Summe klassifiziert werden:

$$\sum_{i=t}^n NetIncomeLoss(t) \quad (4.5)$$

Hierbei stellt die Variable n die letzte der betrachteten Perioden, i die erste im Laufindex und t die Periode, von der Gewinn (income) bzw. Verlust (loss) betrachtet werden. Mit einer grafischen Darstellung dieser Daten kann eine Entwicklung, vielleicht sogar ein zyklischer Verlauf gut erkannt werden – je nachdem ob man Monats oder Jahresdaten betrachtet. Die Signifikanz dieser Aufzählung im Banksektor erkannten auch Albertazzi und Gambacorta in ihrer Arbeit zur Profitabilität der Arbeit von Banken aus dem Jahre 2009 ([AG09]).

7. Präzision von Angaben verschiedener Quellen

Wenn verschiedene Quellen Werte zu gleichen Daten angeben, gibt es oftmals kleinere Rundungsfehler, manchmal aber auch enorme Diskrepanzen zwischen den Daten. Auch wenn es sich hierbei nicht direkt um eine Kennzahl handelt, ist die Verifizierung der Richtigkeit von benutzten Daten ein essentieller Schritt.

Diese Abfrage klärt, ob es sich bei den Unterschieden zwischen zwei Daten, die eigentlich den gleichen Inhalt ausdrücken sollen, eventuell um Rundungsfehler oder eben doch um “falsche“ Daten handelt. Hierbei ist im vorliegenden Kontext die Anzahl aller Beschäftigten eines Unternehmens durch die SEC und auch DBpedia ein angemessenes Beispiel.

8. Aufbau des Gesamtunternehmens

Durch die zur Verfügung stehenden Daten, kann auch eine grobe Unternehmensstruktur aufgestellt werden. Über die Branche der Unternehmen aus der SEC Datenbank kann man auf die jeweiligen DBpedia Einträge schließen.

In einem zweiten Schritt werden anhand dieser durch das “parentCompany²¹“ Attribut alle Tochterunternehmen der betrachteten Bank ausgewählt. Von diesen werden nun, ebenfalls über die DBpedia Datenbank, Name, Anzahl der Angestellten sowie die Stadt des Hauptsitzes zurückgegeben²². Die Anzahl der Mitarbeiter ist deswegen sinnvoll, da diese eine ungefähre Einordnung der Größe der Tochtergesellschaft möglich macht.

²¹<http://dbpedia.org/ontology/parentCompany> Abruf: 13.09.2014

²²Natürlich lassen sich auch eine Menge weitere Informationen bestimmen – für die exemplarische Vorgehensweise soll jedoch eine Auflistung der drei genannten Elemente genügen.

Diese Kennziffer ist nicht zuletzt auch deswegen interessant, da der Einfluss geographischer Ereignisse auf Unternehmen (bzw. Banken im Anwendungsfall) beispielsweise anhand der Hauptsitze von etwaigen Tochtergesellschaften bestimmt werden könnte.

9. Entwicklung des Aktienkurses

Durch Daten, die mit dem Yahoo! Finance Wrapper²³ in RDF abgerufen werden können, ist es möglich, die höchst sensiblen Aktienkurse in die Analyse mit einzubeziehen. Vor allem in Kombination mit der Analyse der Gesamtstruktur der Unternehmens, können beispielsweise Einflüsse auf einzelne Unternehmensteile (beispielsweise Tochterunternehmen, Partnerunternehmen) als Gründe für die etwaigen Schwankungen identifiziert werden.

4.4 Zusammenfassung und Klassifizierung

Zur Zusammenfassung dieses Absatzes dient eine Tabelle, welche übersichtlich alle für den Anwendungsfall ausgewählten Kennzahlen und die Datenquellen, aus denen jeweils Daten bezogen werden, darstellt.

Nr.	Kennzahl	Quelle(n)	URI(s)
1	(total)Assets	SEC	us-gaap2009:Assets
2	$\frac{Revenues(t)}{numberOfEmployees}$	SEC DBpedia	– dbpedia-owl:numberOfEmployees
3	$\frac{Profits(t)}{numberOfEmployees}$	SEC DBpedia	us-gaap2009:NetIncomeLoss dbpedia-owl:numberOfEmployees
4	$\frac{Revenues(t)}{numberOfLocations}$	SEC DBpedia	– dbpedia-owl:numberOfLocations
5	$\frac{Non-interestIncome}{PersonnelCosts}$	SEC SEC	us-gaap2009:LaborAndRelatedExpense us-gaap2009:noninterestIncome
6	$\sum_{i=t}^n NetIncomeLoss(t)$	SEC	us-gaap2009:NetIncomeLoss
7	Abgleich versch. Quellen	SEC DBpedia	us-gaap2009:NetIncomeLoss dbpedia-owl:netIncome
8	Gesamtunternehmen	SEC DBpedia	– dbpedia-owl:parentCompany
9	Aktienkurs	Yahoo	yahoo:Close

An den Stellen, an denen keine Einträge vorhanden sind, war keine eindeutige Zuordnung einer URI durchführbar. Dies hat den Grund, dass die Zahl entweder aus mehreren Teilen besteht, oder die Unternehmen diese in der Regel unterschiedlich publizieren. Ferner

²³<http://yahoofinancewrap.appspot.com/> Abruf: 09.10.2014

sind die beiden zu Abfrage 7 angegebenen URIs lediglich Beispiele für einen möglichen Datenabgleich. Aus Gründen der Lesbarkeit wurden verschiedene Präfixe verwendet, die zum Verständnis der Tabelle zu beachten sind:

```
@prefix yahoo: <http://yahoofinancewrap.appspot.com/vocab/stock#> .  
@prefix dbpedia-owl: <http://dbpedia.org/ontology/> .  
@prefix us-gaap2009: <http://edgarwrap.ontologycentral.com/vocab/us-gaap-2009-01-31#> .
```

Abbildung 15: Benutzte Präfixe der zusammenfassenden Tabelle

Abgeschlossen wird das Kapitel, indem die zuvor genannten neun Kennzahlen in vier verschiedene Gruppen klassifiziert werden.

- Zu Beginn wird als Heranführung an das Anwendungsszenario eine relativ schlichte Abfrage zur Ermittlung der totalAssets eines beliebigen Unternehmens aus bestehenden XBRL-Dokumenten durchgeführt. Die ermittelten Daten können als Zeitreihe dargestellt werden, um eine Aussage zur Entwicklung in diesem Bereich tätigen zu können.
- Die nächste Klasse wird vorwiegend durch das Konzept der DEA bestimmt. Mittels verschiedener In-/ und Outputs werden unterschiedliche Kennzahlen gebildet. Diese geben die Unternehmensperformance oder Unternehmenseffizienz als Verhältnis zwischen diesen beiden Zahlen an.
- Die siebte Kennzahl ist eine Summe. Analog zur ersten Kennzahl werden die Werte verschiedener Unternehmen über mehrere Perioden hinweg betrachtet. Da diese positive (Gewinn) wie negative (Verlust) Werte annehmen können, zeigt erst eine langfristige Betrachtung der Werte einen Trend an.
- Der letzte Teil der Kennziffern betrifft weniger die finanziellen Fakten, also Aufbau und Umfeld des Unternehmens. Interessant für Analysten ist stets die Präzision von Werten zu “gleichen“ Daten aus verschiedenen Quellen – diese Frage wird von Kennzahl 8 betrachtet. Zusätzlich wird die Struktur des Gesamtunternehmens betrachtet, ehe die Schwankungen in der Entwicklung des Aktienkurses betrachtet werden.

5 Anwendungsfall

Dieses Kapitel behandelt den Anwendungsfall der Arbeit. Es wird anhand der International Standard Industrial Classification (ISIC) bzw. Standard Industrial Classification (SIC) eine Unternehmensbranche genauer betrachtet werden. Anhand dieser werden dann auf Basis der in Kapitel 4 genannten Kennzahlen einerseits die Integration von Daten verschiedener Systeme durchgeführt und andererseits branchenspezifische Informationen zusammengetragen und ausgewertet.

5.1 Abgrenzung ISIC und SIC

Um ein Unternehmen oder eine natürliche Person anhand einer Identifikationsnummer identifizieren zu können, wird im Zusammenhang von publizierten Finanzberichten auf CIKs zurückgegriffen. Soll hingegen international eine spezielle Branche ausfindig gemacht oder bestimmt werden, können bereits die Möglichkeiten von sowohl ISIC, als auch SIC gewählt werden.

Die 1937 in den USA eingeführte SIC diente dazu, verschiedene Unternehmensbranchen identifizieren zu können²⁴. 1997 wurde dieses System durch das sechsstellige North American Industry Classification System (NAICS) ersetzt, wird jedoch beispielsweise von der SEC weiter verwendet²⁵. Da diese also auch in den XBRL-Datensätzen noch heute verwendet wird, ist es in dieser Arbeit ein relevanter Faktor, um eine entsprechende Unternehmensbranche auszuwählen. Jedes Unternehmen kann anhand einer zwei- bis vierstelligen Nummer einer Hauptgruppe zugeordnet werden, in welcher sich weitere Unternehmen aus dieser Branche befinden. Ähnliche Hauptgruppen lassen sich wiederum zusammenfassen und in eine von zehn Divisionen einordnen²⁶.

Die zweite Möglichkeit, einen Unternehmensbereich auszumachen, ist die ISIC. Diese ist ein Konstrukt der UNO, welches hauptsächlich in Europa angewendet wird. Sie klassifiziert jeden Industrie- oder Wirtschaftszweig in eine Hauptgruppe mit anderen Unternehmen ähnlichen Hintergrunds. Es gibt insgesamt 21 verschiedene Hauptgruppen²⁷. Ferner findet die ISIC in den EU-Ländern, sowie ca. zehn weiteren Staaten außerhalb der EU Anwendung²⁸.

Da ein großer Anteil der zur Verfügung stehenden Daten aus EDGAR-dumps stammt, verwendet diese Arbeit das Konzept der SICs. Wichtig ist jedoch im Hinterkopf zu be-

²⁴<http://www.sec.gov/investor/pubs/edgarguide.htm> Abruf: 09.10.2014

²⁵http://en.wikipedia.org/wiki/North_American_Industry_Classification_System Abruf: 09.10.2014

²⁶https://www.osha.gov/pls/imis/sic_manual.html Abruf: 09.10.2014

²⁷<http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27> Abruf: 09.10.2014

²⁸http://de.wikipedia.org/wiki/International_Standard_Industrial_Classification Abruf: 09.10.2014

halten, dass, wenn beispielsweise Daten aus dem europäischen Raum einbezogen werden sollen, nicht alle Differenzierungen über die SIC durchführbar sind.

5.2 Bestimmen von Unternehmen der gleichen Klassifizierung

Um im Anwendungsfall effizient Arbeiten zu können, muss die Auswahl einer speziellen Unternehmensbranche möglich sein. Dies kann durch SPARQL-Abfragen erreicht werden.

5.2.1 Generelles Vorgehen zur Unternehmensbestimmung

Die allgemeine Vorgehensweise bei gegebener SIC relevante Unternehmen aus den SEC Datensätzen abfragen zu können, ist relativ intuitiv. Hierzu wird die RDF-Struktur (siehe Abbildung 16) genutzt. Auf Basis dieser können alle bindings²⁹ zurückgegeben werden, die die Einschränkungen der Abfrage erfüllen (siehe Abbildungen 17 und 18).

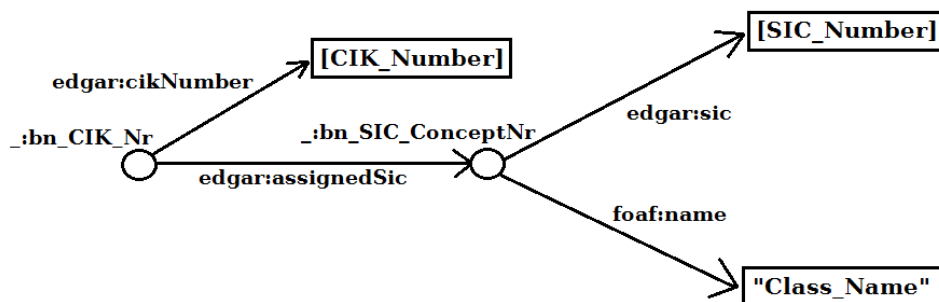


Abbildung 16: Zusammenhang CIK und SIC in der SEC Datenbank – Visualisierte Darstellung der RDF-Triples

Hierbei sind die in Anführungszeichen gesetzten Namen in den Literalen als Platzhalter zu verstehen, die durch die Kenntnis der CIK oder SIC mit Werten ersetzt werden. Die Werte der Variablen [SIC_Number] und [CIK_Number] sind hierbei vom Datentyp xsd:integer³⁰, der Wert des Literals "Class_Name" vom Typ xsd:string³¹.

Jetzt, wo diese generelle Darstellung bekannt ist, können mithilfe von zwei SPARQL-Abfragen die CIKs aller Unternehmen zurückgegeben werden, die in der selben Branche wie ein spezielles Unternehmen, markiert durch die SIC, tätig sind. Sollte die SIC nicht bekannt sein, so muss diese als erstes auf Basis des entsprechenden CIK abgerufen werden:

²⁹ Als "binding" werden alle Ausgaben bezeichnet, die den Voraussetzungen der Abfrage genügen.

³⁰ <http://www.w3.org/TR/xmlschema11-2/#integer> Abruf: 09.10.2014

³¹ <http://www.w3.org/TR/xmlschema11-2/#string> Abruf: 09.10.2014

```
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .

SELECT ?number
WHERE { ?cikID edgar:cikNumber [number] .
        ?cikID edgar:assignedSic ?sicNr .
        ?sicNr edgar:sic ?number . }
```

Abbildung 17: SPARQL-Abfrage der SIC bei gegebenem CIK

Hierbei nimmt die Variable [number] den Wert (als Integer) des CIK des Ausgangsunternehmens ein. Sollte die SIC bereits bekannt sein, gestaltet sich die Abfrage wie folgt:

```
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .

SELECT ?company ?cikNr
WHERE { ?concept edgar:sic [number] .
        ?company edgar:assignedSic ?concept .
        ?company edgar:cikNumber ?cikNr . }
```

Abbildung 18: SPARQL-Abfrage aller Unternehmen in einer SIC-Gruppe

Hierbei wird die Variable [number] durch die SIC (als Integer) der gewünschten Branche ersetzt. Diese ist entweder im Voraus bekannt, oder wurde durch die Abfrage, dargestellt in Abbildung 16, ermittelt. Als Ergebnis werden die Uniform Resource Identifiers (URIs) der zur SIC gehörenden Unternehmen, zusammen mit ihrem jeweiligen CIK als Integer (sofern angegeben) zurückgegeben.

5.2.2 Probleme bei der Bestimmung von Unternehmen der gleichen Branche

Leider gibt es bei dem oben beschriebenen Vorgehen zur Bestimmung aller gelisteten Unternehmen einer speziellen Branche offenkundig zwei Probleme:

1. Es werden offensichtlich nur die Unternehmen zurückgegeben, die auch eine SIC in ihren Angaben vermerkt haben. Die Vollständigkeit kann somit nie zu 100% gewährleistet werden.
2. Unternehmen können mehrere Einträge mit verschiedenen CIKs aufweisen. Dies bedeutet, dass sie als verschiedene Filings gelistet sind, es sich jedoch um das gleiche Unternehmen handelt³². Dies kann jedoch nur äußerst schwierig abgefangen werden, sodass diese dann mehrfach in der Auflistung der Abfrageergebnisse enthalten sind.

³²<http://www.sec.gov/cgi-bin/browse-edgar?company=APPLE+HOSPITALITY&owner=exclude&action=getcompany> Abruf: 09.10.2014

5.3 Anforderungen

Bevor die Details des Anwendungsfalls genannt und erläutert werden, gilt es diese zu begründen. Natürlich besteht der generelle Gedanke, eine theoretische Thematik für den Leser und alle Beteiligten greifbarer zu machen – speziell in diesem Kontext, also der automatisierten Finanzdatenanalyse, gilt es jedoch noch weitere Dinge zu beachten.

1. Der Spagat zwischen nötiger Detailtiefe und Übersicht muss gelingen. Das heißt im vorliegenden Zusammenhang, dass ein Szenario gefunden werden muss, dass auf der einen Seite alle etwaigen Problematiken, die sich im Detail befinden können, beleuchtet, aber auf der anderen Seite nicht zu spezifisch ist, um auf andere Bereiche mit wenigen Adjustierungen angewandt werden zu können.
2. Die Anzahl der Unternehmen in der gewählten Branche darf weder zu klein, noch zu groß sein. Es bringt nichts die These anhand eines Nischenbereichs zu testen, in welchem lediglich Daten von 2 oder 3 Unternehmen zur Verfügung stehen. Ebenso macht es wenig Sinn, 10 aus 1.000 (oder mehr) Unternehmen in einer Branche auswählen zu müssen, da die Repräsentativität von Abfragen, die branchenspezifisch sind, zumindest fragwürdig wäre.
3. Der Anwendungsfall muss Datenintegration aus mehreren Quellsystemen beinhalten. Nur, wenn diese Problematik im betrachteten Szenario auch behandelt wird, kann im Nachhinein eine Aussage getroffen werden, ob sich der Aufwand des Mapings in eine kanonische Form aus Daten verschiedener Systeme überhaupt lohnt.
4. Die Branche sollte greifbar sein. Das bedeutet, dass im besten Fall ein Szenario gewählt wird, welches häufigen Kontakt in Alltagssituationen bietet. Dies könnten beispielsweise beliebige retailer im Bereich Mode oder auch Nahrung sein, während Unternehmen, die Erze fördern und verarbeiten, eher nicht dazu zählen.

Die Auflistung lässt erkennen, dass die Anforderungen vielfältig sind. Ein weiteres Problem ist, dass sich manche dieser Anforderungen zumindest teilweise gegenseitig ausschließen können, was die Auswahl noch anspruchsvoller gestaltet.

Im nächsten Kapitel gilt es also nun, die Wahl einer Branche zu tätigen und auch zu begründen. Hierbei sollte schon vorher klar sein, dass es nicht DIE richtige oder DIE falsche Wahl gibt. Vielmehr wird in der weiteren Aufarbeitung dieses Kapitels der Versuch unternommen, aus 80 verschiedenen Branchen³³ genau die zu finden, welche möglichst viele Punkte der oben genannten Liste möglichst zufriedenstellend erfüllt.

³³Die Daten stellen Informationen für weit mehr als 100 verschiedene Unternehmensbranchen bereit. Eine Auflistung von 80 Beispielen sind im Anhang A.2.

5.4 Wahl und Beschreibung des Anwendungsfalls

Dieser Teil der Arbeit stellt das Anwendungsbeispiel vor und begründet die Auswahl. Aufgrund der Vielfältigkeit der Anforderungen an das Auswahlscenario, wurden zuerst sechs der 80 Branchen als angemessen identifiziert, die alle ihre Vorteile, bezogen auf die Anforderungen, beschrieben im vorherigen Abschnitt, besitzen.

5.4.1 Auswahl und Begründung

Aus diesen sechs Kandidaten wurde nun die Branche “STATE COMMERCIAL BANKS“ mit der SIC “6022“ ausgewählt. Diese Branche beinhaltet alle amerikanischen Banken sowie Treuhandgesellschaften, die von einem der US-Staaten oder einem Regierungsgebiet zugelassen wurden³⁴. Die Begründungen für diese Wahl beziehen sich auf die in Kapitel 5.3 beschriebenen Anforderungen:

1. Die nötige Detailtiefe bleibt komplett erhalten. Die ausgewählte Branche ist von allen verfügbaren mit 294 Einträgen die, mit den meisten Vertretern (vgl. Anhang A.2). Offensichtlich werden in dieser Arbeit nicht alle 294 mit dieser SIC assoziierten Unternehmen betrachtet, sodass die Auswahlmöglichkeiten eine angemessen-detaillierte Darstellung ermöglichen.
2. Wie bereits im vorherigen Punkt aufgegriffen, gibt es nahezu keine Begrenzung der Anzahl an Unternehmen, die gewählt werden können, da eine Auswahl von 294 betrachteten Unternehmen (oder mehr) für die Veranschaulichung der Arbeitsschritte nicht in einem angemessenen Rahmen durchgeführt werden könnte.
3. Die erwähnte Notwendigkeit, mehrere Datenquellen für die Beschreibung der Datenintegration aus diesen, war der ausschlaggebende Punkte für den Anwendungsfall. Neben der angesprochenen Grundlage, also den Daten der SEC, können weitere Informationen über die Unternehmen beispielsweise aus den Quellen “DBpedia“ oder auch “Yahoo! Finance“ extrahiert und zur Unternehmensanalyse hinzugezogen werden.
4. Die Greifbarkeit ist unter Umständen der einzige Punkt, bei der die gewählte Branche Schwächen zeigt. Auch, wenn man im Alltag schon in vielen Angelegenheiten mit Banken in Kontakt kommt, sind gerade amerikanische Banken doch für den gemeinen Leser dieser Arbeit etwas weiter entfernt. Nichtsdestotrotz lassen sich viele der Erkenntnisse auch auf Banken anderer Länder beziehen.

³⁴https://www.osha.gov/pls/imis/sic_manual.display?id=67&tab=description
09.10.2014

5.4.2 Datenintegration der gewählten Branche

Mit der Festlegung der betrachteten Branche samt ihrer SIC, können nun nach dem allgemeinen Vorgehen, welches in Abbildung 18 gezeigt wurde, die Unternehmen des Anwendungsfalls abgefragt werden. Hierfür muss im bereits angegebenen Code der SPARQL-Abfrage lediglich die Variable “number“ durch die SIC des Anwendungsszenarios, also in diesem Fall die Nummer “6022“ ersetzt werden.

```
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .  
  
SELECT ?company ?cikNr  
WHERE { ?concept edgar:sic "6022" .  
        ?company edgar:assignedSic ?concept .  
        ?company edgar:cikNumber ?cikNr . }
```

Abbildung 19: SPARQL-Abfrage für alle Unternehmen assoziiert zur SIC 6022

Die Ergebnisse dieser Anfrage teilen sich auf in die URI (?company) zum CIK des Unternehmens, sowie dem CIK als Integer (?cikNr). Aus Gründen der Lesbarkeit ist ein veranschaulichender Teil der Auflistung dieser Tabelle in Anhang A.3.1 zu finden.

Ferner ist durch die Festlegung der SIC der Branche für den Anwendungsfall die Möglichkeit gegeben, die Basis des datestores für die Abfragen der Kennzahlen, welche im folgenden Abschnitt stattfinden werden, zu bilden. Diese besteht aus dem dump aller Daten der SEC, die in deren Datenbank aus XBRL-Dokumenten mit der Branche “STATE COMMERCIAL BANKS“ bzw. mit der SIC “6022“ assoziiert sind.

6 Umsetzung

In diesem Kapitel werden die zuvor bestimmten Kennzahlen aus Kapitel 4.3 mit SPARQL aus der Datengrundlage abgefragt und die Ergebnisse angegeben. Ferner werden diese Ergebnisse im Kontext von Finanzanalysen interpretiert und die Möglichkeiten, die vor allem durch das Einbeziehen von Daten aus verschiedenen Quellen ermöglicht werden, aufgezeigt. Generell publizieren die Unternehmen mithilfe der US-GAAPs von 2009 und 2011. Aus Gründen der Übersicht und Konsistenz, wurde in dieser Arbeit mithilfe des “`rdfs:subClassOf`” property die jeweilige Kennzahl über eine eigens definierte Klasse (`us-gaap-ab:NameDerKennzahl`) abgefragt. Zu diesem Zwecke mussten für die benutzten Variablen eine Regel:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .

{ ?c1 rdfs:subClassOf ?c2 . ?x edgar:subject ?c1 . } =>
{ ?x edgar:subject ?c2 . } .
```

Abbildung 20: Regel zur Abfrage von Daten mit gleicher Bezeichnung aus mehreren GAAPs

und diverse `rdfs:subClassOf`-Beziehungen, hier allgemeine Form mit Platzhalter, eingeführt werden:

```
us-gaap2011:name rdfs:subClassOf us-gaap-ab:name .
us-gaap2009:name rdfs:subClassOf us-gaap-ab:name .
```

Abbildung 21: `rdfs:subClassOf`-Beziehungen (allgemein)

Zum Test der Abfragen wurde das Programm Linked Data-Fu in der Version 0.9.0 genutzt, mit welchem ohne endpoints SPARQL-Abfragen über ein beliebiges Datenset durchgeführt werden können ([SSHS13]). Der Nachteil dieser Methode: Es wird ein datastore benötigt, welcher die URIs verschiedener Ressourcen von gleichen Unternehmen miteinander verknüpft, sofern dies nicht von der jeweiligen Datenquelle ohnehin getan wird (vgl. “RDF Store“ Illustration Abbildung 1). Im vorliegenden Fall besteht der datastore aus einem dump von SEC Publikationen verschiedener Unternehmen des Zeitraums 2007–2010 in XBRL. Das bedeutet für diese Arbeit, dass in allen folgenden Kennzahlen des Anwendungsszenarios, welche Daten über die in XBRL veröffentlichten Berichte der SEC hinaus benötigen, eine Verbindung über das `owl:sameAs` property in den datastore eingefügt werden musste. Dieser Fall tritt erstmals bei Kennzahl 2 auf und wird dort nochmals erklärt (vgl. Abbildung 25).

6.1 Assets (oder totalAssets) - Aktivposten der Bilanz

Daten der Aktivseite der Bilanz können beispielsweise als Zeitreihe eines einzelnen (oder mehrerer) Unternehmen dargestellt werden. Die relevanten Daten, nämlich Wert (value) der Assets, Name (bankName) und CIK (cik) der zugehörigen Bank, bilden die Datengrundlagen dieser Kennzahl. Die zugehörige Abfrage sieht wie folgt aus:

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

SELECT ?value ?cik ?bankName
WHERE { ?bn edgar:subject us-gaap-ab:Assets .
        ?bn measure:obsValue ?value .
        ?bn edgar:issuer ?company .
        ?company foaf:name ?bankName .
        ?company edgar:cikNumber ?cik .
}
```

Abbildung 22: SPARQL-Abfrage Assets und dazugehörige Banken

Die Ergebnisse dieser SPARQL-Query lassen viele Analysemöglichkeiten zu. Ein exemplarischer Auszug der Ergebnisse dieser Abfrage findet sich in Anhang A.3.2. Das bisherige Problem: viele verschiedene Werte zum gleichen Unternehmen. Es gilt also, die Abfrage so anzupassen, dass der Zeitraum (oder der Zeitpunkt) mitbestimmt wird und so eine grafische Darstellung des Verlaufs des Wertes wiedergegeben werden kann. Die folgende Abfrage modelliert beispielhaft die Ausgabe aller Assets der Fifth Third Bancorp³⁵, inklusive Start- und Endpunkt der betrachteten Periode. Bei identischen Start- und Endpunkten handelt es sich um einen Wert zu einem speziellen Zeitpunkt. Die Abfrage des beschriebenen Falles lautet:

³⁵<https://www.53.com/site> Abruf: 09.10.2014

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

SELECT ?value ?dateBegin ?dateEnd
WHERE { ?company edgar:cikNumber "35527" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn .
        ?bn edgar:subject us-gaap-ab:Assets .
        ?bn ical:dtstart ?dateBegin .
        ?bn ical:dtend ?dateEnd .
        ?bn measure:obsValue ?value .
      }

```

Abbildung 23: SPARQL-Abfrage Assets der Fifth Third Bancorp

der dazugehörige Output ergibt sich folgendermaßen:

?value	?dateBegin	?dateEnd
119313000000	2009-03-31	2009-03-31
112651000000	2010-03-31	2010-03-31
113380000000	2009-12-31	2009-12-31
116294000000	2008-09-30	2008-09-30
110740000000	2009-09-30	2009-09-30
119764000000	2008-12-31	2008-12-31
119764000000	2008-12-31	2008-12-31
113380000000	2009-12-31	2009-12-31

Da bei allen Ergebnissen der Abfrage-Start- und Endzeitpunkt identisch sind, kann von der Erhebung zu einem festen Zeitpunkt ausgegangen werden. Somit sind nun, nach Bereinigung der Dubletten, für die Darstellung der Zeitlinie noch sechs Werte relevant. Diese ergeben folgenden grafischen Verlauf:

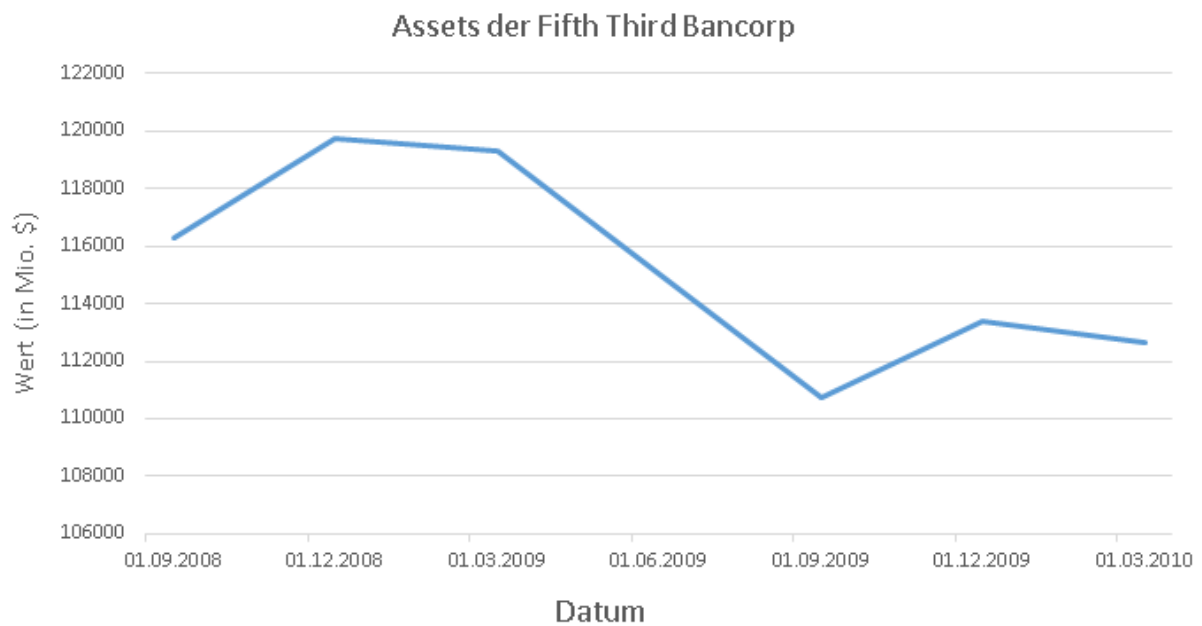


Abbildung 24: Verlauf des Wertes der Assets der Fifth Third Bancorp

Bereits dieses einführende Beispiel zeigt Möglichkeiten auf, die dieser Ansatz besitzt. Gewiss kann die Kennzahl Assets durch weitere, individuell auf andere Unternehmen oder Branchen und deren Anforderungen anpassbare Kennziffern erweitert oder ersetzt werden. Die Betrachtung der nächsten Kennzahlen wird zeigen, dass noch viel mehr Möglichkeiten umsetzbar sind. Beispielsweise können alle Unternehmen ausgegeben werden, die in einer bestimmten Kennzahl einen höheren (tieferen) Wert, als den des Beispielunternehmens besitzen. Da dies allerdings im Kontext der Assets analysetechnisch nur bedingt sinnvoll ist, wird dies in einem der folgenden Teile durchgeführt werden.

6.2 Unternehmenseffizienz – Gesamtumsatz pro Mitarbeiter

Wie zur Einleitung dieses Kapitels beschrieben, müssen bei der Integration von Daten aus verschiedenen Quellen, über den SEC Datensatz hinaus, weitere Triple dem RDF Store hinzugefügt werden. Im schlechtesten Falle müssen diese Triple per Hand eingetragen werden.

Im ersten Schritt werden die beiden Datenquellen durch das owl:sameAs property verbunden:

```
<http://edgarwrap.ontologycentral.com/cik/35527#id>
<http://www.w3.org/2002/07/owl#sameAs>
<http://dbpedia.org/resource/Fifth\_Third\_Bank> .
```

Abbildung 25: Verbinden der Datenquellen SEC und DBpedia (Beispiel: Fifth Third Bank)

Diese Verbindung ist in der Regel noch manuell einzusetzen. Es gibt einige wenige Ausnahmen, die bereits innerhalb der DBpedia Datenbank ein secCik property³⁶ zur entsprechenden URI des CIK besitzen.

Im nächsten Schritt können der zugeordneten DBpedia Seite des Unternehmens verschiedene Daten samt Werten hinzugefügt werden. DBpedia bietet hierzu die Möglichkeit an, die Informationen einer Seite im RDF Format als ntriples zu exportieren³⁷. Die weitaus elegantere Methode, wie sie auch in der query präsentiert wird, ist jedoch das Abfragen der benötigten Informationen durch den GET Befehl. Diese Vorgehensweise wird auch in den folgenden Abfragen benutzt werden.

Bei der Zahl der Mitarbeiter geht diese Arbeit unter anderem auch davon aus, dass DBpedia einen aktuellen Wert besitzt und somit zur Ermittlung der Kennzahl der aktuellste zur Verfügung stehende Wert die beste Wahl ist. Als erstes interessiert natürlich, welchen Wert die Kennziffer des Referenzunternehmens, also der Fifth Third Bancorp, aufweist. Wie in der Abfrage zu erkennen ist, setzen sich in diesem Kontext die revenues aus drei Kennzahlen zusammen, und können nicht über das GAAP Element “totalRevenues“ abgefragt werden. Außerdem werden in diesem Fall erstmals Zeiträume betrachtet. Da der Umsatzwert ein Posten in der Gewinn- und Verlustrechnung ist, und somit keinen Wert zu einem festen Zeitpunkt besitzt, muss auf den zugehörigen Zeitraum geachtet werden. Dies geschieht mit der Spezifizierung der Abfrage durch die properties “ical:dtstart“ und “ical:dtend“. In dieser Abfrage wird die Effizienz im Jahre 2009 betrachtet:

³⁶<http://dbpedia.org/property/secCik> Abruf: 18.09.2014

³⁷http://dbpedia.org/data/Fifth_Third_Bank.ntriples Abruf: 18.09.2014

```

@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .
@prefix ab: <http://alexander-buescher.de/> .

SELECT (((?revenues1+?revenues2+?revenues3)/?employees) AS ?result)
      ?revenues1 ?revenues2 ?revenues3 ?employees
WHERE { ?dbcompany ?p ?employees.
      ?p rdf:type ab:DBnumberOfEmployees .
      ?company edgar:cikNumber "35527" .
      ?company edgar:issued ?facts .
      ?facts rdfs:seeAlso ?bn1 .
      ?facts rdfs:seeAlso ?bn2 .
      ?facts rdfs:seeAlso ?bn3 .
      ?bn1 edgar:subject us-gaap-ab:InterestAndDividendIncomeOperating .
      ?bn1 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 measure:obsValue ?revenues1 .
      ?bn2 edgar:subject us-gaap-ab:InterestAndDividendIncomeSecurities .
      ?bn2 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 measure:obsValue ?revenues2 .
      ?bn3 edgar:subject us-gaap-ab:InterestAndFeeIncomeLoansAndLeases .
      ?bn3 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 measure:obsValue ?revenues3 .
    }

    { ?x edgar:cikNumber "35527" .
      ?x owl:sameAs ?dbcompany .
    } => {
    [] http:mthd httpm:GET ;
      http:requestURI ?dbcompany . } .

```

Abbildung 26: Ermittlung des Jahresumsatzes pro Mitarbeiter am Beispiel der Fifth Third Bank 2009

Output der Abfrage:

?result	?revenues1	?revenues2	?revenues3	?employees
431915	4668000000	733000000	3934000000	21613

DBpedia kann in manchen Fällen³⁸ die gleichen Informationen mit verschiedenen properties verknüpfen. Außerdem kann es vorkommen, dass manche Seiten das eine, andere Seiten jedoch ein anderes property zur Verknüpfung der von uns benötigten Informationen benutzen. Aus diesem Grund sind zwei Schritte nötig, um einerseits nicht unnötigerweise alle properties in jeder Abfrage einzeln aufzulisten und andererseits etwaige neue properties schnell in die Vorgehensweisen integrieren zu können. Im ersten Schritt müssen alle (bekannten) properties als spezieller Typ deklariert werden:

```
dbpedia-owl:numberOfEmployees rdf:type ex:NumberOfEmployees .
dbpprop:numEmployees rdf:type ex:NumberOfEmployees .
dbpprop:employees rdf:type ex:NumberOfEmployees .
```

Abbildung 27: Deklaration mehrerer properties zu gleichen Informationen zu einem Typ (Beispiel: Mitarbeiteranzahl)

In einem zweiten Schritt darf nun in den Abfragen nicht mehr das spezielle property verwendet werden, sondern der zuvor deklarierte Typ. Dadurch werden alle mit diesem verknüpften properties ausgewählt:

```
?companyDB ?p ?employees .
?p rdf:type ex:NumberOfEmployees .
```

Abbildung 28: Auswahl des deklarierten Typs in Abfragen (Beispiel: Mitarbeiteranzahl)

Diese beiden Abbildungen zeigen das exemplarische Vorgehen zur Einführung eines eigenen Typs, der nun benutzt wird, um verschiedene properties möglichst einfach in die Abfrage zu integrieren. Wenn nun beispielsweise andere properties ebenfalls zusammengefasst werden sollen, da sie dieselben Daten bereitstellen, muss dieses Vorgehen logischerweise einmalig für diese durchgeführt und die Abfragen dahingehend angepasst werden.

Der nun folgende zweite Teil dieses Abschnitts präsentiert eine Abfrage zur direkten Ermittlung derjenigen Unternehmen innerhalb der Branche des Anwendungsfalls, die einen höheren (bzw. je nach Beispiel niedrigeren) Wert besitzen und somit dem Beispielunternehmen im Bereich dieser Kennziffer überlegen sind (Konkurrenzanalyse).

³⁸Beispiel: Fifth Third Bancorp – http://dbpedia.org/page/Fifth_Third_Bank Abruf: 01.11.2014

```

@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT (((?revenues1+?revenues2+?revenues3)/?employees) AS ?result)
      ?cikNr ?name
WHERE {?company edgar:issued ?facts .
      ?facts rdfs:seeAlso ?bn1 .
      ?facts rdfs:seeAlso ?bn2 .
      ?facts rdfs:seeAlso ?bn3 .
      ?bn1 edgar:subject us-gaap-ab:InterestAndDividendIncomeOperating .
      ?bn1 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 measure:obsValue ?revenues1 .
      ?bn2 edgar:subject us-gaap-ab:InterestAndDividendIncomeSecurities .
      ?bn2 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 measure:obsValue ?revenues2 .
      ?bn3 edgar:subject us-gaap-ab:InterestAndFeeIncomeLoansAndLeases .
      ?bn3 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 measure:obsValue ?revenues3 .
      ?company owl:sameAs ?dbcompany .
      ?dbcompany ?p ?employees .
      ?p rdf:type ab:DBnumberOfEmployees .
      ?company edgar:cikNumber ?cikNr .
      ?company foaf:name ?name .
      FILTER (?result > 431915)
    }

    { ?company edgar:cikNumber ?number .
      ?company owl:sameAs ?dbcompany .
    } => {
    [] http:mthd httpm:GET ;
      http:requestURI ?dbcompany . } .

```

Abbildung 29: Konkurrenzanalyse im Bereich des Gesamtumsatzes pro Mitarbeiter

Bei dieser Abfrage wird die “FILTER“ Option in SPARQL genutzt, die – wie der Name bereits vermuten lässt – alle Ausgaben nach einer bestimmten Voraussetzung filtert. Hierbei werden in erster Linie arithmetische Einschränkungen festgelegt, die eine Identifi-

zierung der Konkurrenz vereinfachen sollen. Die Ermittlung der eigenen Kennziffer, sowie der Vergleich mit der Konkurrenz, wurden aus Gründen der Lesbarkeit in zwei Abfragen aufgeteilt, sind jedoch auch als Einzelabfrage durchführbar (vgl. Anhang Abbildung 58).

Die Ausführung wurde in diesem Fall mit den drei verschiedenen Informationen zu den Erträgen angepasst. Andere Unternehmen können ihre Erträge auch durch eine einzelne Zahl publizieren (beispielsweise `totalRevenues`), sodass die Abfrage sich diesbezüglich natürlich minimal ändern kann – das generelle Vorgehen bleibt jedoch gleich.

Das Ergebnis der Konkurrenzanalyse im Bereich des Gesamtumsatzes pro Mitarbeiter, ergibt den folgenden Output, **vor** Anwendung der FILTER Bedingung:

?result	?cikNr	?name
313435	872821	WILMINGTON TRUST CORP
322478	36270	M&T BANK CORP
332087	22356	COMMERCE BANCSHARES INC
383777	7789	ASSOCIATED BANC-CORP

Nach Anwendung der FILTER Bedingung, würde das Ergebnis der Abfrage leer sein. Grund hierfür ist, dass alle³⁹ übrigen Banken, die zur exemplarischen Beschreibung der Vorgehensweise betrachtet werden, im Bereich der aktuellen Kennzahl unter dem Wert von “431915“ der Fifth Third Bank liegen.

Die oben aufgezeigte Vorgehensweise ist in der Analyse potentieller Konkurrenten zwar in der Regel zielführend, birgt jedoch auch zwei potentielle Probleme: Erstens ist fraglich, ob die FILTER Bedingung den gewünschten Effekt hat – Unternehmen, die nur knapp unter der Kennzahl des Beispielunternehmens liegen, oder dieselbe besitzen, sind nämlich eindeutig Konkurrenten, werden so aber nicht in den Ergebnissen aufgeführt.

Weiter geben manche Banken ihre Anzahl der Mitarbeiter nicht als Integer, sondern als String (beispielsweise First Financial Bank “More than 1000“⁴⁰) an. Dieser führt logischerweise zu Problemen in dem Schritt der Abfrage, in dem die Zahlen dividiert und an die Variable ?result gebinded werden.

³⁹Da zum exemplarischen Aufzeigen der Vorgehensweise nur insgesamt fünf Banken durch das nötige owl:sameAs property mit ihren DBpedia Daten verbunden wurden, ist diese Ausgabe nicht notwendigerweise vollständig.

⁴⁰http://dbpedia.org/page/First_Financial_Bank Abruf: 21.10.2014

6.3 Unternehmenseffizienz – Pro-Kopf-Gewinn

Diese Kennziffer geht nach einem ähnlichen Muster wie die Vorherige vor. Der Fokus liegt wieder auf der Integration von Daten aus mehr als einer statischen Quelle. Wieder werden Finanzdaten aus den XBRL-Dokumenten der SEC bezogen, die Anzahl der Angestellten wiederholt aus den von DBpedia zur Verfügung gestellten Daten. Ein weiterer Aspekt, den diese Kennzahl unterstreichen soll, ist der signifikante Unterschied zwischen Umsatz und Gewinn (Verlust).

Analog zum vorhergehenden Szenario lässt sich mit einer leicht modifizierten Abfrage der eigene Wert dieser Kennzahl ermitteln:

```
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2001/http#> .
@prefix httpm: <http://www.w3.org/2001/http-methods#> .
@prefix ab: <http://alexander-buescher.de/> .

SELECT ((?profit/?employees) AS ?result) ?profit ?employees
WHERE { ?company edgar:cikNumber "35527" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn .
        ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
        ?bn ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn measure:obsValue ?profit .
        ?dbcompany ?p ?employees .
        ?p rdf:type ab:DBnumberOfEmployees .
      }

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .
```

Abbildung 30: Ermittlung des Jahresgewinns pro Mitarbeiter am Beispiel der Fifth Third Bank 2009

Wie zu erkennen ist, ist die Abfrage lediglich durch eine Änderung des zu betrachtenden Konzepts modifiziert worden. Essentiell ist wieder die Spezifizierung auf ein Anfangs- und Enddatum, um nicht versehentlich zwischen Quartals- und Jahresberichten hin und her

zu wechseln. Die Abfrage liefert folgendes Ergebnis:

?result	?profit	?employees
34099	737000000	21613

Der Wert des Pro-Kopf-Gewinns derselben Periode (2009-01-01 bis 2009-12-31) ist bereits deutlich geringer als der Umsatz pro Mitarbeiter. Als nächstes gilt es zu untersuchen, ob sich auch in der Analyse der Konkurrenz mit Änderung der Kennzahl starke Unterschiede ergeben, oder eine ähnliche Aufzählung wie in der vorhergehenden Kennzahl zu erkennen ist. Durch eine geringe Modifikation der zweiten Abfrage der vorhergehenden Kennzahl erhält man eine neue Abfrage, die Aufschluss über die zuvor gestellte Frage gibt:

```

@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .
@prefix ab: <http://alexander-buescher.de/> .

SELECT ((?profit/?employees) AS ?result) ?cikNr ?name
WHERE { ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn .
        ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
        ?bn ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn measure:obsValue ?profit .
        ?dbcompany ?p ?employees .
        ?p rdf:type ab:DBnumberOfEmployees .
        ?company edgar:cikNumber ?cikNr .
        ?company foaf:name ?name .
        FILTER (?result > 34099)
      }

{ ?x edgar:cikNumber ?number .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .

```

Abbildung 31: Konkurrenzanalyse im Bereich des Jahresgewinns pro Mitarbeiter

Auch in diesem Fall wurde in einer ersten Abfrage der eigene Wert der Kennzahl ermittelt, um diesen dann mit denen der Konkurrenz vergleichen zu können und somit Aussagen

über die mögliche Konkurrenz in diesem Segment tätigen zu können. Eine Abfrage, die beide Schritte ausführt, findet sich in Abbildung 59 im Anhang.

Auch bei dieser Kennzahl werden zuerst die Abfragen **ohne** FILTER Bedingung präsentiert:

?result	?cikNr	?name
-1646	872821	WILMINGTON TRUST CORP
24249	36270	M&T BANK CORP
35790	22356	COMMERCE BANCSHARES INC
-29191	7789	ASSOCIATED BANC-CORP

Wie erwähnt dürfen Inputs und Outputs bei einer DEA lediglich positive Werte annehmen – dies wurde im vorherigen Kapitel dadurch versucht darzustellen, da explizit von “Profits(t)” gesprochen wurde. Die XBRL Dokumenten bieten jedoch lediglich Werte mit Informationen zum “netIncomeLoss“, weswegen negative Werte in diesem Falle ausgeklammert werden müssen:

?result	?cikNr	?name
24249	36270	M&T BANK CORP
35790	22356	COMMERCE BANCSHARES INC

Reduziert man diese Tabelle durch die Anwendung des FILTERs, bleibt lediglich die Commerce Bancshares übrig, die als einzige in diesem Fall einen höheren Wert der Kennzahl besitzt, als das betrachtete Beispielunternehmen:

?result	?cikNr	?name
35790	22356	COMMERCE BANCSHARES INC

Vergleicht man nun die Ergebnisse dieser Konkurrenzanalyse mit der des vorherigen Abschnitts, kann man erkennen, dass sowohl Vorgehensweise als auch Outputvariablen sich gleichen – die Ergebnisse jedoch unterschiedlich sind. Die COMMERCE BANCSHARES INC wird als einzige Bank bei Kennzahl 3, nicht jedoch bei Kennzahl 2, nach Anwendung der FILTER Bedingung noch aufgeführt. Dies zeigt, dass diese beiden Kennziffern nicht nur dazu geeignet waren, die Vorgehensweise bei der Integration von Daten mehrerer Quellen zu zeigen, sondern auch, den Unterschied zwischen Umsatz und Profit im Kontext von Finanzanalysen nochmals zu hervorzuheben.

Ein Nachteil dieser Zahl ist das zuvor bereits angesprochene Problem, dass aufgrund der einzuhaltenden Voraussetzungen bei der Durchführung einer DEA lediglich positive Werte betrachtet werden können. Dieses Problem könnte dadurch angegangen werden, dass via XBRL gezwungenermaßen die Werte zu “netIncomeLoss“ auf beispielsweise “netIncome“ und “netLoss“ aufgeschlüsselt werden müssten.

6.4 Umsatz pro Geschäftsstelle

Analog zu den Vorgängern ergibt sich auch diese Kennzahl auf Basis der Verbindung der Datensätze der SEC und DBpedia. Auf Grundlage der Erkenntnisse der Studie von Devlin und Gerrard aus dem Jahre 2005 ([DG05]), liegt das Hauptaugenmerk auf der Anzahl der Standorte, die das Unternehmen besitzt.

Diese Kennzahl wird ebenfalls auf Basis der DEA-Methode durchgeführt werden, in welcher das Verhältnis zwischen Input (Anzahl der Standorte – bereitgestellt durch DBpedia) und Output (Umsatz in Periode t – bereitgestellt durch die SEC) angegeben wird. Mithilfe folgender SPARQL-Abfrage werden die benötigten Werte ermittelt:

```
@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT (((?revenues1+?revenues2+?revenues3)/?locations) AS ?result)
       ?revenues1 ?revenues2 ?revenues3 ?locations
WHERE { ?company edgar:cikNumber "35527" .
       ?company edgar:issued ?facts .
       ?facts rdfs:seeAlso ?bn1 .
       ?facts rdfs:seeAlso ?bn2 .
       ?facts rdfs:seeAlso ?bn3 .
       ?bn1 edgar:subject us-gaap-ab:InterestAndDividendIncomeOperating .
       ?bn1 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn1 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn1 measure:obsValue ?revenues1 .
       ?bn2 edgar:subject us-gaap-ab:InterestAndDividendIncomeSecurities .
       ?bn2 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn2 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn2 measure:obsValue ?revenues2 .
       ?bn3 edgar:subject us-gaap-ab:InterestAndFeeIncomeLoansAndLeases .
       ?bn3 ical:dtstart "2009-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn3 ical:dtend "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
       ?bn3 measure:obsValue ?revenues3 .
       ?dbcompany ?p ?locations .
       ?p rdf:type ab:DBlocations .
}

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .
```

Abbildung 32: Abfrage Umsätze sowie Anzahl der Standorte (Beispiel: Fifth Third Bank)

Zusätzlich zu dieser Abfrage, die den vorhergehenden Abfragen sehr ähnlich ist, soll gezeigt werden, dass auch in Form dieser Kennzahlen eine Entwicklung angegeben werden kann. Hierzu gilt es die Kennzahlen mit Werten der darauffolgenden Jahre zu vergleichen. Deren Abfrage kann je nach gewünschtem Zeitraum durch die Veränderung der Werte von ical:dtstart bzw. ical:dtend modifiziert werden⁴¹. Die Ausgabe für die Kalenderjahre 2009, 2010 und 2011 ergibt sich folgendermaßen:

Jahr	?result	?revenues1	?revenues2	?revenues3	?locations
2009	7131398	4668000000	733000000	3934000000	1309
2010	6836890	4489000000	658000000	3823000000	1312
2011	6406534	4218000000	600000000	3613000000	1316

Aus diesen Daten kann nun wieder eine grafische Entwicklung dargestellt werden:

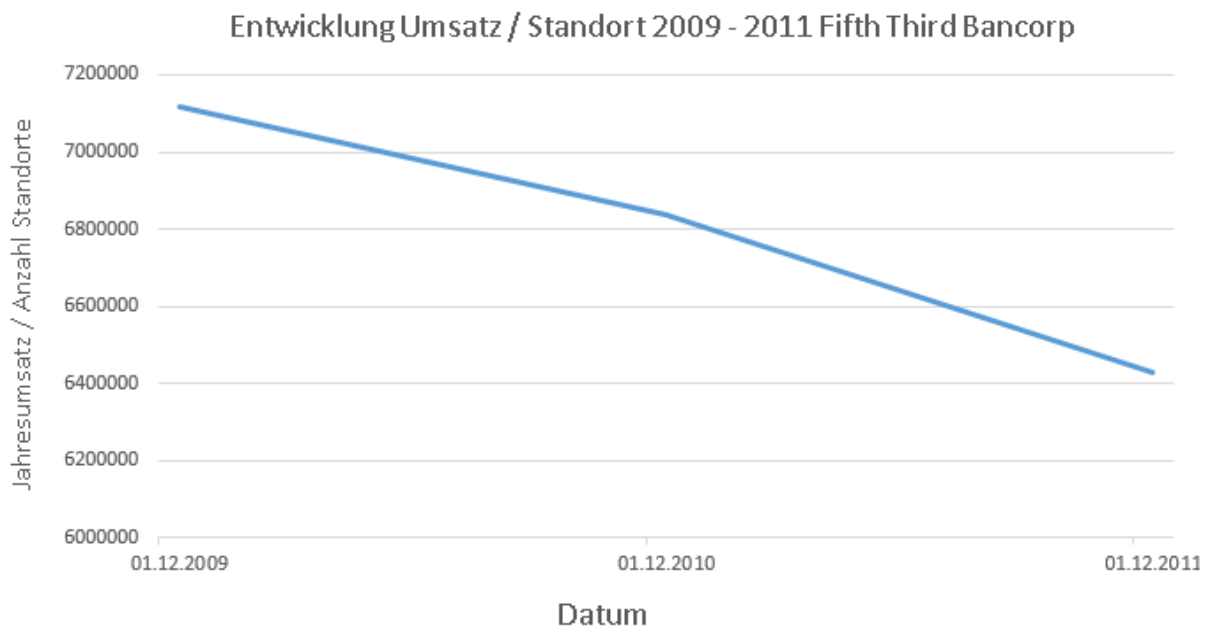


Abbildung 33: Entwicklung der Kennzahl $\frac{\text{UmsatzPeriode}(t)}{\text{AnzahlStandorte}}$ 2009 - 2011 (Beispiel: Fifth Third Bank)

Natürlich kann analog zu den beiden vorhergehenden Kennzahlen, auch mit geringer Modifikation der dortigen Abfrage, eine Analyse über die Konkurrenz durchgeführt werden (vgl. Abbildungen 29 und 31).

Ein Problem bei dieser Art der Beschreibung zur Entwicklung einer Kennzahl ist jedoch, dass mit den Zahlen gearbeitet werden muss, die man bekommt. Die Anzahl der Niederlassungen bezieht DBpedia über die Dokumente zum Jahresabschluss des Unternehmens

⁴¹Aus Gründen der Lesbarkeit finden sich diese Abfragen in Anhang A.4.

bei der SEC ⁴² von 2010. Eine Historie von jährlichen Daten gibt es nicht. An dieser Stelle wäre es wohl besser, die Anzahl der Standorte ebenfalls aus den SEC Dokumenten zu beziehen, jedoch ist der Gedanke hinter dieser Abfrage exemplarisch um die Integration von Daten für eine Kennzahl aus zwei verschiedenen Datenquellen zu zeigen.

Bezieht man die Daten der Standorte jedoch ebenfalls aus den entsprechenden Dokumenten der SEC, ändern sich Output sowie Grafik:

Jahr	?result	?revenues1	?revenues2	?revenues3	?locations
2009	7115091	4668000000	733000000	3934000000	1312
2010	6836890	4489000000	658000000	3823000000	1312
2011	6426067	4218000000	600000000	3613000000	1312

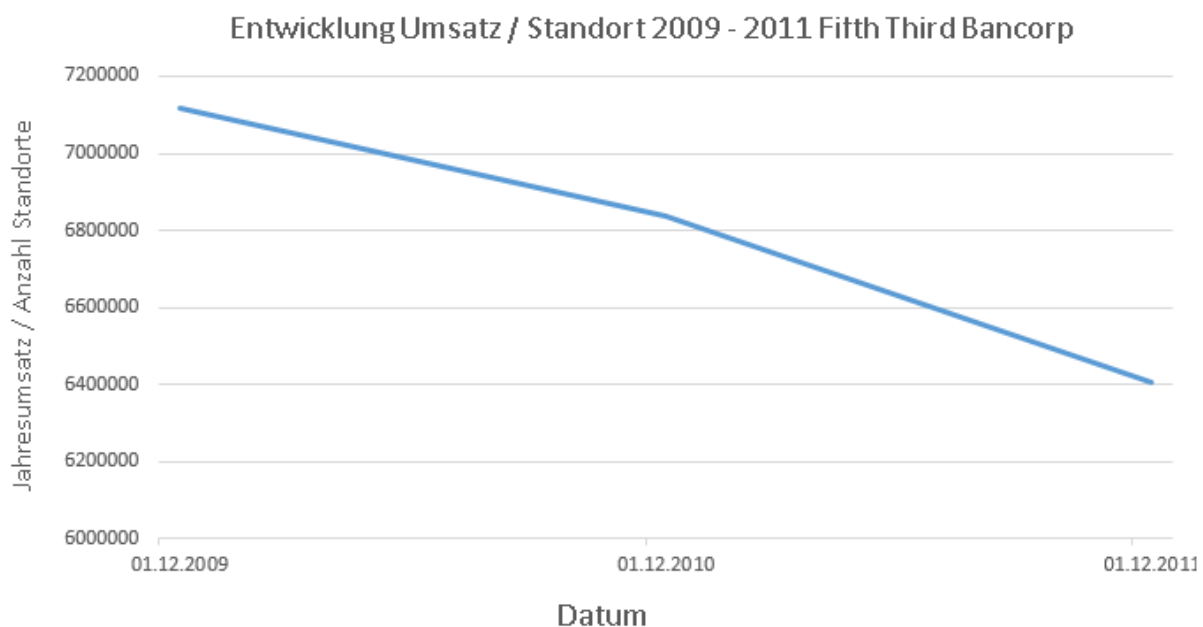


Abbildung 34: Entwicklung der Kennzahl $\frac{UmsatzPeriode(t)}{AnzahlStandorte}$ 2009 - 2011 (Beispiel: Fifth Third Bank)

Auch, wenn der Unterschied nur marginal ist, kann eine Ungenauigkeit nicht von der Hand gewiesen werden. An dieser Stelle muss also wiederholt der tradeoff zwischen gewünschter Detailtiefe sowie gewünschtem Grad der automatischen Verarbeitung stattfinden.

Da die Daten der Standorte lediglich im Fließtext der Finanzberichte, nicht jedoch als eigene XBRL-Konzepte auftauchen, wird bei maximaler Detailtiefe immer ein – wenn auch geringer – manueller Aufwand stattfinden müssen. Es gilt also situationsbedingt die

⁴²http://en.wikipedia.org/wiki/Fifth_Third_Bank#cite_note-10K-1 Abruf: 24.09.2014

Entscheidung zu treffen, ob man sich auf die Präzision der Werte, wie in diesem Fall von DBpedia, verlassen kann und möchte oder eben nicht. Die Problematik “Präzision vs. Automatisierbarkeit“ wird auch in Kapitel 7, welches sich mit der Notwendigkeit weiterer Forschung beschäftigt, betrachtet.

6.5 Effektivität der Produktion

Bei der Kennzahl “Produktion“ denkt man in der Regel nicht an Unternehmen in der Bankenbranche. Unter anderem aus diesem Grund wird diese Arbeit im nachfolgenden Teil eine Kennzahl des Anwendungsszenarios betrachten, die der Effektivität einer Produktion extrem Nahe kommt, wie auch Giokas im Jahre 2008 in seinen Ausführungen beschrieb ([Gio08]).

Während der Punkt des non-interest incomes in der Regel als eigener Wert in den XBRL-Dokumenten der Unternehmen vorkommt und somit leicht ermittelt werden kann, muss die Zusammensetzung der personnel costs geklärt werden. Diese werden per Definition als Auszahlungen an Mitarbeiter des Unternehmens bezeichnet, die im entsprechenden Zeitraum als Entlohnung für verrichtete Arbeit getätigt werden – hierzu gehören auch Prämien und Bonuszahlungen. Ferner zählen auch Steuern, Abgaben für die Sozialversicherung sowie freiwillige als auch verordnete Sozialabgaben dazu⁴³.

Diese Zusammensetzung zu klären ist wichtig. Das Konzept der personnel costs, welches durch die US-GAAP bereitgestellt wird, heißt “LaborAndRelatedExpense“, wird jedoch eben nicht von jedem Unternehmen in den XBRL-Dokumenten verwendet. Ist dies der Fall, muss diese Zahl eigens aus den oben genannten Werten zusammengesetzt werden. Im Falle der Angabe der Zahl, wie es beispielsweise die Wilmington Trust Corp.⁴⁴ in ihren Berichten macht, ist die Abfrage dieser Kennzahl simpel:

```
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

SELECT (xsd:double((xsd:double(?nonInterestIncome))/(xsd:double(?personnelCosts))) AS ?result)
       ?nonInterestIncome ?personnelCosts
WHERE { ?company edgar:cikNumber "872821" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn1 .
        ?facts rdfs:seeAlso ?bn2 .
        ?bn1 edgar:subject us-gaap-ab:LaborAndRelatedExpense .
        ?bn1 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn1 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn1 measure:obsValue ?personnelCosts .
        ?bn2 edgar:subject us-gaap-ab:NoninterestIncome .
        ?bn2 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn2 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn2 measure:obsValue ?nonInterestIncome .
}
```

Abbildung 35: Abfrage Personalaufwand und nicht zu verzinsendes Einkommen (Beispiel: Wilmington Trust Corp.)

⁴³<http://www.likeforex.com/glossary/w/personnel-costs-127346> Abruf: 29.09.2014

⁴⁴<https://www.wilmingtontrust.com/wtcom/> Abruf: 29.09.2014

Bei dieser Kennzahl ist es wichtig mit den entsprechenden Einheiten zu arbeiten, um eine angemessene Detailtiefe zur Analyse gewährleisten zu können⁴⁵. Der dazugehörige Output ergibt:

?result	?nonInterestIncome	?personnelCosts
1.262	3627000000	2874000000

Wie bereits zuvor erwähnt, muss die Abfrage angepasst werden, wenn das publizierende Unternehmen keine US-GAAP verwendet, wie es beispielsweise bei der Fifth Third Bancorp der Fall ist. Hierbei muss die Abfrage modelliert werden:

```
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix fitb: <http://www.sec.gov/Archives/edgar/data/35527/000119312511049817/fitb-20101231.xsd#> .

SELECT (xsd:double((xsd:double(?nonInterestIncome))/(xsd:double(?wages+?benefits))) AS ?result)
       ?nonInterestIncome ?wages ?benefits
WHERE { ?company edgar:cikNumber "35527" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn1 .
        ?facts rdfs:seeAlso ?bn2 .
        ?facts rdfs:seeAlso ?bn3 .
        ?bn1 edgar:subject us-gaap-ab:OtherLaborRelatedExpenses .
        ?bn1 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn1 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn1 measure:obsValue ?benefits .
        ?bn2 edgar:subject us-gaap-ab:NoninterestIncome .
        ?bn2 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn2 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn2 measure:obsValue ?nonInterestIncome .
        ?bn3 edgar:subject fitb:SalariesWagesAndIncentives .
        ?bn3 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn3 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn3 measure:obsValue ?wages .
}
```

Abbildung 36: Abfrage Personalaufwand und nicht zu verzinsendes Einkommen (Beispiel: Fifth Third Bancorp)

Der dazugehörige Output:

?result	?nonInterestIncome	?wages	?benefits
2.90	4782000000	1339000000	311000000

Der größte Unterschied in dieser Abfrage ist der erstmalige Einbezug eines eigenen Konzeptes für dieses Anwendungsszenario. Auf die Vorteile, die eigene Konzepte mit sich

⁴⁵Der Wert der Variable ?result wurde aus Gründen der Lesbarkeit auf drei Nachkommastellen gerundet.

bringen, wurde in der Einleitung dieser Arbeit bereits hingewiesen – hinsichtlich des Ziels, sich von manuellen Abfragen zu entfernen und diese möglichst automatisierbar zu machen, können sie eher hinderlich sein, wie auch in Kapitel 7, welches sich mit Gebieten, in denen weiter geforscht werden muss beschäftigt, aufgezeigt.

Analog zur vorhergehenden Kennzahl, kann auch hier durch geringe Modifikationen der Start- und Endzeitpunkte die Abfrage über mehrere Jahresberichte durchgeführt werden, welche zu folgenden Werten führt⁴⁶:

Jahr	?result	?nonInterestIncome	?personnelCosts
2009	1.262	3627000000	2874000000
2010	1.338	3967000000	2965000000

Selbstverständlich können auch Kennzahlen als Entwicklungsreihe dargestellt und interpretiert werden:

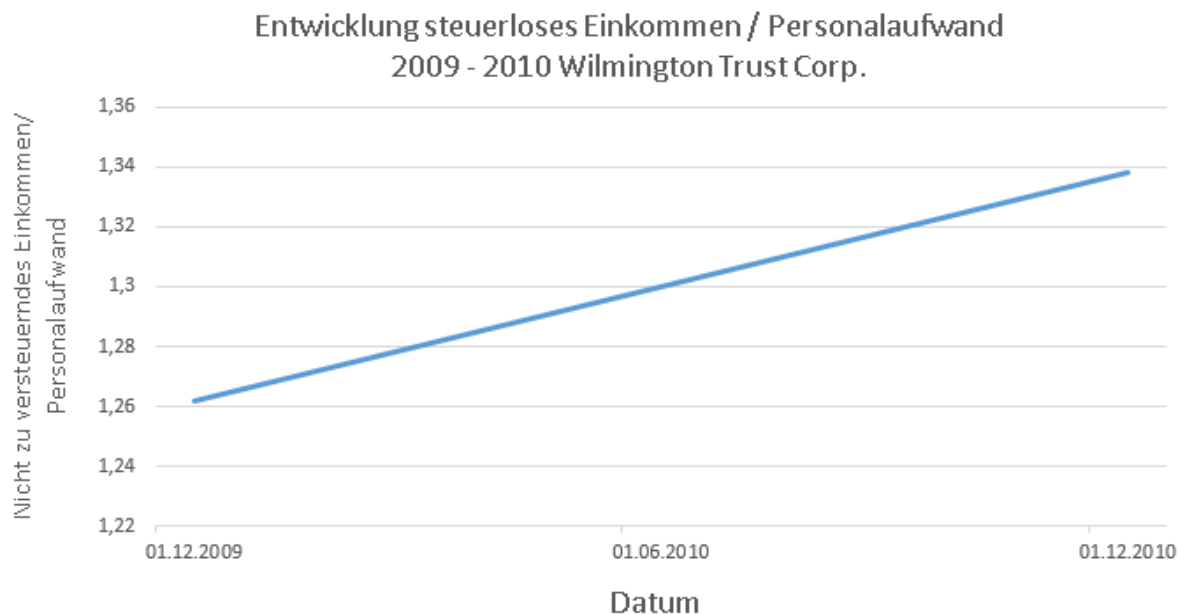


Abbildung 37: Entwicklung der Kennzahl $\frac{nonInterestIncome}{personnelCosts}$ 2009 - 2010 (Beispiel: Wilmington Trust Corp.)

Ebenfalls ist für diese Zahl eine Analyse der Kennzahlen der Konkurrenz, wie in den vorherigen Teilen 2 und 3 dieses Kapitels, denkbar und sinnvoll. Diese wäre mit geringen Konfigurationen der Abfragen, wie sie in Anhang A.4 durch die Abbildungen 58 und 59 präsentiert werden, möglich.

⁴⁶An dieser Stelle wurde zu Illustrationszwecken das erste Abfragebeispiel dieser Kennzahl gewählt und modifiziert.

6.6 Erzielte Gewinne / Verluste über mehrere Perioden

Im Gegensatz zur Methode der ersten Kennzahl, macht es bei betrachteten Änderungen von Werten eines bestimmten Konzepts innerhalb einer Periode wenig Sinn eine Zeitreihe aufzustellen. Dies ist insbesondere aus dem Grund sinnvoll, da bei Variablen, die in den Dokumenten positive (Gewinne) wie auch negative (Verluste) Werte annehmen können, eine Entwicklungsreihe nicht aussagekräftig ist. Summiert man diese nun über einen längeren Zeitraum auf, erhält man eine aussagekräftigere Zahl, als wenn nur einzelne Wertveränderungen in Perioden betrachtet werden. Der erste Teil dieses Abschnitts betrachtet jedoch die Abfrage einzelner Werte, ehe im zweiten Teil dann gezeigt wird, warum der Vergleich über einen längeren Zeitraum sinnvoller erscheint. Die Abfrage der einzelnen Werte via SPARQL gestaltet sich relativ simpel, jedoch macht es in diesem Zusammenhang durchaus Sinn, die Ergebnisse der Abfrage hinsichtlich der bindings der Variable `?dateBegin` zu sortieren, um möglichst leicht Quartals- und Jahresberichte identifizieren und voneinander trennen zu können:

```
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

SELECT ?value ?dateBegin ?dateEnd
WHERE { ?company edgar:cikNumber "35527" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn .
        ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
        ?bn ical:dtstart ?dateBegin .
        ?bn ical:dtend ?dateEnd .
        ?bn measure:obsValue ?value .
      }
ORDER BY ASC(?dateBegin)
```

Abbildung 38: Abfrage Gewinne (Verluste) einzeln (Beispiel: Fifth Third Bank)

?value	?dateBegin	?dateEnd
1076000000	2007-01-01	2007-12-31
-2113000000	2008-01-01	2008-12-31
290000000	2008-01-01	2008-09-30
-560000000	2008-07-01	2008-09-30
500000000	2009-01-01	2009-03-31
737000000	2009-01-01	2009-12-31
835000000	2009-01-01	2009-09-30
-970000000	2009-07-01	2009-09-30
-100000000	2010-01-01	2010-03-31

Diese Daten liefern Informationen darüber, in welchen Perioden ein Unternehmen hinsichtlich Profit bzw. Verlust besonders gut oder besonders schlecht abgeschnitten hat. Besonders wichtig ist die Betrachtung des Zeitraums. In dieser Tabelle wird beispielsweise nicht zwischen Werten aus Jahres- und Quartalsberichten unterschieden (siehe beispielsweise Zeile 1 und 3).

Beispielsweise für Investoren ist insbesondere die langfristige Profitabilität eines Unternehmens interessant – eine Aussage über diese kann mithilfe der Ergebnisse der zweiten Abfrage gegeben werden. Wieder wird die Abfrage durch arithmetische Operationen zwischen den einzelnen Variablen ergänzt, um das gewünschte Gesamtergebnis zu erhalten:

```
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .

SELECT ((?value1+?value2+?value3+[...]) AS ?result)
WHERE { ?bnFact1 measure:obsValue ?value1 .
        ?bnFact2 measure:obsValue ?value2 .
        ?bnFact3 measure:obsValue ?value3 .
        ...
}
```

Abbildung 39: Allgemeine Berechnung von Wertveränderungen über mehrere Perioden

Im vorliegenden Fall würde sich das binding der Variable ?result wie folgt errechnen: $1076000000 - 2113000000 + 290000000 - 560000000 + 500000000 + 737000000 + 835000000 - 970000000 - 100000000 = 451000000$ Wieder besteht das Problem, dass sich Zahlen aus verschiedenen Perioden überschneiden. Werden also nur die Werte der Jahresabschlüsse betrachtet, die am Start- und Endzeitpunkt der jeweiligen Periode zu erkennen sind, kann aus dem Zeitraum 2007-01-01 bis 2009-12-31 ein Gesamtwert von -300.000.000\$ ($1076000000 + (-2113000000) + 737000000$) erhalten werden.

Die beiden großen Nachteile bei dieser Vorgehensweise liegen auf der Hand: Ohne genauere Kenntnis über die Daten, die die relevanten financial Facts eingrenzen, ist eine derartige Abfrage nicht möglich. Daher ist es unerlässlich, zuerst alle relevanten Werte – bestenfalls sortiert (siehe Abbildung 38) – ermitteln zu lassen. Ferner muss die Problematik betrachtet werden, dass bei Abfragen die Daten von Quartals- sowie Jahresberichten zusammen ausgegeben werden können. Hier ist es wichtig, mit den properties “ical:dtstart“ und “ical:dtend“ möglichst präzise Abfragen durchzuführen.

6.7 Präzision von Angaben verschiedener Quellen

Die Integration von Daten aus verschiedenen Quellen bringt vor allem für Finanzanalysten viele weitere Möglichkeiten mit sich. Wie die vorangehenden Zahlen gezeigt haben, können nicht mehr nur starre Finanzzahlen und eingeschränkte Kennzahlen (beispielsweise ROA, ROE) berechnet werden, sondern es eröffnen sich ganz neue Kombinations- und somit Interpretationsmöglichkeiten bezüglich der Effizienz von Unternehmen.

Allerdings birgt diese Integration auch potentielle Gefahren und Schwachstellen. Für Finanzanalysten ist einer der wichtigsten Punkte, dass die Daten, die sie in ihre Berichte einbeziehen, verifizierbar und zuverlässig sind. Zu diesen Punkten gehört auch, dass diese Konsistenz in Bezug auf die Verfügbarkeit mehrerer Quellen aufweisen können – ein Punkt, der oftmals unterschätzt wird. Ein Bericht wirkt enorm unglaubwürdig, wenn bereits nach kurzer Suche in einer anderen Quelle ein unterschiedlicher Wert zu einem bestimmten Datum gefunden wird, der eigentlich identisch sein sollte. Daher gilt es zu versuchen, Daten aus verschiedenen Quellen zeitlich möglichst genau abgleichen zu können. Dieser Schritt wird in diesem Abschnitt des Anwendungsfalls präsentiert.

Ein Abgleich kann im Szenario der Arbeit nur mit sehr allgemeinen Angaben geschehen, da detaillierte Finanzberichte von DBpedia nicht bereitgestellt werden. Somit beschränken sich die Ansätze zur Veranschaulichung auf die Daten des Netto-Einkommens (`netIncome`) der Fifth Third Bancorp. Problematisch ist in diesem Zusammenhang wieder der Zeitraum der Erhebung. Da in den Datensätzen von DBpedia selbst kein Zeitpunkt bestimmt ist, gilt es dies auf der entsprechenden Wikipedia Seite in den Quellen zu prüfen. Die hier betrachteten Werte stammen laut Wikipedia aus der 10-K Form des Jahres 2010⁴⁷. Es gibt zwei Abfragen, die prüfen können, ob Werte aus verschiedenen Quellen übereinstimmen:

⁴⁷http://en.wikipedia.org/wiki/Fifth_Third_Bank#cite_note-10K-1 Abruf: 24.09.2014

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix ab: <http://alexander-buescher.de/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ?netIncomeLossSec ?netIncomeLossDBpedia
WHERE {
  ?company edgar:cikNumber "35527" .
  ?company edgar:issued ?facts .
  ?facts rdfs:seeAlso ?bn .
  ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
  ?bn measure:obsValue ?netIncomeLossSec .
  ?bn ical:dtstart "2010-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn ical:dtend "2010-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
  ?dbcompany ?p ?netIncomeLossDBpedia .
  ?p rdf:type ab:DBnetIncome .

}

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?dbcompany . } .

```

Abbildung 40: Vergleich von Angaben zum netIncome von DBpedia und SEC im Jahresbericht 2010 der Fifth Third Bancorp

Anhand des Outputs dieser Abfrage kann manuell überprüft werden, ob die beiden Werte übereinstimmen. Sollte dies nicht der Fall sein, kann außerdem entschieden werden, ob der Unterschied eventuell Rundungsfehlern zuzuordnen ist, oder, ob es sich tatsächlich um zwei unterschiedliche Werte handelt, die normalerweise gleich sein sollten.

?netIncomeLossSec	?netIncomeLossDBpedia
753000000	7.53E8

Die zweite Möglichkeit ist nicht mehr manuell. Hingegen lässt sich durch die Abfrage automatisch prüfen, ob die Werte identisch sind. Hierbei wird der ASK Operator⁴⁸ benutzt, welcher weder Triple konstruiert, noch bindings zurückgibt. Hingegen prüft ASK, ob es Ergebnisse gibt, die der Abfrage genügen. Ist dies der Fall, wird *true* zurückgegeben, andernfalls *false*.

⁴⁸<http://www.w3.org/TR/sparql11-query/#ask> Abruf: 24.09.2014

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix ab: <http://alexander-buescher.de/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

ASK
WHERE { ?company edgar:cikNumber "35527" .
        ?company edgar:issued ?facts .
        ?facts rdfs:seeAlso ?bn .
        ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
        ?bn measure:obsValue ?netIncomeLossSec .
        ?bn ical:dtstart "2010-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn ical:dtend "2010-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?dbcompany ?p ?netIncomeLossDBpedia .
        ?p rdf:type ab:DBnetIncome .
        FILTER (?netIncomeLossSec = ?netIncomeLossDBpedia)
}

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?dbcompany . } .

```

Abbildung 41: Vergleich von Angaben zum netIncome von DBpedia und SEC im Jahresbericht 2010 der Fifth Third Bancorp mithilfe des ASK Operators

Diese Abfrage illustriert zugleich angemessen das Problem, das zwischen den beiden Möglichkeiten der Überprüfung entstehen kann. Wie durch die obige Tabelle, die die Ausgabe von Abbildung 40 angibt, gezeigt wird, sind die beiden Werte inhaltlich zwar gleich, werden jedoch auf unterschiedliche Art und Weise dargestellt. Die ASK-Abfrage würde *false* zurückgeben, obwohl die Informationen der Werte identisch sind. Die Möglichkeit dies zu umgehen, ist jedoch gegeben. Beide Werte könnten (bzw. müssten) vor dem Vergleich zuerst in den selben Datentyp gecasted werden, um die Problematik von gleichen Informationen in unterschiedlicher Darstellung zu umgehen.

Bei diesem Vorgehen muss also abgewogen werden, ob man die manuelle Vergleichsarbeit oder eher “ungenauere” Ausgaben bereit ist in Kauf zu nehmen. Eine allgemeine Empfehlung ist hierbei schwer zu geben – die Entscheidung sollte wohl in Abhängigkeit der Anzahl der zu prüfenden Werte und auch des Aufwands der Abfrage, also gänzlich situationsbedingt, getroffen werden.

6.8 Aufbau des Gesamtunternehmens

Mit den zur Verfügung stehenden Daten der DBpedia Datenbank, ist es möglich die Gesamtstruktur aller dort gelisteten Unternehmen detailliert aufzuschlüsseln. Die Idee dahinter ist einleuchtend: Nicht nur die geographische Bestimmung der größten Abnehmermärkte ist wichtig, auch eigene Teile des Unternehmens können durch lokale Krisen (vgl. Nahost Konflikt / Russland Embargo) erheblich beeinträchtigt werden. Daher ist ein Finanzbericht erst dann als vollständig anzusehen, wenn auch dieser Teil betrachtet wurde. Glücklicherweise bietet DBpedia die Möglichkeit, Tochterunternehmen (bzw. Elternunternehmen, je nach Ansicht) mit einzutragen. Durch die bereits zuvor eingerichtete Verbindung innerhalb des Datastores zwischen DBpedia und SEC Dokumenten, ist es ohne größeren Aufwand möglich, alle Eltern- und Tochterunternehmen, samt ihren Hauptsitzen zu erhalten:

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix ab: <http://alexander-buescher.de/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ?subsidiary ?subLocation ?subEmp ?parent ?parLocation ?parEmp
WHERE { ?company edgar:cikNumber [number] .
        ?dbcompany ?p0 ?parent .
        ?p0 rdf:type ab:DBparentCompany .
OPTIONAL(?parent ?p1 ?parLocation .
        ?p1 rdf:type ab:DBlocation .)
OPTIONAL(?parent ?p2 ?parEmp .
        ?p2 rdf:type ab:DBnumberOfEmployees .)
OPTIONAL(?subsidiary ?p3 ?dbcompany .
        ?p3 rdf:type ab:DBparentCompany .)
OPTIONAL(?subsidiary ?p4 ?subLocation .
        ?p4 rdf:type ab:DBlocation .)
OPTIONAL(?subsidiary ?p5 ?subEmp .
        ?p5 rdf:type ab:DBnumberOfEmployees .)
}

{ ?x edgar:cikNumber [number] .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .

{ ?dbcompany ?p6 ?parent .
  ?p6 rdf:type ab:DBparentCompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?parent . } .

{ ?subsidiary ?p7 ?dbcompany .
  ?p7 rdf:type ab:DBparentCompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?subsidiary . } .

```

Abbildung 42: Abfrage zu zugehörigen Unternehmen der Gruppe, zusammen mit deren Hauptsitz und Anzahl der Mitarbeiter

Der Wert “number“ in eckigen Klammern nimmt den CIK des gewünschten Unternehmens als Integer ein. Die OPTIONAL tags ermöglichen es, die URIs der entsprechenden Unternehmen auch dann als Ergebnis der Abfrage zu erhalten, wenn in der Datenbank keine Werte für die “location“ sowie “numberOfEmployees“ properties gesetzt wurden. Ferner ist es wichtig, wie in Kapitel 4 bereits erläutert, zusätzlich die Anzahl der Mitarbeiter abzufragen, um eine Aussage über die Größe des betrachteten Unternehmens tätigen zu können. Danach kann auch der mögliche Einfluss eines Ereignisses auf die gesamte Unternehmenshierarchie abgeschätzt werden.

Im Anwendungsfall sähe eine spezifizierte Abfrage der Wilmington Trust⁴⁹ folgendermaßen aus:⁵⁰

```

@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix ab: <http://alexander-buescher.de/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ?subsidiary ?subLocation ?subEmp ?parent ?parLocation ?parEmp
WHERE { ?company edgar:cikNumber "872821" .
        ?dbcompany ?p0 ?parent .
        ?p0 rdf:type ab:DBparentCompany .
OPTIONAL(?parent ?p1 ?parLocation .
        ?p1 rdf:type ab:DBlocation .)
OPTIONAL(?parent ?p2 ?parEmp .
        ?p2 rdf:type ab:DBnumberOfEmployees .)
OPTIONAL(?subsidiary ?p3 ?dbcompany .
        ?p3 rdf:type ab:DBparentCompany .)
OPTIONAL(?subsidiary ?p4 ?subLocation .
        ?p4 rdf:type ab:DBlocation .)
OPTIONAL(?subsidiary ?p5 ?subEmp .
        ?p5 rdf:type ab:DBnumberOfEmployees .)
}

{ ?x edgar:cikNumber "872821" .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .

{ ?dbcompany ?p6 ?parent .
  ?p6 rdf:type ab:DBparentCompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?parent . } .

{ ?subsidiary ?p7 ?dbcompany .
  ?p7 rdf:type ab:DBparentCompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?subsidiary . } .

```

Abbildung 43: Abfrage zu zugehörigen Unternehmen der Gruppe, zusammen mit deren Hauptsitz und Anzahl der Mitarbeiter (Beispiel: Wilmington Trust)

⁴⁹<https://www.wilmingtontrust.com> Abruf: 23.09.2014

⁵⁰In diesem Fall wurde vom Beispiel der Fifth Third Bank abgewichen, da diese in ihrem DBpedia Eintrag keine Tochterunternehmen angibt.

Aufgrund nur eines Eintrags ist der Output dieser Abfrage sehr überschaubar:

?parent	?parLocation	?parEmp
dbpedia:M&T_Bank	dbpedia:Buffalo_New_York	15666

Da der Anwendungsfall zwar einerseits branchenspezifisch sein soll, zugleich aber auch auf andere Unternehmensbranchen übertragbar sein muss, ist dieser Punkt trotz des knappen Outputs interessant. So ist beispielsweise über den DBpedia Eintrag von Ford⁵¹ eine Liste von mehr als 15 Tochterunternehmen aufzufinden, die alle ihren eigenen Teil zum Gesamtunternehmen beitragen. Sollte also eines von diesen mit erheblichen wirtschaftlichen oder sonstigen negativen Einflüssen zu kämpfen haben, kann und wird dies je nach Größe die Gesamtunternehmung mehr oder weniger empfindlich treffen.

⁵¹http://dbpedia.org/page/Ford_Motor_CompanyAbruf: 23.09.2014

6.9 Entwicklung des Aktienkurses

Die neunte und zugleich letzte der betrachteten Kennzahlen im Anwendungsszenario dieser Arbeit birgt eine weitere Datenquelle, die es zu integrieren gilt. Historische Daten des Aktienkurses sind aus der Yahoo! Finance Datenbank abrufbar⁵². Einziges Problem: Diese sind nicht im RDF Format, können also nicht ohne Weiteres integriert werden. Abhilfe hierbei schafft der Yahoo! Finance wrapper⁵³, der die benötigten Daten im RDF / XML Format bereitstellen kann. Zusätzlich gibt es eine ähnliche Problematik wie bei Kennzahl Nummer 2: Die Verbindung zwischen SEC sowie DBpedia Daten mit den Daten des Finance wrappers muss wieder durch das owl:sameAs property festgelegt werden. Unternehmen werden in der Yahoo! Finance Datenbank eindeutig durch sogenannte “tags“ identifiziert – im Fall der Fifth Third Bancorp wurde der tag “FITB“ festgelegt⁵⁴.

Da diese Arbeit zur Verbindung der beiden Instanzen auf das owl:sameAs property zurückgreift, reicht es eine der beiden bereits gekoppelten Datenquellen mit der jeweiligen Yahoo! Finance ID zu verbinden. Dies hat den Grund, dass owl:sameAs transitive Eigenschaften besitzt (vgl. [HH10]). Eine Möglichkeit zum Verbinden der beiden Quellen, und somit dem ersten Schritt der Datenintegration, zeigen die folgenden Zeilen:

```
<http://edgarwrap.ontologycentral.com/cik/35527#id>
<http://www.w3.org/2002/07/owl#sameAs>
<http://yahoofinancewrap.appspot.com/ticker/FITB#id> .
```

Abbildung 44: Verbindung der Yahoo! Finance Daten in RDF mit dem datastore (Beispiel: Fifth Third Bancorp)

In diesem Teil werden nun die Aktienkurse der Fifth Third Bank des Jahres 2009 betrachtet. Dies stellt einen angemessenen Abschluss des Anwendungsszenarios dar, da nach den Funden zu den vorhergehenden Kennzahlen nun Mutmaßungen und Abschätzungen über die Gründe der Schwankungen im Aktienkurs möglich sind. Ebenfalls ergibt sich die Möglichkeit, vorherige Verlaufskurven (wie beispielsweise Abbildung 24) mit dem Verlauf des Aktienkurses zu vergleichen und daraus Schlüsse ziehen zu können.

Die folgende Abfrage ermittelt nun in einem definierten Zeitraum (im vorliegenden Fall von 2009-01-01 bis 2009-12-31) den Aktienwert bei Börsenschluss für jeden Werktag, wie durch Yahoo! bereitgestellt:

⁵²<http://finance.yahoo.com/> Abruf: 23.09.2014

⁵³<http://yahoofinancewrap.appspot.com/> Abruf: 23.09.2014

⁵⁴<http://finance.yahoo.com/q?s=FITB> Abruf: 23.09.2014

```

@prefix yhof: <http://yahoofinancewrap.appspot.com/vocab/yahoo#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ?date ?value
WHERE {
  <http://yahoofinancewrap.appspot.com/archive/FITB/2009#ds> rdfs:seeAlso ?bn .
  ?bn yhof:subject <http://yahoofinancewrap.appspot.com/vocab/stock#Close> .
  ?bn dcterms:date ?date .
  ?bn measure:obsValue ?value .
}

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?yhcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?yhcompany . } .

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?yahooDB .
  ?yahooDB yhvocab:issued <http://yahoofinancewrap.appspot.com/archive/FITB/2009#ds> .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI <http://yahoofinancewrap.appspot.com/archive/FITB/2009#ds> . } .

```

Abbildung 45: Abfrage von notierten Werten der Fifth Third Bancorp bei Börsenschluss über das Jahr 2009

Durch den CIK eines Unternehmens kann die Abfrage auch dann durchgeführt werden, wenn ein owl:sameAs property nur zwischen DBpedia und Yahoo! eingefügt wurde. Dies ist aufgrund der Transitivität von owl:sameAs möglich. Ein Auszug des Outputs zur oben beschriebenen SPARQL-Query sieht wie folgt aus:

?date	?value
2009-12-31	9.75
2009-12-30	9.79
2009-12-29	9.92
2009-12-28	9.97
2009-12-24	10.23
.	.
.	.
.	.

Mit Eckpunkten aus den Werten der gesamten Tabelle, lässt sich der Verlauf des Aktienkurses für das Jahr 2009 durch folgendes Diagramm illustrieren:

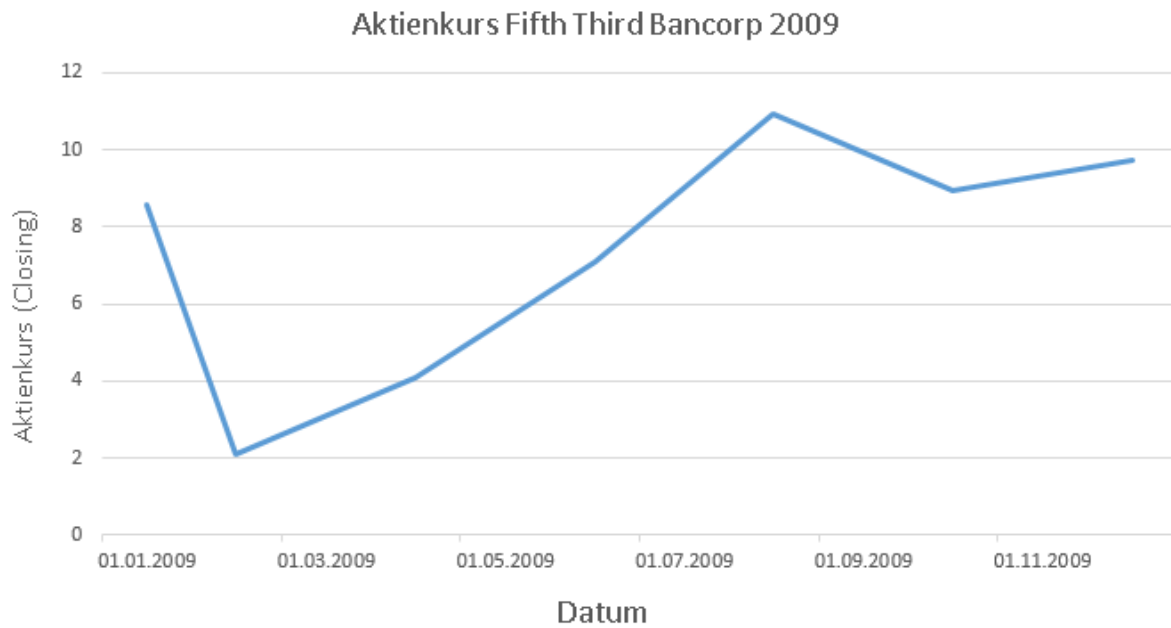


Abbildung 46: Verlauf des Aktienkurses 2009 (Beispiel: Fifth Third Bancorp)

Ebenfalls ist es durch die Abfrage einfach möglich, die Kursveränderung in einer vorgegebenen Periode zu betrachten - hierbei vorausgesetzt, dass man den ersten und letzten Werktag der gewünschten Periode genau benennen kann:

```

@prefix yhof: <http://yahoofinancwrap.appspot.com/vocab/yahoo#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ((?valueBegin - ?valueEnd) AS ?value) ?valueBegin ?valueEnd
WHERE { ?fact rdfs:seeAlso ?bn1 .
        ?fact rdfs:seeAlso ?bn2 .

        ?bn1 yhof:subject <http://yahoofinancwrap.appspot.com/vocab/stock#Close> .
        ?bn1 dcterms:date "2009-01-02" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn1 measure:obsValue ?valueBegin .

        ?bn2 yhof:subject <http://yahoofinancwrap.appspot.com/vocab/stock#Close> .
        ?bn2 dcterms:date "2009-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
        ?bn2 measure:obsValue ?valueEnd .
      }

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?yhcompany .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?yhcompany . } .

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?yahooDB .
  ?yahooDB yhvocab:issued ?fact .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?fact . } .

```

Abbildung 47: Abfrage zur Änderung des Aktienkurses im Jahres 2009 der Fifth Third Bancorp

Der zugehörige Output der Abfrage:

?value	?valueBegin	?valueEnd
1.16	8.59	9.75

Logischerweise ist dieses Vorgehen auf beliebige von Yahoo! Finance notierte Werte, sowie auf beliebig lange (kurze) Perioden adaptierbar.

Zugegeben – die Illustration eines Aktienkurses ist, wie bereits erwähnt, keine bahnbrechende Neuerung. Allerdings gilt es in jedem Finanzbericht auf diesen Kurs zu achten und zu versuchen, Schwankungen, wie sie in dieser Abbildung sehr gut zu erkennen sind, erklären zu können. Aus diesem Grund ist die Ermittlung des Aktienkurses aus den Yahoo! Finance Daten zur Abrundung des exemplarischen Finanzberichtes angemessen.

Eine exemplarische Zusammenfassung der durchgeführten Abfragen im Anwendungsfall könnte für verschiedene Unternehmen folgendermaßen aussehen (alle Angaben beziehen sich auf die Periode 2009-01-01 bis 2009-12-31 bzw. wurden zum Zeitpunkt 2009-12-31 betrachtet):

Bank Name	totalAssets (T\$)	$\frac{Profits(t)}{numberOfEmployees}$	$\frac{netIncomeLoss(t)}{numberOfEmployees}$
Fifth Third Bank	113380000	431915	34099
Wilmington Trust	11097100	–	–1646
M&T Bank Corp	68880399	–	24249
Commerce Bancshares	18120189	–	35767
Associated Banc–Corp	22874142	–	–29191
Bancorp South	13167867	–	18615
Bank of NY Mellon	212224000	–	–15960

Bank Name	$\frac{Profits(t)}{numberOfLocations}$	$\frac{Non-interestIncome}{PersonnelCosts}$	netIncomeLoss(t)
Fifth Third Bank	561737	–	737000000
Wilmington Trust	–	1.2620	–4400000
M&T Bank Corp	–	1.0461	379891000
Commerce Bancshares	–	1.1640	169075000
Associated Banc–Corp	–580876	1.1530	–131859000
Bancorp South	–	0.9876	82729000
Bank of NY Mellon	–	1.0153	–814000000

Bank Name	Unternehmensaufbau	Änderung Aktienwert(t)
Fifth Third Bank		1.16
Wilmington Trust	parent: M&T Bank Corp employees: 15666 location: New_York	–
M&T Bank Corp		9.57
Commerce Bancshares		–4.86
Associated Banc–Corp		–10.38
Bancorp South		–0.21
Bank of NY Mellon		–0.55

Die Teil der Tabelle, in denen keine Werte eingetragen sind, deuten auf Lücken in den Datensätzen hin. Dies kann auch den Grund haben, dass das Unternehmen die jeweilige Kennzahl unter einem anderen (bzw. eigenen) Konzept publiziert, als dem in der Abfrage angegebenen. Diese Problematik wird auch am Ende von Punkt 2 des nächsten Kapitels, das Ansätze zur weiteren Nachforschung beinhaltet, angesprochen.

7 Ansätze zur weiteren Forschung

Der im vorherigen Kapitel präsentierte Anwendungsfall zeigt ein Potential, welches die Integration verschiedener Datenquellen mittels Linked Data im Bereich der Analyse von Finanzperformance birgt. Neben den zweifelsohne vorhanden Vorteilen, gibt es allerdings in diesem forschungstechnisch relativ jungen Gebiet sicherlich auch weitere Aspekte, die noch betrachtet werden müssen. In diesen müssten zusätzliche Nachforschungen betrieben werden, die jedoch den Rahmen dieser Arbeit sprengen würden:

7.1 Globale Anwendbarkeit

Die Grundlage des im Anwendungsfall benutzten Datastores sind Finanzdokumente der SEC, welche ihre publizierenden Unternehmen über das System des Central Index Key (CIK) eindeutig identifiziert. Problematisch hierbei ist, dass diese Art der Identifikation lediglich auf solche Unternehmen beschränkt ist, die als US-Aktiengesellschaften zählen. Möchte man nun also die Performance von Unternehmen, die lediglich in Übersee an der Börse gemeldet sind mit solchen vergleichen, die in ausschließlich in Europa eingetragene Aktienunternehmen sind, wird eine andere Identifikationsmethode gewählt werden müssen.

Ein weiteres Problem im Zusammenhang der CIKs, ist deren Vergabe. In der Datenbank der SEC gibt es diverse Unternehmen, die zum gleichen Namen mehrere und unterschiedliche CIKs haben. Dies kann aus verschiedenen Gründen geschehen sein.

Die veröffentlichten Finanzberichte von Unternehmen werden von der SEC chronologisch angeordnet. Manche Unternehmen teilen ihre Berichte in Perioden (wie zum Beispiel 1991–1996, 1996–2001 usw.) auf und bekommen nach Änderung der Unternehmensstruktur, egal ob gering oder sehr ausschlaggebend, einen neuen CIK zugewiesen (gekennzeichnet durch den Vermerk “filings through yyyy-mm-dd“)⁵⁵.

Die zweite Möglichkeit ist die Angabe von Tochterunternehmen, was auch in Punkt 8 des Anwendungsfalls beleuchtet wurde. In diesem Fall tragen die Aktiengesellschaften zwar die Bezeichnung des hierarchisch übergeordneten Unternehmens mit im Namen, veröffentlichen aber ihre Berichte unter einem eigenen CIK⁵⁶.

Auf der Website der SEC wird angegeben, wie Unternehmen oder Privatpersonen einen

⁵⁵Beispiel Lombard Odier Bank <http://www.sec.gov/cgi-bin/browse-edgar?company=Lombard+odier&owner=exclude&action=getcompany> Abruf: 30.09.2014

⁵⁶Gute Beispiele hierfür sind Automobilunternehmen mit diversen Untermarken - hier: Ford <http://www.sec.gov/cgi-bin/browse-edgar?company=Ford&owner=exclude&action=getcompany> Abruf: 30.09.2014

individuellen CIK erhalten können. Der Bewerbungsprozess läuft über eine gewisse Form⁵⁷ ab, welche ausgefüllt und der SEC geschickt werden muss – in dieser ist lediglich angegeben, dass ein filer “neu“ in der SEC Datenbank sein muss. Wann genau dieser als “neu“ gilt, wird nicht erwähnt⁵⁸. Auch sind die Kriterien, welche über Ablehnung oder Annahme der Bewerbung auf einen CIK entscheiden, nicht genannt, sodass der gesamte Prozess intransparent wirkt und dementsprechend nur schwer nachzuvollziehen ist.

Ein weiteres, aber dennoch ähnliches Problem wurde bereits in Kapitel 5.1 in Bezug auf die Standard Industrial Classification (SIC) zur Identifizierung unterschiedlicher Unternehmensbranche beschrieben: Die SIC ist ein ziemlich veraltetes Modell, welches bis auf wenige Ausnahmen kaum mehr Anwendung findet. Die SEC ist von diesen wenigen noch mit die wichtigste Institution, welche diese Klassifizierung verwendet. Im europäischen Raum werden Branchenklassifizierungen durch die ISIC getätigt, welche ein Konstrukt der UNO ist. Auch wenn bei genauer Betrachtung diverse Parallelen beider Vorgehensweisen auftreten, so sollte doch eine einheitliche Darstellung, bzw. das mapping dieser beiden Ansätze in eine global verwendete Darstellung das langfristige Ziel sein.

Die Arbeit an der Problematik einer fehlenden Vergleichbarkeit über Bundes- und Landesgrenzen hinaus, hat sich der “six degrees”⁵⁹ der Sunlight Foundation⁶⁰ auf die Segel geschrieben. Dieser beschreibt auf seiner Internetpräsenz zwar erst die Problematik, gibt jedoch auch anhand des Beispiels “Singapur“ ein Land an, welches dieses Problem nicht nur in Angriff genommen hat, sondern es auch überwinden konnte⁶¹. In diesem Zusammenhang wird auch das Konzept der CIKs, welche einen zentralen Teil dieser Arbeit darstellen, in die Gesamtbetrachtung mit einbezogen, indem unterstrichen wird, wie nötig eine Vereinheitlichung von identifizieren ist⁶².

7.2 Veröffentlichung eigener XBRL-Konzepte / Nicht-einheitliche Veröffentlichungen

So schön die Möglichkeit eigene Konzepte zu entwerfen (und im Endeffekt dann zu publizieren) für Unternehmen ist, so enorm erschwert dies die Einführung von automatischen Abläufen. Dort, wo keine genauen Einschränkungen – z.B. welcher Wert unter welchem Konzept veröffentlicht werden muss – gegeben sind, ist eine 100%ige Automatisierung enorm schwierig, wenn nicht sogar unmöglich.

⁵⁷<https://www.filermanagement.edgarfiling.sec.gov/servlet/FilerMgmtServlet> Abruf: 10.10.2014

⁵⁸<https://www.filermanagement.edgarfiling.sec.gov/servlet/FilerMgmtServlet> Abruf: 10.10.2014

⁵⁹<http://sunlightfoundation.com/sixdegrees/> Abruf: 10.10.2014

⁶⁰<http://sunlightfoundation.com/> Abruf: 10.10.2014

⁶¹<http://sunlightfoundation.com/sixdegrees/about/> Abruf: 10.10.2014

⁶²<http://sunlightfoundation.com/sixdegrees/existing-identifiers/> Abruf: 10.10.2014

Die Möglichkeiten im Zusammenhang eigens erstellter Finanzkonzepte in den Berichten, wurde bereits in Kapitel 2.1.2 betrachtet. Entscheidend bei diesen Konzepten ist die Betrachtung der Syntax: Eigene Konzepte werden durch den sogenannten “tag“ des jeweiligen Unternehmens gekennzeichnet (siehe Abbildung 6) – über diesen könnte also die Möglichkeit bestehen, die unternehmenseigenen Konzepte gesammelt ausgeben zu lassen. Auch wenn durch die zusätzlich publizierte Taxonomie die Zusammensetzung der eigenen publizierten Daten der Unternehmen in der Regel erläutert wird, so muss deren Einordnung in den Gesamtzusammenhang der Berichte jedoch trotzdem weitestgehend manuell stattfinden.

In diesem Kontext hat sich im Laufe der Arbeit eine Auffälligkeit ergeben: Die tags der Konzepte sowie die tags, die in den Daten der Yahoo! Finance Datenbank als eindeutiges Identifikationsmerkmal galten, stimmten immer⁶³ überein. Vielleicht befindet sich in dieser Richtung ein möglicher Ansatz zum Umgang mit individuellen Konzepten.

Ein ähnliches Problem ist die mangelnde Vereinheitlichung von publizierten Daten. Diese unterschiedlichen Konzepte könnten beispielsweise mit einer großen Anzahl an durch OPTIONAL tags geschachtelte Teil betrachtet werden, was jedoch ebenfalls Probleme bereitet: Erstens müsste jeder Fall das erste Mal eingetragen werden (Lernprozess) und zweitens würde die Performance aufgrund extrem vieler, unnötige OPTIONAL tags extrem leiden. Die Präzisierung der Vorgaben zur Veröffentlichung scheint also hinsichtlich der Möglichkeiten zur Automatisierbarkeit unumgänglich zu sein.

7.3 Präzision der Daten vs. Automatisierbarkeit

Wie bereits in vielen Evaluationen der Ergebnisse der Abfragen des Anwendungsszenarios angedeutet, ist ein erheblicher Punkt, welcher weitere Betrachtung erfordert: Der tradeoff zwischen dem Grad der Automatisierung und der Präzision der Daten. In dieser Austauschbeziehung leidet automatisch einer der beiden Punkte, sobald auf den anderen der verstärkte Fokus gelegt wird.

Auch wenn weitere Arbeit in diesem Punkt nicht direkt als Forschung gewertet werden kann, so ist dieser doch hier aufzulisten. Nötig wäre eine ausführliche Analyse beider Vorgehensweisen, die vor allem auch die jeweiligen Vor- und Nachteile beinhaltet. Ferner ist festzustellen, ob unter bestimmten Voraussetzungen oder für bestimmte Ziele, der eine bzw. andere Ansatz zielführender ist, damit nicht für jeden Punkt, den man in Zukunft im Kontext der Integration von Daten verschiedener Quellsysteme angeht, eine eigene Abwägung stattfinden muss.

Ein möglicher zu verfolgender Ansatz könnte hier die Bestimmung des Grades der Abweichung sein. In diesem Falle müsste eine Kennziffer eingeführt werden, die einen (zur

⁶³Betrachtet wurden 5 verschiedene Unternehmen bzw. deren tags.

Gewährleistung der Vergleichbarkeit bestenfalls prozentualen) Wert der Abweichung zwischen aktuellen, manuell ermittelten Daten und automatisch abfragbaren angibt. Für diesen Wert der Abweichung könnte beispielsweise eine Konstante k gewählt werden, ab welcher manuelle Abfragen stattfinden sollten. Für alle Werte bis k wäre die Abweichung noch “angemessen“, zumindest verglichen zum sonstigen Mehraufwand.

7.4 Diskrepanz zwischen verschiedenen Publikationsformen der Finanzberichte

Ein Teil dieser Arbeit beinhaltet auch den Vergleich, zwischen den Daten, die in Form von XBRL Dokumenten veröffentlicht werden, und solchen, die als Jahresbericht in Textform bereitgestellt werden. Kapitel 2.1 illustriert sowohl die Einführung als auch Vorteile von XBRL. Aus diesen geht auch hervor, dass eigentlich eine inhaltliche Äquivalenz zwischen beiden Publikationsformen bestehen sollte.

Nach genauer Betrachtung ist dies jedoch nicht der Fall⁶⁴. Dieser Fakt mindert erstens die allgemeine Vergleichbarkeit der Dokumente und erschwert an zweiter Stelle die Arbeit bezüglich der Auswertung. Zusammen mit Punkt 3 ist eine automatische Auswertung der Dokumente auf Basis der schriftlichen Versionen der Finanzberichte natürlich nur dann möglich, wenn alle relevanten Fakten dieser Dokumente auch in XBRL erreichbar sind.

An dieser Stelle sind zukünftig nun mehrere Aspekte zu betrachten. Es gilt herauszufinden, welche Daten nicht in XBRL Dokumente übernommen werden und warum dies nicht geschieht. Ferner müssen Überlegungen angestellt werden, ob die Verpflichtung zu deren Angabe Sinn macht. Während der Bearbeitung hat sich gezeigt, dass vor allem solche Zahlen, die im Fließtext der Berichte auftauchen, oftmals nicht in XBRL Dokumente übernommen werden⁶⁵ – vielleicht kann auch hier eine einheitliche Regelung gefunden werden. Zuletzt müsste geprüft werden, ob die “fehlenden“ Zahlen aus den bereits gegebenen Informationen geschlossen werden können. Bei einer Zahl wie den Niederlassungen ist dies Gewiss nicht möglich, bei Angaben, die sich aus mehreren Werten zusammensetzen vielleicht schon.

⁶⁴Für ein Beispiel können unter anderem die Jahresberichte der Fifth Third Bancorp des Jahres 2009 nach den Werten der “Home Equity“ durchsucht werden <http://www.sec.gov/Archives/edgar/data/35527/000119312510042347/0001193125-10-042347-index.htm> Abruf: 01.10.2014

⁶⁵Siehe Anzahl der Niederlassungen (Beispiel: Fifth Third Bancorp Jahresberichte 2009–2011)<http://www.sec.gov/cgi-bin/browse-edgar?action=getcompany&CIK=0000035527&type=10-K&dateb=2012-06-06&owner=exclude&count=10> Abruf: 01.10.2014

7.5 Visualisierbarkeit

Hinsichtlich der Benutzerfreundlichkeit ist die Umsetzung zur Handhabung der Abfragen aus Kapitel 5 durch ein Graphical User Interface (GUI) unabdingbar. Abfragen in SPARQL Syntax sind für Laien dieser Thematik nicht so handlich, wie ein angemessenes GUI. Hierbei muss darauf geachtet werden, dass alle Aspekte bezüglich der individuellen Konfiguration einzelner Abfragen, beispielsweise durch Drop-Down Menüs, integriert werden. Logischerweise wäre dies nach Abarbeitung der zuvor genannten, die generelle Logik betreffenden Ansätze auch der nächste Schritt, wenn die Vermarktung einer Software mit dem beschriebenen Vorgehen zur Analyse von Finanzdaten in Planung wäre.

8 Diskussion und Evaluation

Dieser Teil der Arbeit ist in zwei Abschnitte aufgeteilt. Im ersten Abschnitt findet eine Diskussion statt, in der aus Autorensicht Vor- und Nachteile der Vorgehensweise, wie sie im Anwendungsfall in Kapitel 5 präsentiert wurde, beleuchtet werden. Der zweite Teil beschäftigt sich dann mit der Performance der Ausführungen, welche durch eine Befragung und die Betrachtung der Laufzeiten der Abfragen beleuchtet und evaluiert wird.

8.1 Diskussion

Zuerst findet an dieser Stelle die Diskussion über Durchführung und Ergebnisse des Anwendungsszenarios statt. Hierbei werden sowohl die Komplexität, die generelle Durchführbarkeit aber auch die Probleme und mögliche Verbesserungen, die sich bei der Durchführung ergeben haben, eine große Rolle spielen.

Bei der Suche nach einer “passenden“ Unternehmensbranche mussten glücklicherweise noch nicht allzu viele Dinge penibel beachtet werden. Aufgrund der Masse an möglichen Branchen war es möglich, eine für dieses Szenario angemessene zu finden. Ferner gibt es nicht “die richtige“ oder “die falsche“ Branche – musste doch nach der Wahl lediglich bei den Kennzahlen auf die Anwendungsmöglichkeiten bzw. Eigenschaften der gewählten Branche geachtet werden. Die Bankenbranche war, nicht zuletzt auch vor dem finanziellen Hintergrund der Gesamtarbeit, naheliegend und wurde somit für den Anwendungsteil der Arbeit als sinnvoll erachtet.

Da bei der Wahl der Kennzahlen zeitgleich mehrere Dinge beachtet werden mussten, war das Treffen dieser Auswahl bereits problematischer. Hierbei galt es neben speziellen Eigenschaften der Bankenbranche noch weitere Voraussetzungen zu beachten: Es sollten vor allem die technischen Möglichkeiten, aber auch deren Grenzen aufgezeigt werden. Ferner war es wichtig, möglichst verschiedene Kennzahlen zu finden, um so viele unterschiedliche Aspekte betrachten zu können wie möglich. Nicht zuletzt mussten Kennzahlen gewählt werden, die durch angegebene Daten überhaupt dargestellt werden konnten!

Auch wenn der letzte Punkt trivial klingt, stellte er doch im Laufe der Bearbeitung des Anwendungsfalls das größte Problem dar. Dies hat hauptsächlich den Grund, dass die textuelle Version der Finanzberichte sich von den XBRL Publikationen unterscheidet, wie es bereits in Kapitel 7 beschrieben wird. Nach dem Fund einer angemessenen Kennzahl wurde in dieser Arbeit als erstes betrachtet, ob die – für Menschen deutlich angenehmer zu lesenden – Finanzberichte im Textformat diese Kennzahlen enthielten. Als es dann an die Arbeit der Auswertungen mithilfe der Abfragen ging, waren die benötigten Daten stellenweise zum Teil oder gänzlich nicht verfügbar. An dieser Stelle muss also die Frage gestellt werden, ob die Vorgehensweise der Findung von Kennzahlen angemessen war.

Auch kann nicht gewährleistet werden, dass an dieser Stelle die “besten“⁶⁶ Kennziffern ausgewählt wurden. Die beispielhafte Extraktion **aller** Finanzkonzepte aus einem Bericht (mit anschließender Bestimmung der Kennzahlen auf diesen Daten als Grundlage) wäre zwar deutlich aufwändiger, jedoch möglicherweise auch zielführender gewesen.

Trotz der durchaus ausbaufähigen Auswahl der Kennzahlen decken diese die verschiedenen technischen Möglichkeiten des Ansatzes angemessen ab. Die erste Kennzahl gibt zeigt bereits als einleitendes Beispiel, wie die Entwicklung einer beliebig zu wählenden Variable betrachtet werden kann. Durch die Einführung einer eigenen Klasse (vgl. Abbildung 21), ist auch die Implementierung von Kennzahlen, die sich auf neue Versionen der US-GAAP beziehen, problemlos möglich.

Der Grundsatz der Data Envelopment Analyse (DEA) erscheint im Zusammenhang der Messung der Performance von Unternehmen sinnvoll und ist durch diverse Paper belegt. Da dieses Konstrukt die Grundlage eines großen Teils des Anwendungsfalls (also somit das Kernstück des Hauptteils der gesamten Arbeit) bildet, ist auch eine thematische Einführung gegeben.

Die Kennzahlen 2 bis 5, welche in den Bereich der DEA fallen, stellen die Möglichkeiten der Abfragen über integrierte Daten mehrerer Quellen am deutlichsten dar. Die Aussicht, unternehmensinterne Werte verschiedenster Kennzahlen mit Werten aus verschiedenen Quellen, nicht nur zu ermitteln, sondern auch im selben Schritt mit denen anderer Unternehmen innerhalb einer einzelnen Abfrage vergleichen zu können, ist nicht nur für Unternehmer, sondern auch für Analysten überaus attraktiv. Jedoch wird Kapitel 8.2.3 zeigen, dass bei komplexeren queries die Performance sinkt. Besonders signifikant ist dies der Fall, wenn rules (vgl. Abbildung 20) mit eingebunden werden, oder GETs⁶⁷ auf URIs von Datenquellen außerhalb des RDF Store durchgeführt werden, um die mit diesen Ressourcen verknüpften Daten einbeziehen zu können (vgl. zum Beispiel Abbildung 31).

Die durch Kennzahl 6 repräsentierte Vorgehensweise zur Aufsummierung von Daten aus mehreren Perioden ist sowohl für die externe, als auch für die interne Umgebung des Unternehmens interessant. Extern kann durch die Wahl verschiedener Faktoren eine langfristige Profitabilität nachgewiesen (oder verworfen) werden. Intern ist dies ein angemessenes Instrument, um die strategische Ausrichtung des Unternehmens zu überdenken oder gänzlich neu auszurichten. Da die Abfrage relativ simpel gestaltet ist, sind die Möglichkeiten zur Abfrage verschiedener Werte flexibel und leicht umsetzbar. Nichtsdestotrotz birgt diese Kennzahl jedoch auch gewisse Probleme in der Umsetzung. Eine Abfrage, die beispielsweise nur Werte aus Jahresberichten zurückgibt, ist ohne weiteres nicht möglich. Stattdessen muss für jedes Jahr über die properties ical:dtstart bzw. ical:dtend der genaue Zeitraum definiert werden, um diesen anschließend aufsummieren zu können. Bei der

⁶⁶Auf die Problematik der Bewertung von Kennziffern als “gut“ oder “schlecht“ wurde in dieser Arbeit bereits eingegangen.

⁶⁷http://www.w3schools.com/tags/ref_httpmethods.asp Abruf: 05.10.2014

Prüfung der Summe eines Wertes von Quartalsberichten ist analog vorzugehen. Aus genau diesem Grund ist es wichtig, die financial facts, welche man erfragen will, zu kennen. Eine nachträgliche Umstellung von Jahres- auf Quartalsberichte ist ohne Abänderung der SPARQL-Abfrage nicht möglich. Diese Problematik wird in Abbildung 39 durch die Pünktchen symbolisiert. Die präzise Repräsentation von längeren Verläufen ist also in diesem Fall nur schwierig vollautomatisch durchzuführen – um das gewünschte Ergebnis zu erhalten, muss also bereits vorher der komplette Zeitraum, sowie die Herkunft der Werte (Jahres- oder Quartalszahlen) besprochen werden. Dieses Problem basiert ebenfalls auf dem in Kapitel 7 angesprochenen Ansatz von “Präzision der Daten vs. Automatisierbarkeit”.

Die Kennzahlen 8 und 9 sind eher unterstützende als zentrale Punkte und sollen die technischen Möglichkeiten aufzeigen. Wie auch im Hauptteil selbst erläutert, bietet Kennzahl 8 über Daten von DBpedia die Möglichkeit, eine – wenn auch sehr oberflächliche – Strukturanalyse des Gesamtunternehmens zu geben. Da die Standorte von Tochtergesellschaften gegeben werden, ist zusätzlich zu den reinen finanziellen Aspekten in dieser Betrachtung ebenfalls die Möglichkeit gegeben, auf geographische Besonderheiten eingehen bzw. reagieren zu können. Kennzahl 9 stellt den Aktienkurs und seine Entwicklung da. Trotz der Integration einer weiteren Datenquelle (Yahoo! Finance) muss die Frage gestellt werden, ob diese nicht auch auf andere Art und Weise hätte gezeigt werden können, als mit einer Kennzahl, die so bereits schon oft verwendet wurde und keine Neuerung darstellt.

Kennzahl 7 beinhaltet mit dem Vergleich der Akkuratessse von Werten zur selben Variable (beispielsweise Aktiva oder Eigenkapital) aus verschiedenen Quellen die Möglichkeit für Analysten, ihre Quellen zu validieren. Im Zusammenhang des Anwendungsfalls ist dies ebenfalls eine unterstützende Abfrage, die ohne den Kontext der anderen Kennzahlen keine große Aussagekraft besitzt. Die Ergänzung scheint jedoch vor allem bei der Benutzung mehrerer verschiedener Quellen mit (angeblich) gleichen Angaben sinnvoll und unterstreicht die Glaubwürdigkeit der präsentierten Werte. In Kapitel 6 werden zu den Ausführungen dieses Ansatzes zwei Vorgehensweisen aufgezeigt und erläutert, die ebenfalls die bereits angesprochene Problematik “Präzision der Daten vs. Automatisierbarkeit” betonen und diese zu einem wichtigen weiteren Forschungsansatz machen.

Abschließend bleibt festzuhalten, dass der Anwendungsfall generell sowohl technische Möglichkeiten als auch Probleme und Grenzen des Ansatzes darstellen kann. Dem gegenüber steht das eingangs dieses Kapitels beschriebene Problem der Kennzahlenauswahl. Auch wenn richtigerweise betont wird, dass die Wahl stets auf subjektiven Grundlagen getätigt wird, besteht hier Verbesserungspotential. Ebenfalls wäre wünschenswert, dass ein anderes Vorgehen zur Verbindung von Datenquellen, als das manuelle Einfügen von “owl:sameAs” Verknüpfungen (siehe Abbildung 25), bedacht werden könnte. Weitere wichtige Ansätze zur Forschung und Verbesserung werden in Kapitel 7 beschrieben.

8.2 Evaluation des Nutzens und Analyse der Performance

Dieser Abschnitt des Diskussionsteils beschäftigt sich mit der Evaluation der durchgeführten Umfrage, sowie mit der Performance verschiedener Abfragen, wie sie in Kapitel 6 beschrieben wurden.

8.2.1 Beschreibung der Grundgesamtheit der Umfrage

Bevor die Umfrage dieses Kapitels präsentiert wird, gilt es die Eigenschaften der Grundgesamtheit, auf der die Umfrage durchgeführt wurde, zu beschreiben. Um sowohl den wirtschaftlichen als auch den technologischen Aspekten des Fragebogens (siehe A.5) gerecht werden zu können, werden also Teilnehmer benötigt, die nicht nur über technisches Know-How, sondern auch über eine Grundausbildung der Betriebswirtschaftslehre verfügen. Diese beiden Disziplinen vereinen vor allem Studierende der Informationswirtschaft ([GST03]), welche die Grundgesamtheit dieser Umfrage darstellen. Aus dieser wurden zufällig 15 Teilnehmer im Alter von 20 bis 25 (vgl. Anhang A.6) ausgewählt, die einen prozentualen Anteil von ca. 2,85% aller Informationswirte des KIT ausmachen⁶⁸. Zusätzlich wurde hier im Besonderen darauf geachtet, dass alle Teilnehmer der Befragung mindestens das zweite Semester des Bachelorstudiums abgeschlossen hatten, um wenigstens die notwendigsten Grundkenntnisse der BWL und Informatik gewährleisten zu können.

8.2.2 Methodologie

Den Teilnehmern wurde der Hauptteil dieser Arbeit unter neutralen Gesichtspunkten vorgestellt. Hierbei wurde gleichsam auf Vorteile, aber auch auf potenzielle Verbesserungen ausführlich eingegangen. Die Vorstellung der Arbeitsschritte nahm je nach Person ca. 15 bis 20 Minuten Zeit in Anspruch. Anschließend wurde den Teilnehmern der ausgedruckte Fragebogen (vgl. A.5) zum Ausfüllen vorgelegt, wozu diesen ein Zeitraum von zehn Minuten gegeben wurde.

Das Design des Fragebogens wurde auf Basis der Funde von Weijters et al. aus dem Jahre 2010 ([WCS10]) sowie Weathers et al. von 2005 ([WSN05]) erstellt. Diese berichten übereinstimmend, dass hinsichtlich hoher Validität und möglichst geringem Verhaltensbias, Fragebögen mit 5 Antwortmöglichkeiten, welche alle beschriftet sind, für die besten Ergebnisse sorgen. Des Weiteren ist zu erwähnen, dass im Kernteil des Fragebogens 3 sogenannte “reversed items“-Paare ([SWN08]) eingefügt wurden, um die Qualität der Ergebnisse mit diesem Ansatz prüfen zu können.

⁶⁸Stand: Sommersemester 2014. http://www.kit.edu/downloads/Statistik_SS2014.pdf Abruf: 21.10.2014

In der Umfrage selber wurden alle Teilnehmer zuerst gebeten, ihr Alter, Geschlecht sowie eine Einschätzung bezüglich der eigenen Fähigkeiten in den Bereichen Informatik und BWL generell, sowie Semantic Web und Finanzwirtschaft & Rechnungswesen im Speziellen zu geben. Anschließend wurden neun Aussagen angegeben, denen gegenüber die Probanden ihre Meinung entweder als zustimmend, ablehnend oder neutral geben sollten.

Während der zehn Minuten, in denen die Probanden den Fragebogen ausfüllen sollten, wurden nochmals die Überschriften aller neun Kennzahlen bzw. deren Vorgehensweisen eingeblendet, um die Fragen bezüglich der Vollständigkeit beantworten zu können. Zusätzlich wurde der Code der Abfrage aus Abbildung 23 eingeblendet, um Einschätzungen zu Aufbau und Komplexität der Abfragen geben zu können.

8.2.3 Evaluation der Umfrageergebnisse

Die Befragung zielte darauf ab, zu evaluieren, ob ein wirklicher Mehrwert des in Kapitel 5 präsentierten Ansatzes für Unternehmen sowie Analysten gesehen wird. Ferner sollte betrachtet werden, in welchem Umfang Vorkenntnisse zum Verständnis der Vorgehensweise nötig sind, ehe eine kurze Aussage über einen Test der Datenqualität angegeben wird.

Bevor an dieser Stelle auf den geschaffenen Mehrwert für Unternehmen und Analysten eingegangen wird, soll zuerst die Notwendigkeit spezieller Vorkenntnisse betrachtet werden. Die Relevanz der Vorkenntnisse wurde in dieser Befragung durch zwei Dinge betrachtet: Erstens wurden die Teilnehmer um eine realistische Einschätzung ihrer Kenntnisse gebeten, um zweitens die jeweiligen Antworten bezüglich der Komplexität der Abfragen mit diesen Einschätzungen zu vergleichen. Erfreulicherweise konnte das Ziel von generellen Kenntnissen in den Bereichen BWL und Informatik erzielt werden, sodass im Bereich Informatik eine durchschnittliche, im Bereich der BWL sogar eine überdurchschnittliche Kenntnis zu Grunde gelegt werden kann (vgl. Anhang A.6).

Deutlich interessanter sind allerdings die genaueren Auswertungen zu den Vorkenntnissen der Thematik Semantic Web sowie im Bereich Finanzwirtschaft und Rechnungswesen. Hierbei wiesen die meisten Teilnehmer durchschnittlich zwar auch im Bereich Finanzwirtschaft & Rechnungswesen durchschnittliche bis überdurchschnittliche Kenntnisse auf, im Bereich des Semantic Web waren die Kenntnisse jedoch deutlich breiter gestreut und eine Tendenz in Richtung wenigen Vorkenntnissen war zu erkennen:

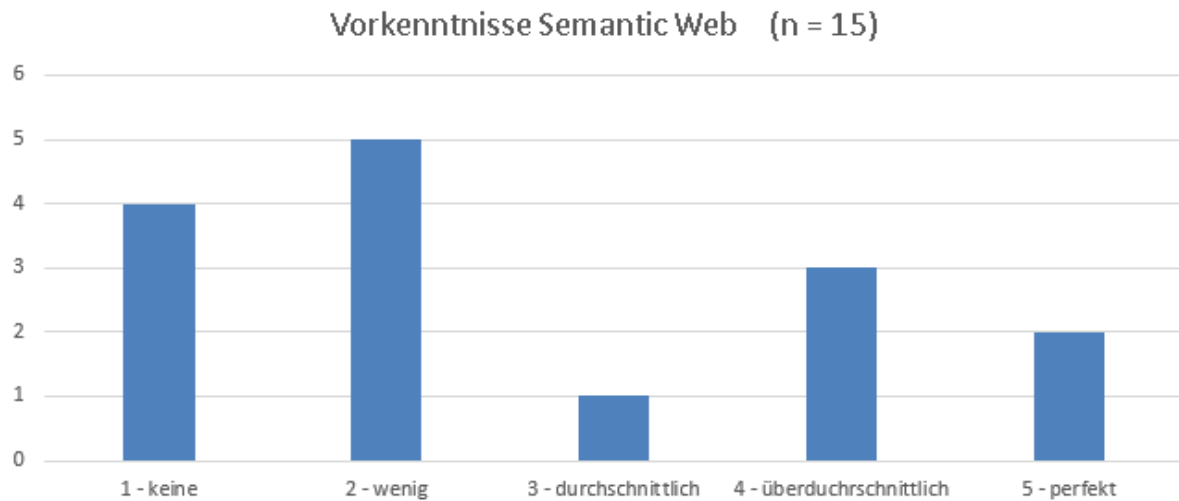


Abbildung 48: Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich Semantic Web

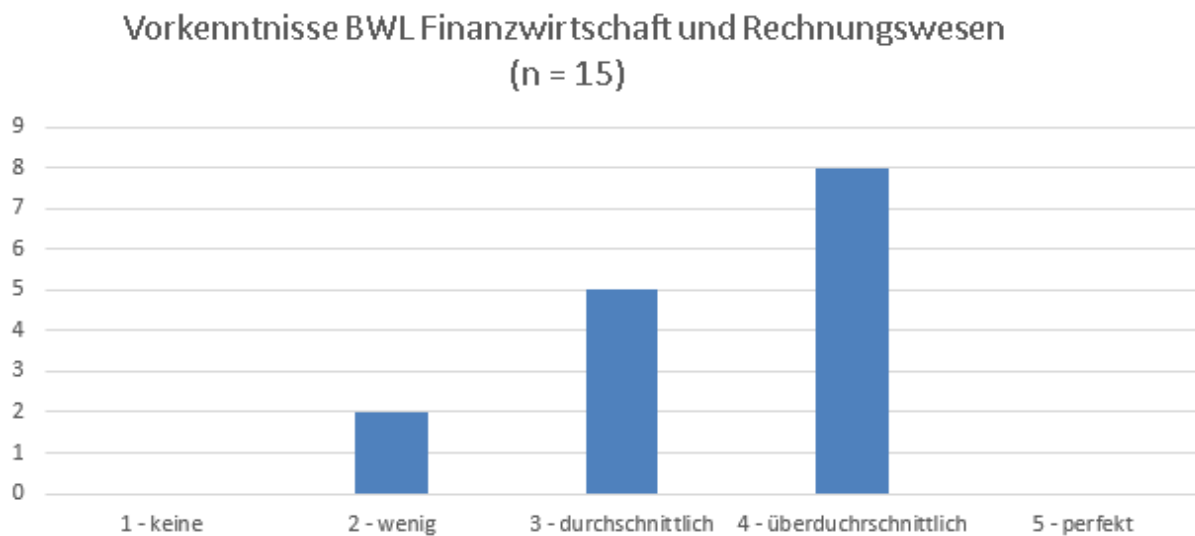


Abbildung 49: Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich BWL Finanzwirtschaft & Rechnungswesen

In diesem Zusammenhang ist nun interessant zu betrachten, als wie anspruchsvoll die Teilnehmer die generelle Vorgehensweise der Abfragen erachtet haben. Hierbei werden zwei Grafiken angegeben, da es sich bei den beiden Fragen dieses Thematik um eines der erwähnten “reversed items“-Paare handelt:

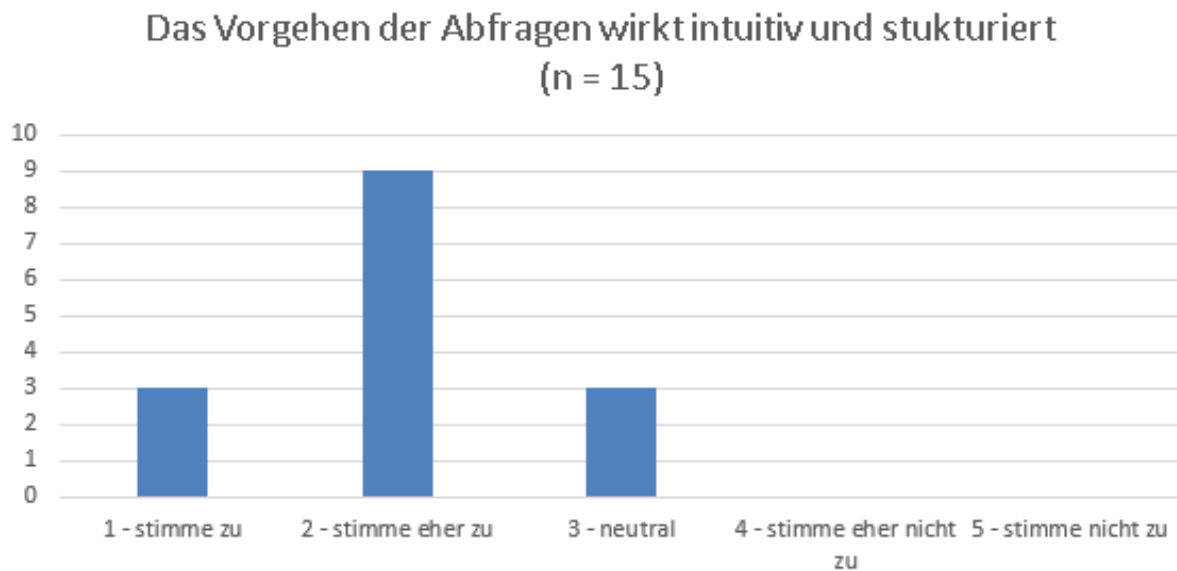


Abbildung 50: Auswertung der Evaluationsergebnisse: Angaben zur Komplexität bzw. dem Verständnis bzgl. der Abfragen (1)

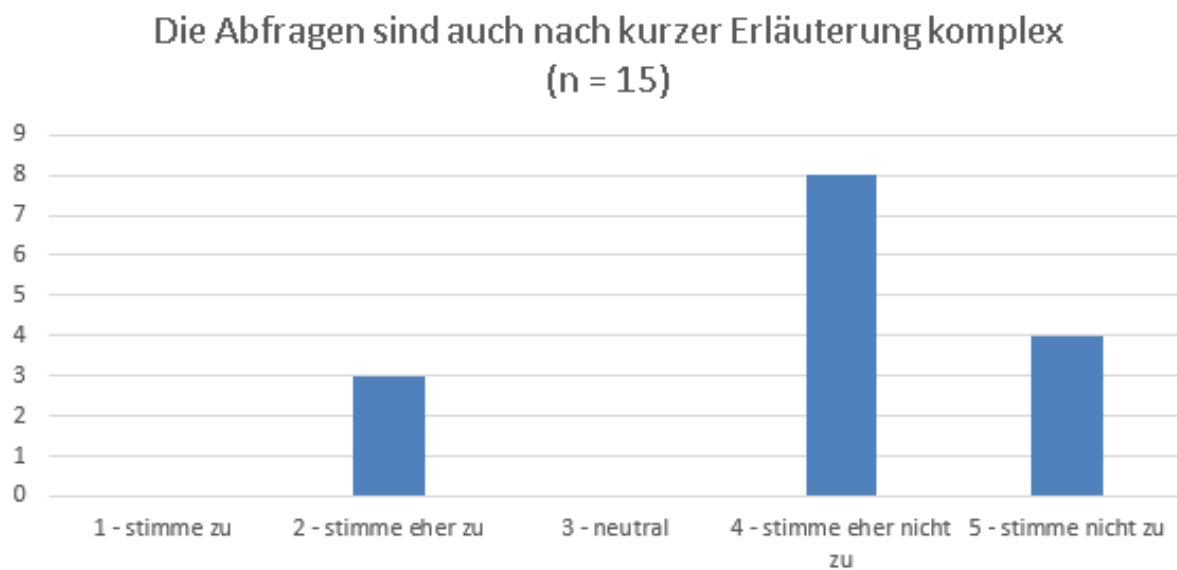


Abbildung 51: Auswertung der Evaluationsergebnisse: Angaben zur Komplexität bzw. dem Verständnis bzgl. der Abfragen (2)

Setzt man diese Ergebnisse nun in Zusammenhang mit den zuvor präsentierten Ergebnissen der Einschätzung der eigenen Vorkenntnisse, vor allem im Bereich des Semantic Web, bleibt festzuhalten, dass ein grundlegendes Verständnis für die Abfragen bei 75%

aller Beteiligten der Fall war. Dies erbringt natürlich keinen endgültigen “Beweis“, dass keinerlei Vorkenntnisse für die Anwendung dieses Ansatzes nötig sind, gibt jedoch die durchaus erfreuliche Tendenz, dass auch Laien im Bereich des Semantic Web die durchgeführten Schritte nachvollziehen können. Nichtsdestotrotz ist grundsätzlich anzumerken, dass weitere Kenntnisse nicht hinderlich sind, wenn das Vorgehen im Anwendungsfall auf eigene, individuelle Bedürfnisse angepasst werden soll.

Der zweite Teil, welcher in der Auswertung der Evaluationsergebnisse besondere Beachtung finden soll, ist die allgemeine Einschätzung der Relevanz für Unternehmen und Analysten, die durch den Eindruck des geschaffenen Mehrwertes beeinflusst wird. Eben diese Einschätzung wurde von den Probanden durch die übrigen sieben Fragen des Fragebogens eingeholt.

Innerhalb dieser sieben Fragen befanden sich auch “reversed items“-Paare. Eines von diesen beschäftigte sich mit der Frage, ob die Teilnehmer den Mehrwert für Unternehmen erkennen konnten – die Ergebnisse sind eindeutig:

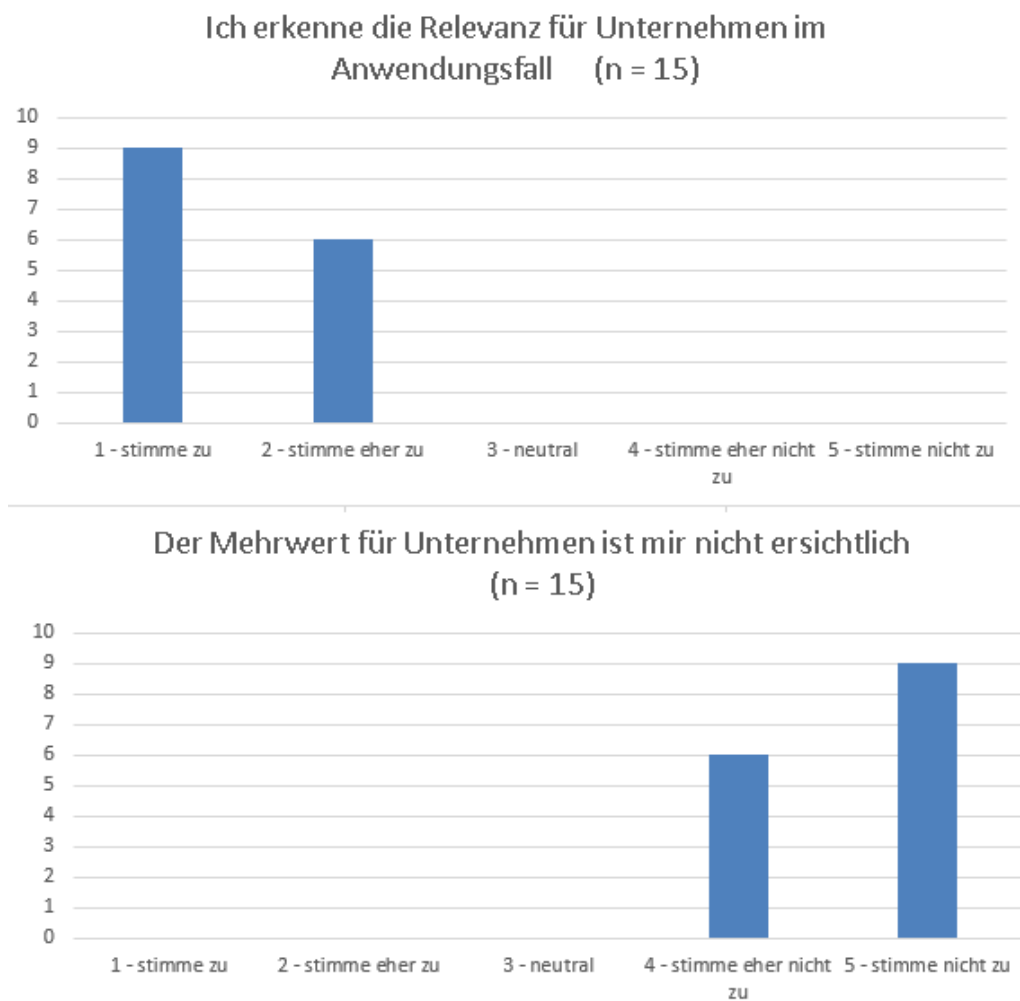


Abbildung 52: Auswertung der Evaluationsergebnisse: Angaben zur Relevanz und dem geschätzten Mehrwert für Unternehmen

Nach diesen Angaben erkennen 100% aller Befragten die Relevanz des Ansatzes zumindest zum größten Teil oder sogar gänzlich. Ferner waren alle Beteiligten sich einig, dass durch den beschriebenen Ansatz – vor allem verglichen zu den Analogen Möglichkeiten mit Stift und Papier – ein erkennbarer Mehrwert für Unternehmen geschaffen werden kann. Auch MR Fehler (die Definition befindet sich weiter unten) wurden in diesem Zusammenhang nicht begangen.

Da der Individualisierungsgrad der Abfragen als immens wichtig angesehen wurde, hatte das zweite “reversed items“-Paar diesen als Thema. Die Meinung der Probanden war auch in diesem Kontext eindeutig:

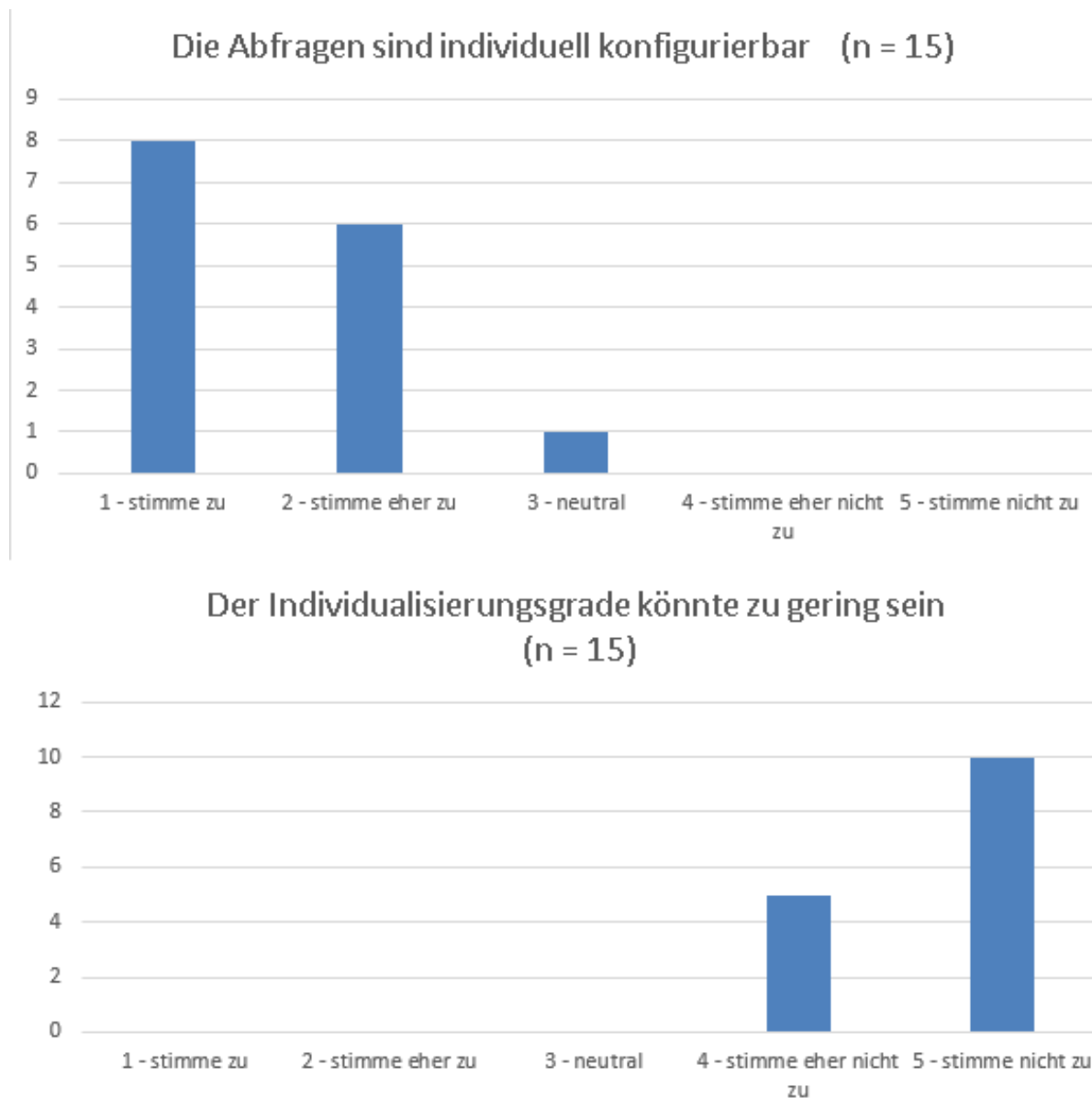


Abbildung 53: Auswertung der Evaluationsergebnisse: Angaben zur Individualisierbarkeit der Abfragen

Die starke Gewichtung dieses Aspekts hatte den Grund, dass auch generelle Vorgehen erst dann wertvoll sind, wenn sie unter verschiedenen, praktischen Umständen bestehen können und den individuellen Anforderungen unterschiedlicher Unternehmen gerecht wurden.

Die Weiteren Fragen beschäftigten sich beispielsweise mit Einschätzungen zum Thema Vollständigkeit der Abfragen oder auch der Präsentation und Darstellbarkeit der Ergebnisse. Am schlechtesten ausgefallen ist die Beurteilung der Frage, ob sich die Teilnehmer vorstellen könnten, dass Unternehmen den präsentierten Ansatz benutzen würden – wobei schlecht an dieser Stelle keine Ablehnung, sondern lediglich den geringsten Zuspruch bedeutet. Die Ergebnisse der in diesem Kapitel nicht präsentierten Fragen finden sich in Anhang A.6.

Ferner wurde die Qualität der Daten hinsichtlich MR betrachtet. Von einem MR Fehler wird an dieser Stelle gesprochen, wenn in einem der drei “reversed items“-Paare (1–8, 2–4, 6–9) Angaben im inhaltlichen Kontrast stehen. Dies ist explizit genau dann der Fall, wenn im ersten Teil des Paares Möglichkeit 1 oder 2 [4 oder 5] angekreuzt wurde, im zweiten Teil des Paares jedoch **nicht** die Antwortmöglichkeit 3, 4 oder 5 [1, 2 oder 3] ausgewählt wurde, logischerweise gilt dies auch genau umgekehrt. Somit kann pro “reversed items“-Paar genau ein MR Fehler auftreten. Daraus folgt, dass pro Fragebogen somit maximal drei MR Fehler auftreten konnten, was bei einer Teilnehmerzahl von 15 Probanden eine maximale Fehlerzahl von 45 ergibt. Von diesen 45 möglichen wurde genau 1 MR Fehler begangen, was zu dem niedrigen MR Level von 2,22% führt.

8.2.4 Performance verschiedener Abfragen

Das Hinzuziehen von rules zum Gewährleisten der allgemeinen Funktionalität macht die Durchführung eines Tests hinsichtlich der Performance der verschiedenen Abfragen unerlässlich. Wie die folgende Tabelle zeigen wird, steigt nämlich mit steigender Komplexität, wozu unter anderem auch das Einsetzen von rules zählt, auch die Abfragezeit:

AbfrageNr.	Datenquellen	mit rules	ohne rules	AbbildungNr.
1	SEC	60s	52s	23
2.1	SEC + DBpedia	70s	60s	25
2.2	SEC + DBpedia	581s	536s	26
3.1	SEC + DBpedia	69s	63s	30
3.2	SEC + DBpedia	523s	516s	31
4	SEC + DBpedia	69s	67s	32
5	SEC	64s	56s	35

AbfrageNr.	Datenquellen	mit rules	ohne rules	AbbildungNr.
6	SEC	71s	69s	38
7	SEC + DBpedia	70s	68s	40
8	SEC + DBpedia	68s	66s	42
9.1	SEC + Yahoo!	k.A.	67s	45
9.2	SEC + Yahoo!	k.A.	68s	47

Positiv überraschend ist an dieser Stelle der geringe Performanceverlust durch die Anwendung der “subClassOf“ rule (vgl. Abbildung 21). Am stärksten ist der Verlust der Performance, wenn direkt mehrere Werte gleichzeitig aus den SEC Dokumenten ermittelt werden müssen, wie es beispielsweise bei den Kennzahlen 2 und 4 der Fall ist. Ebenfalls erfreulich ist die geringe Diskrepanz bezüglich der Performance zwischen Abfragen, die lediglich Daten der SEC beziehen und solchen, die Daten aus mehreren Quellen integrieren. Alles in allem kann man also die Performance hinsichtlich diesen Aspekts als gut bezeichnen.

Die hohen Abfragezeiten in den Abfragen 2.2 sowie 3.2 können dadurch begründet werden, dass nicht mehr nur ein GET auf die DBpedia Seite eines festgelegten Unternehmens mit zuvor angegebenem CIK durchgeführt wird. Genauer werden Informationen **aller** im datastore angegebenen Unternehmen eingeholt, um den Wert von deren Kennzahlen mit dem ebenfalls in der Abfrage angegebenen Wert zu vergleichen. Positiv ist bei diesen Abfragen jedoch hervorzuheben, dass das Hinzufügen von rules die benötigte Zeit lediglich minimal erhöht.

Da in Abfrage 9.1 und 9.2 generell keine subClassOf rule auftaucht, wurde lediglich eine Abfrage ohne Aufruf der Regeln durchgeführt. Dies begründet den Punkte k.A. (keine Angabe) in der Tabelle.

Alle Abfragen wurden auf einem Server mit 128 Gigabyte Ram Speicher und 16 cores durchgeführt. Um zu gewährleisten, dass die Werte sowohl vergleichbar als auch repräsentativ sind, wurden alle Zahlen ermittelt, als die CPU Auslastung zum Startpunkt der Abfrage bei nahezu 0% stand.

9 Fazit

Zum Abschluss der Arbeit erfolgt ein Fazit, welches in Bezug auf die in Kapitel 1.2 präsentierte Fragestellung sowie die in Kapitel 3 aufgestellte Hypothese, die Ergebnisse der Arbeit präsentiert und beurteilt.

Das Ziel der Arbeit war die Betrachtung eines Ansatzes zur Vereinheitlichung von Finanzdaten verschiedener Ursprungsquellen mit anschließendem Versuch zur weitestgehenden Automatisierung unterschiedlicher Abfragen zu diesen Daten. Neben dem Nachweis der allgemeinen Durchführbarkeit, war es ein weiteres Ziel, den vorgestellten Anwendungsfall so zu entwickeln, dass die betrachtete Problematik nicht nur greifbarer wird, sondern auch in realistischen Szenarien performant eingesetzt werden kann, um nicht ein weiterer, nur in der Theorie existierender Gedanke zu sein.

Aufgrund durchgeführter Abwägungen am Ende von Kapitel 2, findet die Arbeit nach der thematischen Einleitung eine Abbildung von XBRL Datensätzen in RDF. Zuvor werden die Vorteile von RDF anscheinlich dargestellt und erläutert. Vor der Einführung in das Anwendungsszenario wird die sich stark am Titel der Arbeit orientierte Hypothese aufgestellt, welche auch am Ende dieses Kapitels nochmals aufgegriffen wird.

Die Ergebnisse der Durchführung des Anwendungsfalls wurden in Kapitel 8 ausführlich betrachtet. Diese stellen anschaulich dar, dass das Anwendungsszenario vor allem den technischen Aspekt mit seinen Möglichkeiten und Einschränkungen gut beschreibt. Ferner wird bei komplexeren Abfragen ein Verlust von Performance festgestellt, welcher als stellenweise zwar als angemessen angesehen werden kann, bei manchen Abfragen jedoch auch zu nicht akzeptablen Abfragezeiten führt (vgl. Kapitel 8.2.4).

Fraglich ist nun, ob die Hypothese nach den Ausführungen dieser Arbeit angenommen, teilweise angenommen oder verworfen werden sollte.

1. Die generelle Durchführbarkeit wird, wie eingangs erwähnt, genau durch den Anwendungsfall betrachtet. Auch wenn die Abfragen komplexer werden, gibt es nur wenige Informationen, die aus einem integrierten RDF datastore nicht mittels SPARQL-Abfragen erhalten werden können. An dieser Stelle ist wichtig zu erwähnen, dass das größte Hindernis der generellen Durchführbarkeit nicht – wie unter Umständen zu erwarten gewesen wäre – technische bzw. syntaktische Unmöglichkeiten waren, sondern die in Kapitel 7 angesprochenen Problematiken bzgl. der Möglichkeit zum Entwurf eigener Konzepte sowie der inhaltlichen Diskrepanz zwischen verschiedenen Berichten. Die generelle Durchführbarkeit kann also, auch wenn die aufgezeigten Grenzen zweifelsohne bestehen, grundsätzlich bejaht werden.
2. Der zweite Teil, der von der Hypothese betrachtet wurde, ist die Frage nach der Performance der unterschiedlichen Abfragen. Auch wenn der Anwendungsfall natürlich

nicht **alle** verschiedenen Formen und Ausprägungen, wie Abfragen gestaltet sein können, abdecken kann, so wird doch ein angemessen breites Spektrum bezüglich der möglichen Komplexität gegeben. Ähnlich breit wie dieses Spektrum ist auch die in Kapitel 8.2.3 dargestellte, unterschiedliche Performance. Vor allem “triviale“ Abfragen ohne rules oder den Einbezug verschiedener Datenquellen, liefern äußerst schnell die gewünschten Ergebnisse. Auf der anderen Seite wird aber vor allem bei den Abfragen, die für eine DEA relevant und somit auch für Analysten interessanter sind als die Entwicklungsreihe eines einzelnen Wertes, mehr Zeit benötigt. Ebenfalls langsamer gestalten sich die Abfragen zur Konkurrenzanalyse, wie sie durch die Kennzahlen 2 und 3 des Anwendungsfalls präsentiert werden. Auch der Teil der Hypothese, der sich mit der Performance beschäftigt, kann somit zwar nicht verworfen, aber auch nur hinsichtlich mancher Gesichtspunkte bestätigt werden.

3. Zuletzt wurde die Hypothese aufgestellt, dass die präsentierten Vorgehensweisen die Arbeit im Bereich der Finanzanalyse erleichtern können. Wie zu Beginn dieses Kapitels erwähnt, handelt es sich um ein relativ junges Forschungsgebiet. Das durchgeführte Szenario birgt vielleicht keine bahnbrechenden Neuerungen, weiß jedoch vor allem durch einfache Analysemöglichkeiten und viel Freiraum zur individuellen Konfiguration zu überzeugen. An dieser Stelle ist erneut zu betonen, dass die vorgestellten Kennzahlen Beispiele sind, die (fast) beliebig auf eigene Anforderungen angepasst werden können. Einschränkungen werden hauptsächlich durch die Punkte zur weiteren Forschung in Kapitel 7 gegeben. Vor allem die Möglichkeiten zur Analyse der Konkurrenz und deren Kennzahlen, eröffnet trotz schlechterer Performance Möglichkeiten zur neuen strategischen Ausrichtung des Unternehmens oder zur Analyse der Effektivität aller Wettbewerber im Markt. Die Vorgehensweise kann Finanzanalysten also durchaus bei ihrer Arbeit unterstützen. Ehe dies geschehen kann, gilt es jedoch, verschiedene Probleme zu lösen. Auch dieser Teil der Hypothese kann also bedingt angenommen werden.

Als abschließendes Ergebnis dieser Arbeit bleibt festzuhalten, dass – mit Bezug auf die präsentierte Fragestellung – im Ansatz die Durchführbarkeit der Integration mehrerer Datenquellen gelingt. Nichtsdestotrotz sind vor allem die Bereiche “Globale Anwendbarkeit“ sowie “Präzision der Daten vs. Automatisierbarkeit“ als Kernthemen der weiteren Forschung festzuhalten.

Ferner ist die in der Arbeit präsentierte Vorgehensweise bzgl. der Finanzanalysen im Kontext des Semantic Web aus meiner Sicht äußerst interessant und verdient es, dass weitere Bemühungen getätigt werden. Diese müssen insbesondere in diversen Detailfragen, die das “Grundgerüst“ umgeben, erledigt werden. Ist dies geschehen, kann von der gewünschten generellen Anwendbarkeit, wie sie für einen solchen Ansatz nötig ist, gesprochen werden und ein Projekt mit – meiner Meinung nach – großem Potential gestartet werden.

A Anhang

A.1 Finanzkonzepte Abbildung 2

```
<ma:AccruedCustomerPrograms
  id="Fact-0F1FB4B2364F97D08ECDEE7C4B62097F"
  contextRef="FI2013Q4"
  unitRef="usd"
  decimals="-6">1286000000</ma:AccruedCustomerPrograms>
```

Abbildung 54: Financial facts aus Abbildung 2 1/4 (Beispiel: Mastercard)

```
<us-gaap:EmployeeRelatedLiabilitiesCurrent
  id="Fact-344745FF5827A4DDEF28EE7C4B607C69"
  contextRef="FI2013Q4"
  unitRef="usd"
  decimals="-6">413000000</us-gaap:EmployeeRelatedLiabilitiesCurrent>
```

Abbildung 55: Financial facts aus Abbildung 2 2/4 (Beispiel: Mastercard)

```
<us-gaap:TaxesPayableCurrent
  id="Fact-A2790DAA8F3745F470BEEE7C4B5E7F07"
  contextRef="FI2013Q4"
  unitRef="usd"
  decimals="-6">95000000</us-gaap:TaxesPayableCurrent>
```

Abbildung 56: Financial facts aus Abbildung 2 3/4 (Beispiel: Mastercard)

```
<us-gaap:OtherAccruedLiabilitiesCurrent
  id="Fact-A5E41E34199C7A75AF57EE7C4B5E452D"
  contextRef="FI2013Q4"
  unitRef="usd"
  decimals="-6">158000000</us-gaap:OtherAccruedLiabilitiesCurrent>
```

Abbildung 57: Financial facts aus Abbildung 2 4/4 (Beispiel: Mastercard)

A.2 80 SIC Gruppen innerhalb der Datenbank

Nr.	SIC	Name	Einträge
1	2780	BLANKBOOKS, LOOSELEAF BINDERS & BOOKBINDG & RELATD WORK	3
2	6029	COMMERCIAL BANKS, NEC	11
3	3620	ELECTRICAL INDUSTRIAL APPARATUS	12
4	4900	ELECTRIC, GAS & SANITARY SERVICES	14
5	5812	RETAIL-EATING PLACES	62
6	3873	WATCHES, CLOCKS, CLOCKWORK OPERATED DEVICES / PARTS	3
7	3721	AIRCRAFT	4
8	5063	WHOLESALE-ELECTRICAL APPARATUS & EQUIPMENT, WIRING SUPPLIES	7
9	7359	SERVICES-EQUIPMENT RENTAL & LEASING, NEC	37
10	5810	RETAIL-EATING & DRINKING PLACES	14
11	6022	STATE COMMERCIAL BANKS	294
12	3728	AIRCRAFT PARTS & AUXILIARY EQUIPMENT, NEC	11
13	5735	RETAIL-RECORD & PRERECORD TAPESTORES	2
14	2300	APPAREL & OTHER FINISHED PRODS OF FABRICS & SIMILAR MATL	21
15	7372	SERVICE-PREPACKAGED SOFTWARE	219
16	6311	LIFE INSURANCE	40
17	6111	FEDERAL & FERDERALLY SPONSORED CREDIT AGENCIES	15
18	2320	MEN'S & BOY'S FURNISHGS, WORK CLOTHG, & ALLIED GARMENTS	10
19	7200	SERVICES-PERSONAL SERVICES	31
20	3861	PHOTOGRAPHIC EQUIPMENT & SUPPLIES	10
21	3743	RAILROAD EQUIPMENT	4
22	2771	GREETING CARDS	2
23	8050	SERVICES-NURSING & PERSONAL CARE FACILITIES	8
24	7361	SERVICES-EMPLOYMENT AGENCIES	9
25	3825	INSTRUMENTS FOR MEAS & TESTING OF ELECTRICITY & ELEC SIGNALS	19
26	5110	WHOLESALE-PAPER & PAPER PRODUCTS	2
27	8741	SERVICES-MANAGEMENT SERVICES	13

Nr.	SIC	Name	Einträge
28	3571	ELECTRONIC COMPUTERS	12
29	4220	PUBLIC WAREHOUSE & STORAGE	2
30	7011	HOTELS & MOTELS	44
31	4700	TRANSPORTATION SERVICES	18
32	3576	COMPUTER COMMUNICATIONS EQUIPMENT	29
33	3711	MOTOR VEHICLES & PASSENGER CAR BODIES	18
34	8711	SERVICES-ENGINEERING SERVICES	19
35	5051	WHOLESALE-METALS SERVICE CENTERS & OFFICES	9
36	2273	CARPETS & RUGS	3
37	6282	INVESTMENT ADVICE	34
38	4922	NATURAL GAS TRANSMISSION	41
39	5661	RETAIL-SHOW STORES	7
40	2082	MALT BEVERAGES	6
41	2911	PETROLEUM REFINING	23
42	4812	RADIOTELEPHONE COMMUNICATIONS	13
43	6361	TITLE INSURANCE	6
44	4731	ARRANGEMENT OF TRANSPORTATION OF FREIGHT & CARGO	14
45	3829	MEASURING & CONTROLLING DEVICES, NEC	31
46	4899	COMMUNICATIONS SERVICES, NEC	49
47	7371	SERVICES-COMPUTER PROGRAMMING SERVICES	40
48	3714	MOTOR VEHICLE PARTS & ACCESSORIES	45
49	6531	REAL ESTATE AGENTS & MANAGERS (FOR OTHERS)	19
50	3674	SEMICONDUCTORS & RELATED DEVICES	147
51	2833	MEDICINAL CHEMICALS & BOTANICAL PRODUCTS	20
52	3613	SWITCHGEAR & SWITCHBOARD APPARATUS	4
53	1040	GOLD AND SILVER ORES	85
54	0100	AGRICULTURAL PRODUCTION-CROPS	13
55	4213	TRUCKING (NO LOCAL)	21
56	5712	RETAIL-FURNITURE STORES	2
57	4841	CABLE & OTHER PAY TELEVISION SERVICES	34
58	3312	STEEL WORKS, BLAST FURNACES & ROLLING MILLS (COKE OVENS)	13
59	6798	REAL ESTATE INVESTMENT TRUSTS	277
60	6153	SHORT-TERM BUSINESS CREDIT INSTITUTIONS	16

Nr.	SIC	Name	Einträge
61	3060	FABRICATED RUBBER PRODUCTS, NEC	8
62	7310	SERVICES-ADVERTISING	34
63	5651	RETAIL-FAMILY CLOTHING STORES	13
64	4833	TELEVISION BROADCASTING STATIONS	39
65	1700	CONSTRUCTION - SPECIAL TRADE CONTRACTORS	15
66	2840	SOAP, DETERGENTS, CLEANING PREPARATIONS, PERFUMES, COSMETICS	6
67	3221	GLASS CONTAINERS	2
68	3272	CONCRETE PRODUCTS, EXCEPT BLOCK & BRICK	7
69	2400	LUMBER & WOOD PRODUCTS (NO FURNITURE)	6
70	6512	OPERATORS OF NONRESIDENTIAL BUILDINGS	20
71	6221	COMMODITY CONTRACTS BROKERS & DEALERS	108
72	5411	RETAIL-GROCERY STORES	15
73	3524	LAWN & GARDEN TRACTORS & HOME LAWN & GARDENS EQUIP	1
74	7311	SERVICES-ADVERTISING AGENCIES	10
75	1381	DRILLING OIL & GAS WELLS	46
76	5940	RETAIL-MISCELLANEOUS SHOPPING GOODS STORES	17
77	8200	SERVICES-EDUCATIONAL SERVICES	44
78	1221	BITUMINOUS COAL & LIGNITE SURFACE MINING	20
79	3541	MACHINE TOOLS, METAL CUTTING TYPES	3
80	5600	RETAIL-APPAREL & ACCESSORY STORES	22
.	.	.	.
.	.	.	.
.	.	.	.

A.3 Ergebnisse SPARQL-Abfragen

Diese Teil des Anhangs präsentiert Ergebnisse der SPARQL-Queries, die aus Gründen der Lesbarkeit nicht im Hauptteil der Arbeit enthalten sind.

A.3.1 Abfrage Ergebnisse zu Abbildung 19

?company	?cikNr
http://edgarwrap.ontologycentral.com/cik/872821#	872821
http://edgarwrap.ontologycentral.com/cik/737210#	737210
http://edgarwrap.ontologycentral.com/cik/730708#	730708
.	.
.	.
.	.

A.3.2 Abfrage Ergebnisse zu Abbildung 22

?value	?cik	?bankName
295308727	1070295	CAPITOL CITY BANCSHARES INC
297823412	1070295	CAPITOL CITY BANCSHARES INC
1497780000	1068300	METROCORP BANCSHARES, INC.
1497780000	1068300	METROCORP BANCSHARES, INC.
1558585000	1068300	METROCORP BANCSHARES, INC.
2234472000	1085706	PENNSYLVANIA COMMERCE BANCORP INC
.	.	.
.	.	.
.	.	.

A.4 SPARQL-Abfragen

Abfrage zur Kennzahl 2 (Umsatz in Periode t pro Mitarbeiter) in einer SPARQL-Query:

```

@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ((?revenues2/?employees2) AS ?result) ?cikNr ?name
WHERE {
  ?company edgar:cikNumber [number] .
  ?company edgar:issued ?facts .
  ?facts rdfs:seeAlso ?bn .
  ?bn edgar:subject us-gaap-ab:totalInterest .
  ?bn ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn measure:obsValue ?revenues .
  ?dbcompany ?p1 ?employees .
  ?p1 rdf:type ab:DBnumberOfEmployees .

  ?company2 edgar:issued ?facts2 .
  ?facts2 rdfs:seeAlso ?bn2 .
  ?bn2 edgar:subject us-gaap-ab:totalInterest .
  ?bn2 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn2 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn2 measure:obsValue ?revenues2 .
  ?company2 owl:sameAs dbcompany2 .
  ?dbcompany2 ?p2 ?employees2 .
  ?p2 rdf:type ab:DBnumberOfEmployees .
  ?company2 edgar:cikNumber ?cikNr .
  ?company2 foaf:name ?name .

  BIND ((?revenues/?employees) AS ?index)
  FILTER (?result > ?index)
}

{ ?x edgar:cikNumber [number] .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .

{ ?y edgar:cikNumber ?number .
  ?y owl:sameAs ?dbcompany2 .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany2 . } .

```

Abbildung 58: Beide Abfragen zu Kennzahl 2 innerhalb einer Query

Abfrage zur Kennzahl 3 (Pro-Kopf-Gewinn / Pro-Kopf-Verlust) in einer SPARQL-Query:

```

@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT ((?profit2/?employees2) AS ?result) ?cikNr ?name
WHERE {
  ?company edgar:cikNumber [number] .
  ?company edgar:issued ?facts .
  ?facts rdfs:seeAlso ?bn .
  ?bn edgar:subject us-gaap-ab:NetIncomeLoss .
  ?bn ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn measure:obsValue ?profit .
  ?dbcompany ?p1 ?employees .
  ?p1 rdf:type ab:DBnumberOfEmployees .

  ?company2 edgar:issued ?facts2 .
  ?facts2 rdfs:seeAlso ?bn2 .
  ?bn2 edgar:subject us-gaap-ab:NetIncomeLoss .
  ?bn2 ical:dtstart "2009-01-01"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn2 ical:dtend "2009-12-31"^^<http://www.w3.org/2001/XMLSchema#date> .
  ?bn2 measure:obsValue ?profit2 .
  ?dbcompany2 ?p2 ?employees2 .
  ?p2 rdf:type ab:DBnumberOfEmployees .
  ?company2 owl:sameAs ?dbcompany2 .
  ?company2 edgar:cikNumber ?cikNr .
  ?company2 foaf:name ?name .

  BIND ((?profit/?employees) AS ?index)
  FILTER (?result > ?index)
}

{ ?x edgar:cikNumber [number] .
  ?x owl:sameAs ?dbcompany .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany . } .

{ ?y edgar:cikNumber ?number .
  ?y owl:sameAs ?dbcompany2 .
} => {
  [] http:mthd httpm:GET ;
    http:requestURI ?dbcompany2 . } .

```

Abbildung 59: Beide Abfragen zu Kennzahl 3 innerhalb einer Query

Abfragen zu Kennzahl 4 (Modifikation des Zeitraums der Abfragen zu Abbildung 32.

```

@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT (((?revenues1+?revenues2+?revenues3)/?locations) AS ?result)
      ?revenues1 ?revenues2 ?revenues3 ?locations
WHERE { ?company edgar:cikNumber "35527" .
      ?company edgar:issued ?facts .
      ?facts rdfs:seeAlso ?bn1 .
      ?facts rdfs:seeAlso ?bn2 .
      ?facts rdfs:seeAlso ?bn3 .
      ?bn1 edgar:subject us-gaap-ab:InterestAndDividendIncomeOperating .
      ?bn1 ical:dtstart "2010-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 ical:dtend "2010-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 measure:obsValue ?revenues1 .
      ?bn2 edgar:subject us-gaap-ab:InterestAndDividendIncomeSecurities .
      ?bn2 ical:dtstart "2010-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 ical:dtend "2010-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 measure:obsValue ?revenues2 .
      ?bn3 edgar:subject us-gaap-ab:InterestAndFeeIncomeLoansAndLeases .
      ?bn3 ical:dtstart "2010-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 ical:dtend "2010-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 measure:obsValue ?revenues3 .
      ?dbcompany ?p ?locations .
      ?p rdf:type ab:DBlocations .
    }

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?dbcompany . } .

```

Abbildung 60: Modifizierte Abfragen zu Kennzahl 4 (Jahr 2010)

```

@prefix ab: <http://alexander-buescher.de/> .
@prefix ical: <http://www.w3.org/2002/12/cal/ical#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix edgar: <http://edgarwrap.ontologycentral.com/vocab/edgar#> .
@prefix us-gaap-ab: <http://alexander-buescher.de/usgaap#> .
@prefix measure: <http://purl.org/linked-data/sdmx/2009/measure#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix http: <http://www.w3.org/2011/http#> .
@prefix httpm: <http://www.w3.org/2011/http-methods#> .

SELECT (((?revenues1+?revenues2+?revenues3)/?locations) AS ?result)
      ?revenues1 ?revenues2 ?revenues3 ?locations
WHERE { ?company edgar:cikNumber "35527" .
      ?company edgar:issued ?facts .
      ?facts rdfs:seeAlso ?bn1 .
      ?facts rdfs:seeAlso ?bn2 .
      ?facts rdfs:seeAlso ?bn3 .
      ?bn1 edgar:subject us-gaap-ab:InterestAndDividendIncomeOperating .
      ?bn1 ical:dtstart "2011-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 ical:dtend "2011-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn1 measure:obsValue ?revenues1 .
      ?bn2 edgar:subject us-gaap-ab:InterestAndDividendIncomeSecurities .
      ?bn2 ical:dtstart "2011-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 ical:dtend "2011-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn2 measure:obsValue ?revenues2 .
      ?bn3 edgar:subject us-gaap-ab:InterestAndFeeIncomeLoansAndLeases .
      ?bn3 ical:dtstart "2011-01-01" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 ical:dtend "2011-12-31" ^^<http://www.w3.org/2001/XMLSchema#date> .
      ?bn3 measure:obsValue ?revenues3 .
      ?dbcompany ?p ?locations .
      ?p rdf:type ab:DBlocations .
}

{ ?x edgar:cikNumber "35527" .
  ?x owl:sameAs ?dbcompany .
} => {
[] http:mthd httpm:GET ;
  http:requestURI ?dbcompany . } .

```

Abbildung 61: Modifizierte Abfragen zu Kennzahl 4 (Jahr 2011)

A.5 Umfrage

Evaluation der Durchführbarkeit und Performance der Integration von Finanzdaten im Semantic Web

Diese Umfrage enthält 6 Fragen.

Allgemeine Fragen zur Person

Dieser Teil der Umfrage sammelt Daten über den Probanden. Hierbei geht es vor allem um die Validierung der Repräsentativität, sowie die Betrachtung der Altersverteilung. Alle Daten werden selbstverständlich gemäß dem aktuellen Datenschutz behandelt.

[]Wie alt sind Sie? *

In dieses Feld dürfen nur Zahlen eingegeben werden.

Bitte geben Sie Ihre Antwort hier ein:

[]Bitte geben Sie Ihr Geschlecht an. *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ☐ weiblich
☐ männlich

[]Die folgenden kurzen Fragen beschäftigen sich mit Ihren Vorkenntnissen zum Zeitpunkt des Ausfüllens der Umfrage *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	1	2	3	4	5
Bitte bewerten Sie Ihre Vorkenntnisse im Bereich Informatik	☐	☐	☐	☐	☐
Bitte bewerten Sie Ihre Vorkenntnisse im Bereich Semantic Web	☐	☐	☐	☐	☐
Bitte bewerten Sie Ihre Vorkenntnisse im Bereich BWL generell	☐	☐	☐	☐	☐
Bitte bewerten Sie Ihre Vorkenntnisse im Bereich Finanzwirtschaft und Rechnungswesen	☐	☐	☐	☐	☐

Anmerkung: Hierbei ist folgende Skala zu verwenden:

[1 = keine Kenntnisse | 2 = wenige Kenntnisse | 3 = durchschnittliche Kenntnisse | 4 = gute Kenntnisse | 5 = herausragende Kenntnisse]

Fragen zur Evaluation des Anwendungsfalls

Auf Basis des vorgestellten Use Cases werden im folgenden Teil der Umfrage 9 Fragen gestellt, deren Beantwortung Auskunft über die Anwendbarkeit, hinsichtlich sowohl wirtschaftlicher als auch technischer Aspekte, geben sollen.

[] Verschiedene Fragen hinsichtlich Wirtschaftlichkeit und technischer Umsetzbarkeit des Anwendungsfalles. Bitte die Fragen genau lesen und auf Formulierungen achten! *

Bitte wählen Sie die zutreffende Antwort für jeden Punkt aus:

	Stimme zu	Stimme eher zu	Neutral	Stimme eher nicht zu	Stimme nicht zu
Ich erkenne die Relevanz für Unternehmen im Anwendungsfall	☐	☐	☐	☐	☐
Die Abfragemöglichkeiten sind angemessen konfigurierbar	☐	☐	☐	☐	☐
Ich könnte mir vorstellen, dass Unternehmen Finanz- und Konkurrenzanalysen mit dem präsentierten Ansatz durchführen	☐	☐	☐	☐	☐
Ich denke, der Individualisierungsgrad könnte zu gering sein	☐	☐	☐	☐	☐
Die präsentierten Vorgehensweisen decken den relevanten Teil für Finanz- und Konkurrenzanalysen ab	☐	☐	☐	☐	☐
Die Vorgehensweise der Abfragen wirken auch nach kurzer Erläuterung nach wie vor komplex	☐	☐	☐	☐	☐
Die Ergebnisse sind angemessen dargestellt und leicht interpretierbar	☐	☐	☐	☐	☐
Der Mehrwert für Unternehmen durch dieses System ist mir nicht ersichtlich	☐	☐	☐	☐	☐
Die Abfragen wirken intuitiv, gut strukturiert und verständlich	☐	☐	☐	☐	☐

[]Machen die Vorgehensweisen der Abfragen einen vollständigen Eindruck auf Sie? *

Bitte wählen Sie nur eine der folgenden Antworten aus:

- ☐ Ja
☐ Nein, mir fehlt:

Bitte schreiben Sie einen Kommentar zu Ihrer Auswahl

[]Platz für weitere Anmerkungen jeglicher Art:

Bitte geben Sie Ihre Antwort hier ein:

A.6 Auswertungen der Befragungen

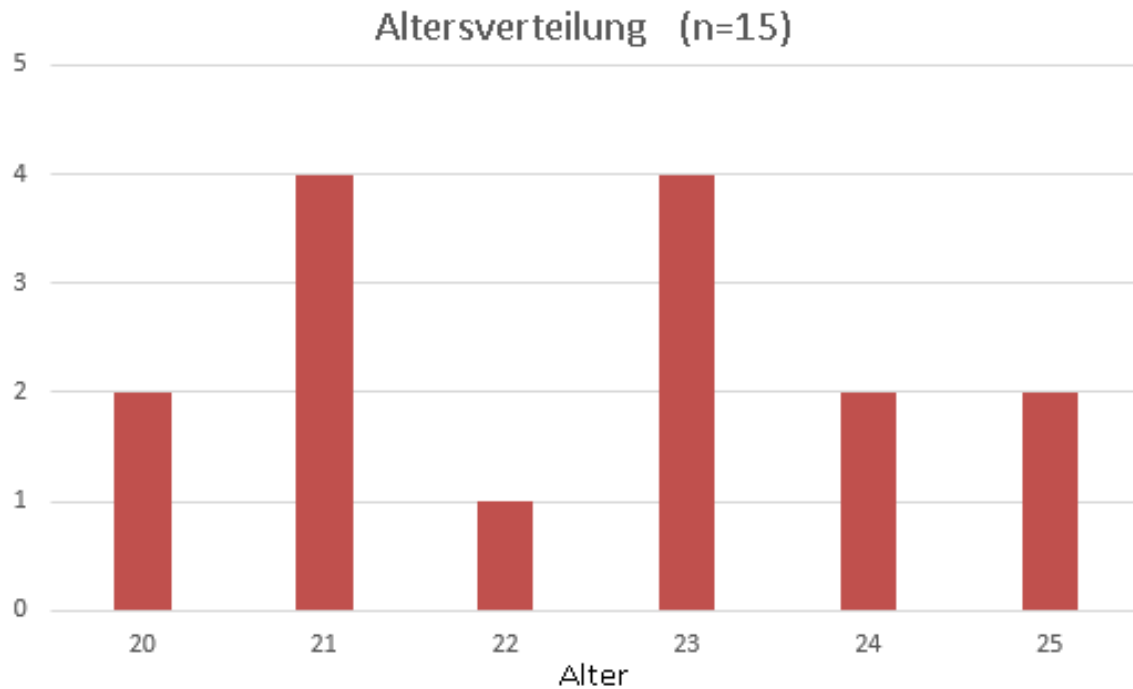


Abbildung 62: Altersverteilung der Teilnehmer

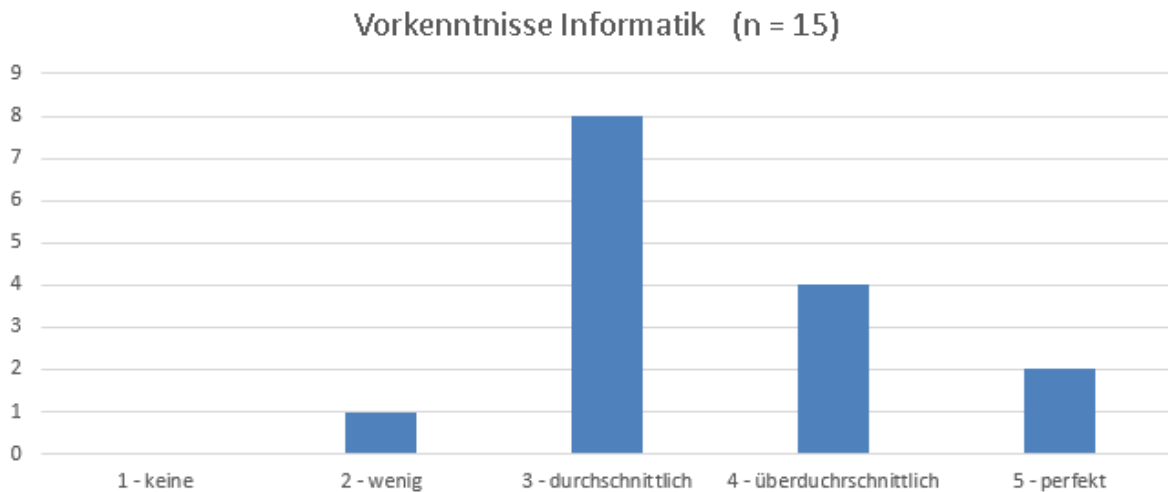


Abbildung 63: Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich Informatik

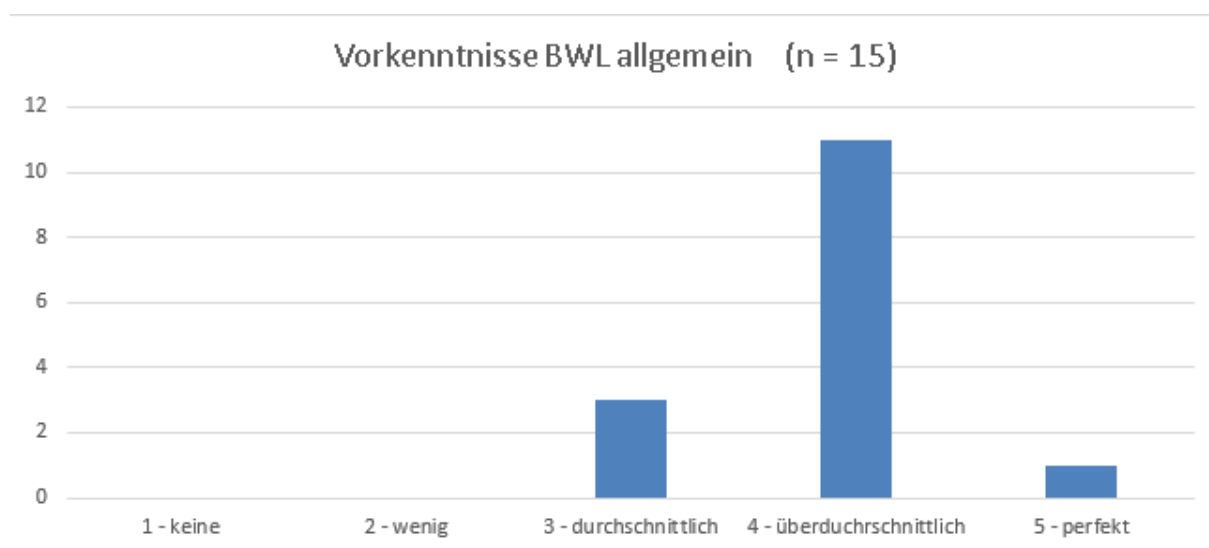


Abbildung 64: Auswertung der Evaluationsergebnisse: Angaben zu den Vorkenntnissen im Bereich BWL allg.

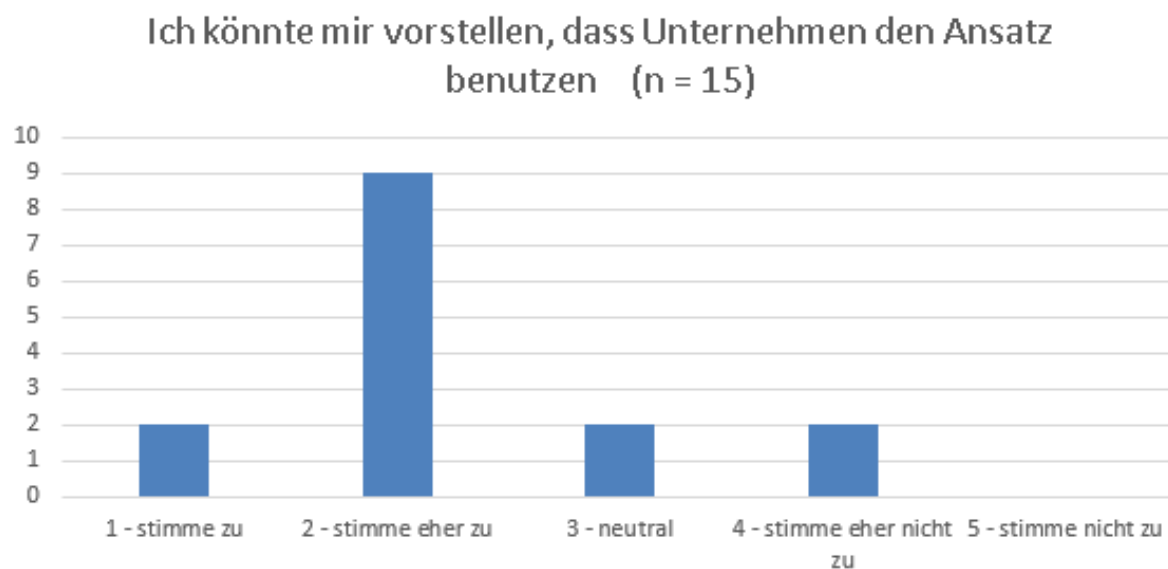


Abbildung 65: Auswertung der Evaluationsergebnisse: Angaben zur Einschätzung, ob Unternehmen den präsentierten Ansatz benutzen

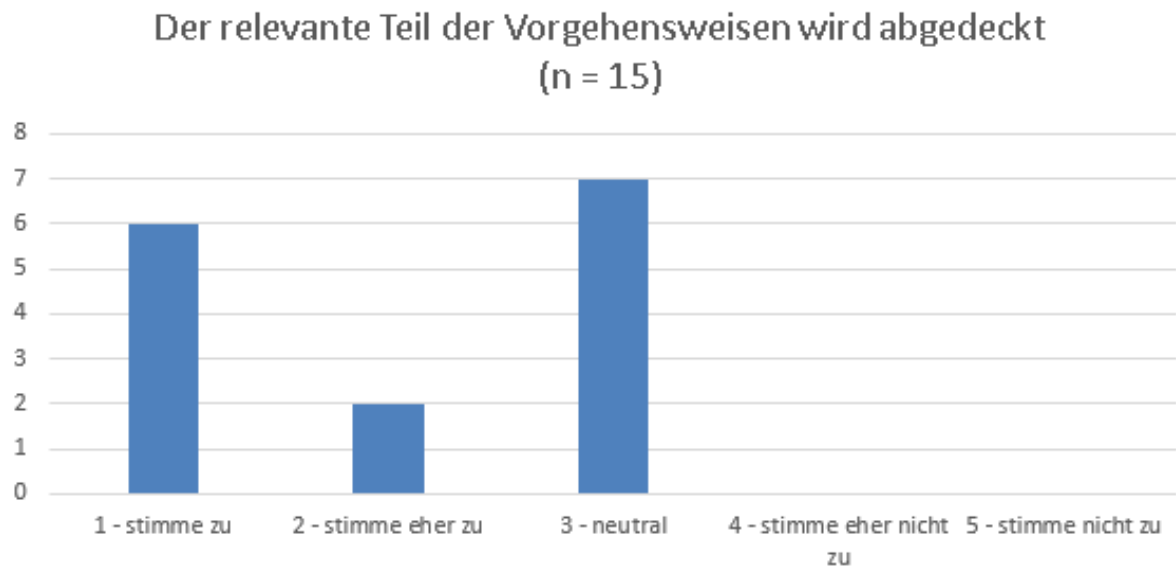


Abbildung 66: Auswertung der Evaluationsergebnisse: Angaben zur Einschätzung, ob durch den Use Case alle relevanten Vorgehensweisen abgedeckt wurden

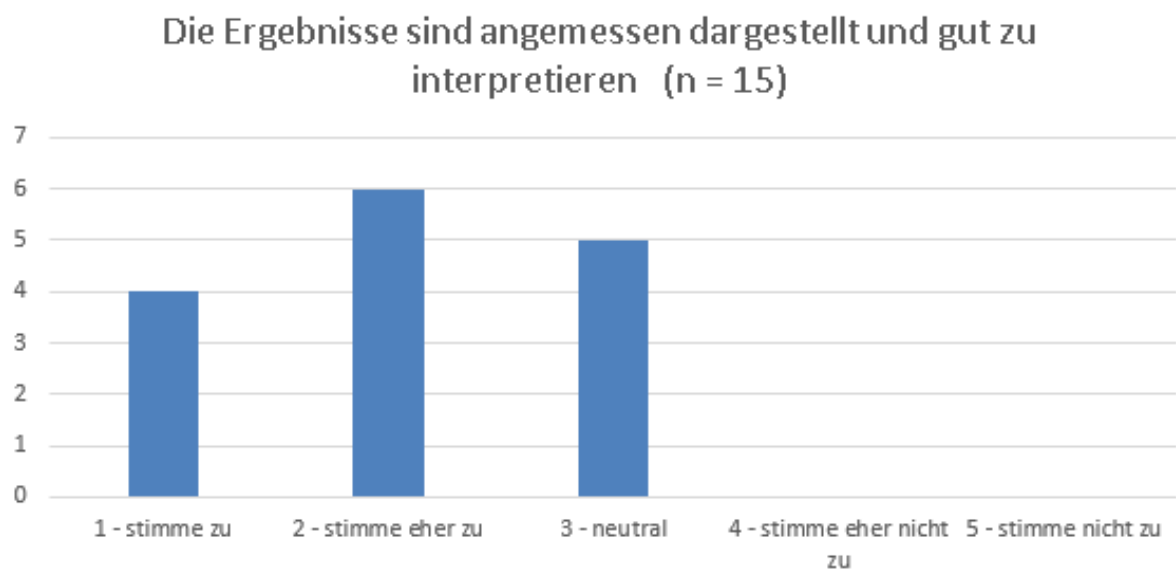


Abbildung 67: Auswertung der Evaluationsergebnisse: Angaben zur Darstellung und Interpretationsmöglichkeit der Ergebnisse

Literatur

- [AG09] ALBERTAZZI, U. und L. GAMBACORTA: *Bank profitability and the business cycle*. Journal of Financial Stability, 5(4):393–409, 2009.
- [BRLD10] BAO, J., G. RONG, X. LI und L. DING: *Representing Financial Reports on the Semantic Web: a Faithful Translation from XBRL to OWL*. In: *Proceedings of the 2010 international conference on Semantic web rules*, 2010.
- [BS07] BHAGWAT, R. und M. SHARMAR: *Performance measurement of supply chain management: A balanced scorecard approach*. Computers & Industrial Engineering, 53(1):43–62, 2007.
- [Bur11] BURDELSKI, DR. T.: *Institut für Finanzwirtschaft Banken und Versicherungen Universität Karlsruhe (TH) – Rechnungswesen I*, 2011.
- [CCP08] CHEN, T., C. CHEN und S. PENG: *Firm operation performance analysis using data envelopment analysis and balanced scorecard: A case study of a credit cooperative bank*. International Journal of Productivity and Performance Management, 57(7):523–539, 2008.
- [DFO⁺09] DEBRCENY, R., C. FELDEN, B. OCHOCKI, M. PIECHOCKI und M. PIECHOCKI: *XBRL for Interactive Data - Engineering the Information Value Chain*. Springer Verlag, Heidelberg, 2009.
- [DG05] DEVLIN, J. und P. GERRARD: *A study of customer choice criteria for multiple bank users*. Journal of Retailing and Consumer services, 12(4):297–306, 2005.
- [GG10] GARCIA, R. und R. GIL: *Triplifying and linking XBRL financial data*. In: *Proceedings of the 6th International Conference on Semantic Systems*, 2010.
- [Gio08] GIOKAS, D.: *Assessing the efficiency in operations of a large Greek bank branch network adopting different economic behaviors*. Economic Modelling, 25(3):559–574, 2008.
- [GS14] GEYER-SCHULZ, A.: *IISM/EM Informationsdienste und elektronische Märkte Karlsruher Institut für Technologie (KIT) – BWL der Informationsunternehmen, Vorlesung 6. Entscheidungen bei Sicherheit (Teil 2)*, 2014.
- [GST03] GEYER-SCHULZ, A. und A. TAUDS: *Informationswirtschaft: Ein Sektor mit Zukunft*. 2003.
- [HH10] HALPIN, H. und P. HAYES: *When owl:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web*. In: *Linked Data on the Web*, 2010.

- [HS97] HANLEY, C. A. und M.P. SUTERL: *Banking's top performance*. ABA Banking Journal, 89(7):36–40, 1997.
- [HS01] HOFFMAN, C. und C. STRAND: *XBRL Essentials: a nontechnical introduction to eXtensible business reporting language (XBRL), the digital language of business reporting*. American Institute of Certified Public Accountants, 2001.
- [Kle02] KLEIN, M.: *Interpreting XML documents via an RDF schema ontology*. In: *Proceedings of the 13th Int. Workshop on Database and Expert Systems Applications*, Seiten 889–894, 2002.
- [KPDZ06] KOSMIDOU, K., F. PASIOURAS, M. DOUMPOSL und C. ZOPOUNIDIS: *Assessing performance factors in the UK banking sector: a multicriteria methodology*. Central European Journal of Operations Research, 14(1):25–44, 2006.
- [KWO⁺14] KÄMPGEN, B., T. WELLER, S. O'RIAIN, C. WEBER und A. HARTH: *Accepting the XBRL Challenge with Linked Data for Financial Data Integration*. In: *International Semantic Web Conference*, Seiten 595–610. Springer, 2014.
- [Mil98] MILLER, E.: *An introduction to the resource description framework*. Bulletin of the American Society for Information Science and Technology, 25(1):15–19, 1998.
- [OCH12] O'RIAIN, S., E. CURRY und A. HARTH: *XBRL and open data for global financial ecosystems: A linked data approach*. International Journal of Accounting Information Systems, (13), Februar 2012.
- [SSHS13] STADTMÜLLER, S., S. SPEISER, A. HARTH und R. STUDER: *Data-Fu: A Language and an Interpreter for Interaction with Read/Write Linked Data*. In: *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, Seiten 1225–1236, 2013.
- [Sta07] STANTIAL, JOHN: *ROI on XBRL: interactive data cuts reporting costs today*. Journal of Accountancy, June 2007.
- [SWN08] SWAIN, S., D. WEATHERS und R. NIEDRICH: *Assessing three sources of misresponse to reversed Likert items*. Journal of Marketing Research, 45(1):116–131, 2008.
- [SZ99] SEIFORD, L. und J. ZHU: *Profitability and marketability of the top 55 US commercial banks*. management science, 45(9):1270–1288, 1999.
- [WCS10] WEIJTERS, B., E. CABOOTER und N. SCHILLEWART: *The effect of rating scale format on response styles: The number of response categories and response category labels*. International Journal of Research in Marketing, 27(3):236–247, 2010.

-
- [Wei01] WEI, QUANLING: *Data envelopment analysis*. Chinese Science Bulletin, 46(16):1321–1332, 2001.
- [WSN05] WEATHERS, D., S. SHARMA und R. W. NIEDRICH: *The impact of the number of scale points, dispositional factors, and the status quo decision heuristic on scale reliability and response accuracy*. Journal of Business Research, 58(11):1516–1524, 2005.
- [Wu12] WU, H.: *Constructing a strategy map for banking institutions with key performance indicators of the balanced scorecard*. Evaluation and program planning, 35(3):303–320, 2012.
- [ZW10] ZHU, H. und H. WU: *Quality of XBRL US GAAP Taxonomy: Empirical Evaluation using SEC Filings*, 2010.