

Projektdokumentation im Modul Semantic Web – Zusammenhang von Terroranschlägen mit Wirtschaft und Politik

Hochschule für Technik, Wirtschaft und Kultur Leipzig
Julius Seiffert - julius.seiffert@stud.htwk-leipzig.de

31.08.2019

Recherchefragestellung: Wie hängen Terroranschläge in Deutschland mit dem deutschen Aktienindex und den deutschen Wahlverhalten zusammen.

1 Inhaltliche Interpretation der Fragestellung

Im Rahmen dieser Forschungsarbeit ist eine konkrete Interpretation, bzw. Schlussforderung natürlich nicht möglich. Alle Zusammenhänge die zwischen den Datenquellen gefunden wurden können durch Zufall entstanden sein.

Natürlich kann es bei dieser Fragestellung auch direkte Zusammenhänge geben. Terroranschläge können in der Gesellschaft einer Nation Angst und Hass schüren, was sich wiederum auf das Wahlverhalten auswirken kann. Auch können Terroranschläge bestimmte Bereiche der Wirtschaft finanziell zu Gute kommen, allerdings auch genauso Schaden. Beispielsweise könnte der Bereich der Waffen- und Militärausrüstung davon profitieren, oder bei vorhandener Angst der Einzelhandel leiden.

2 Relevante Datenquellen

Relevante Datenquellen waren die Terrordatenbank der University Maryland, die Website Yahoo Finance, welche historische Daten des DAX zum Download zur Verfügung stellt und die Sonntagsfrage des FORSA Instituts, welches nahezu jeden Sonntag eine Wahlumfrage erhebt.

2.1 Terrordatenbank der University Maryland

Die Datenbank der University Maryland beinhaltet globale Terroranschläge von 1970 bis 2017. Auf der Website werden die Daten als CSV-Datei zum Download angeboten und können auch direkt über eine Tabelle auf der Website angezeigt werden.

Link	https://www.start.umd.edu/gtd/about/
Datenformat	CSV, HTML
Schnittstelle	Webcarper, manueller Download
Lizenz	https://www.start.umd.edu/gtd/terms-of-use/
Open Data	***

2.2 Yahoo Finance

Die Website Yahoo Finance stellt sämtliche Börsenkurse und auch historische Daten davon bereit. Yahoo Finance hat bis zum März 2019 eine API-Schnittstelle der verfügbaren Finanzdaten angeboten. Leider wurde diese Schnittstelle eingestellt und die Daten können nur durch einen Webcarper, oder einen manuellen Download heruntergeladen werden. Eine Automatisierung des Downloads gestaltet sich durch Sicherheitsmaßnahmen der Website als schwierig. Aufgrund dessen wurden die Tageshöchstsätze des DAX manuell als CSV-Datei heruntergeladen, welche bis zum 01.01.2004 zurückgeht.

Link	https://de.finance.yahoo.com/
Datenformat	CSV, HTML
Schnittstelle	manueller Download
Lizenz	https://policies.oath.com/ie/de/oath/terms/otos/index.html
Open Data	★★

2.3 Sonntagsfrage FORSA

Das FORSA Institut stellt zu jeden Sonntag die Sonntagsfrage. Bei dieser Frage werden die Befragten gefragt, welche Partei sie wählen würden, wenn am nächsten Sonntag die Bundeswahl wäre. Diese Daten gehen bis 1998 zurück. Es gibt allerdings auch Zeiträume, in denen keine Daten vorhanden sind.

Die Daten werden auf der Website <https://www.wahlrecht.de/> in HTML-Tabellen bereitgestellt.

Link	https://www.wahlrecht.de/umfragen/forsa.htm
Datenformat	HTML
Schnittstelle	Webscarper
Lizenz	
Open Data	★★

3 Vokabular

Für den Aufbau der Datenbasis wurde das RDF-Vokabular verwendet. Der Präfix zu dem RDF-Vokabular ist <http://www.w3.org/1999/02/22-rdf-syntax-ns>

Für die spezifischen Daten wurde ein neues Vokabular erstellt. Dieses wurde unter dem Präfix <https://github.com/juuls28/semanticWEB> verwendet.

Listing 1: Präfixe

```

1  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
2  xmlns:j.0="https://github.com/juuls28/semanticWEB#"

```

3.1 Terroranschlag

Die Daten der Terroranschläge sind durch Ihre ID gekennzeichnet.

Aus der Terrordatenbank lassen sich viele Daten zu jedem Terroranschlag auslesen. Hier wurden nur die wichtigsten extrahiert.

happenedOn beschreibt das Datum im Format yyyy-MM-dd an dem das Ereignis geschehen ist.

inCountry beschreibt das Land in dem das Ereignis passiert ist als Zeichenkette.

inCity beschreibt die Stadt in dem das Ereignis passiert ist als Zeichenkette.

hasFatalities beschreibt die Anzahl der getöteten Personen als ganze Zahl.

hasInjuries beschreibt die Anzahl der verletzten Personen als ganze Zahl.

Listing 2: Beispiel RDF Terroranschlag

```

1  <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#201608040029">
2    <j.0:hasInjuries>0</j.0:hasInjuries>
3    <j.0:hasFatalities>0</j.0:hasFatalities>
4    <j.0:inCity>Berlin</j.0:inCity>
5    <j.0:inCountry>Germany</j.0:inCountry>

```

```

6      <j.0:happenedOn>2016-08-04</j.0:happenedOn>
7    </rdf:Description>

```

3.2 DAX-Kurs

Die Daten des Aktienkurses sind durch das Wort DAX und das Datum eindeutig identifizierbar. Die Ressourcen haben daher die Form DAXyyyy-MM-dd.

Bei den historischen Daten des DAX-Kurses wird das Datum und der zugehörige Eröffnungskurs beachtet.

happenedOn beschreibt das Datum im Format yyyy-MM-dd an dem das Ereignis geschehen ist.

value beschreibt den Wert des Eröffnungskurses am Tag der Ressource als Fließkommazahl.

Listing 3: Beispiel RDF DAX

```

1  <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#DAX2009-11-02">
2    <j.0:value>5410.609863</j.0:value>
3    <j.0:happenedOn>2009-11-02</j.0:happenedOn>
4  </rdf:Description>

```

3.3 Wahlumfrage

Die Wahlumfrage wird durch das Datum im Format yyyy-MM-dd gekennzeichnet. Sie beinhaltet das Datum und eine Liste der Parteien mit den zugehörigen Prozentsatz, mit dem diese gewählt worden wären.

happenedOn beschreibt das Datum im Format yyyy-MM-dd an dem das Ereignis geschehen ist.

outcomes beinhaltet eine Liste der Ressource Partei.

3.4 Partei

Die Ressource Partei wird durch das Datum im Format yyyy-MM-dd und den Parteinamen gekennzeichnet. Sie repräsentiert das Wahlergebnis zu jeder Partei an dem Datum der Wahl.

partyName beschreibt den Namen der Partei als Zeichenkette.

partyPercent beschreibt das Wahlergebnis (Prozentsatz) der Partei als Fließkommazahl.

Listing 4: Beispiel RDF Wahlumfrage

```

1  <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21">
2    <j.0:outcomes>
3      <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21Sonstige">
4        <j.0:partyPercent>4.0</j.0:partyPercent>
5        <j.0:partyName>Sonstige</j.0:partyName>
6      </rdf:Description>
7    </j.0:outcomes>
8    <j.0:outcomes>
9      <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21PDS">
10       <j.0:partyPercent>5.0</j.0:partyPercent>
11       <j.0:partyName>PDS</j.0:partyName>
12     </rdf:Description>
13   </j.0:outcomes>
14   <j.0:outcomes>
15     <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21FDP">
16       <j.0:partyPercent>9.0</j.0:partyPercent>
17       <j.0:partyName>FDP</j.0:partyName>
18     </rdf:Description>
19   </j.0:outcomes>
20   <j.0:outcomes>
21     <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21GRÜNE">

```

```

22         <j.0:partyPercent>7.0</j.0:partyPercent>
23         <j.0:partyName>GRÜNE</j.0:partyName>
24     </rdf:Description>
25 </j.0:outcomes>
26 <j.0:outcomes>
27     <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21SPD">
28         <j.0:partyPercent>39.0</j.0:partyPercent>
29         <j.0:partyName>SPD</j.0:partyName>
30     </rdf:Description>
31 </j.0:outcomes>
32 <j.0:outcomes>
33     <rdf:Description rdf:about="https://github.com/juuls28/semanticWEB#2001-06-21CDU/CSU">
34         <j.0:partyPercent>36.0</j.0:partyPercent>
35         <j.0:partyName>CDU/CSU</j.0:partyName>
36     </rdf:Description>
37 </j.0:outcomes>
38 <j.0:happenedOn>2001-06-21</j.0:happenedOn>
39 </rdf:Description>

```

4 Extraktion relevanter Daten und Import in einen Triplestore

Die Datensätze wurde teilweise automatisch und teilweise manuell heruntergeladen und weiterverarbeitet. Als Basis dient hierzu ein Java-Projekt welches unter dem Github Repository <https://github.com/juuls28/semanticWEB> einsehbar ist.

Aus den einzelnen Daten wurde mit Hilfe des Apache Jena Frameworks ein Tripelstore gebildet und SPARQL-Abfragen ausgeführt.

4.1 Extraktion Terrordatenbank

Die CSV-Datei mit den Daten zu den Terroranschlägen kann direkt durch einer URL in das Programm geladen werden. Mit in die URL können Filterparameter wie Startdatum, Enddatum, Land, Stadt etc. angegeben werden.

Hier war nur das Start- und Enddatum relevant, welche dynamisch anhand des Suchzeitraums hinzugefügt werden. Das Land bleibt statisch auf Deutschland eingestellt. Eine Abfrage vom 01.01.2010 bis zum 01.01.2011 würde wie folgt aussehen: https://www.start.umd.edu/gtd/search/ResultsCSV.aspx?csv=1&casualties_type=b&casualties_max=&start_year=2010&start_month=1&start_day=1&end_year=2011&end_month=1&end_day=1&ctp2=all&country=75

Aus den CSV-Dateien konnte dann die Daten zu den Anschlägen ausgelesen werden. Hierbei wurden fehlerhafte Daten bereinigt. Wenn die Daten nicht bereinigt werden konnten, wurden sie nicht übernommen.

4.2 Extraktion DAX-Kurs

Der DAX-Kurs wurde manuell als CSV-Datei heruntergeladen und in das Projekt importiert. Die Daten konnten dann wie bei der Terrordatenbank in das Programm geladen und bereinigt werden.

4.3 Extraktion Wahlumfrage

Die Daten zur Wahlumfrage mussten direkt von der HTML-Tabelle auf der Website ausgelesen werden. Hierfür wurde die Bibliothek jsoup¹ verwendet. Jsoup ist ein HTML-Parser mit dem Webseiten nach bestimmten Merkmalen untersucht werden können. Mithilfe dieser Bibliothek konnten die Daten ausgelesen und in aufbereitet werden.

Leider werden auch auf dieser Website inkonsistente Daten bereitgestellt, weswegen bestimmte Datensätze aufbereitet werden mussten, oder nicht verwertbar waren.

5 Verlinkung von Ressourcen

Die Verschiedenen Ressourcen werden über das Datum, also über die Eigenschaft <https://github.com/juuls28/semanticWEB#happenedOn/> verlinkt. Dazu wurde mit Hilfe des Frameworks Apache Jena ein RDF Model

¹jsoup ist eine OpenSource Bibliothek und ist unter der Verwendung der MIT-Lizenz benutzbar. Informationen hierzu gibt es unter <https://jsoup.org/>

erstellt. Dieses Model fungiert als Tripel Store. Das ganze Model ist in dem Projekt unter dem Pfad /data/-model.rdf exportiert worden, damit es auch in anderen Services wie den SPARQL Server Apache Jena Fuseki verwendet werden kann.

6 Anfrage an die Forschungswissensbasis

Im Rahmen der geschriebenen Software gibt es drei SPARQL-Anfragen

6.1 SPARQL-Anfrage

6.2 Ergebnis der Anfrage

Es handelt sich hier um ein Muster, das Dokument ist daher unvollständig...

7 Interpretation und Zusammenfassung

Es handelt sich hier um ein Muster, das Dokument ist daher unvollständig...