Attribute Exploration on the Web

Robert Jäschke\textsuperscript{1}
Sebastian Rudolph\textsuperscript{2}

\textsuperscript{1}L3S Research Center
\textsuperscript{2}Technische Universität Dresden

ICFCA 2013
Agenda

1. Motivation

2. Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3. Implementation

4. Conclusion + Discussion
Agenda

1. Motivation

2. Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3. Implementation

4. Conclusion + Discussion
## Motivation

![Concept Explorer](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show arrow rel...</td>
<td>don't show</td>
</tr>
<tr>
<td>Compressed</td>
<td></td>
</tr>
<tr>
<td>Object count</td>
<td>3</td>
</tr>
<tr>
<td>Attribute count</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATO</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euro</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schengen</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Re...</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

The Concept Explorer is a tool for exploring attributes on the Web. It allows users to visualize and analyze the relationships between different attributes, such as NATO, EU, Euro, and Schengen. The tool provides a clear dependent update feature, which helps in understanding the dependencies between different attributes.
Motivation

Is it true, that all objects have attribute(s) NATO, Schengen?

Yes  No  Stop Attribute Exploration
Motivation

Is it true, that when object has attribute(s) Euro, that it also has attribute(s) Schengen?

Yes  No  Stop Attribute Exploration
Motivation

How can we support attribute exploration on the web?

- typically, an expert performs attribute exploration
- a huge amount of information is available on the web
- (and many experts, too)
- How can we leverage this knowledge to support the expert?
  - by finding counterexamples
  - by providing context

How can we support attribute exploration on the web?
Attribute Exploration

Basic terms:

- formal context $\mathbb{K} := (G, M, I)$
- an implication $L \rightarrow R$ (for $L, R \subseteq M$) holds in $\mathbb{K}$, iff each object having all attributes from $L$ also has all attributes from $R$
- an object $c$ refutes an implication $L \rightarrow R$, iff $L \subseteq c'$ but $R \nsubseteq c'$ (i.e., there is at least one $m \in R$ with $m \notin c'$).

Attribute Exploration is then:

- computing a set of implications that hold for all objects
- for each new implication $L \rightarrow R$, ask the expert whether $L \rightarrow R$ holds (then add $L \rightarrow R$ to the set of implications), or $L \rightarrow R$ is refuted by a counterexample $c$ (then extend $\mathbb{K}$ with $c$)
Agenda

1. Motivation

2. Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3. Implementation

4. Conclusion + Discussion
Approach

1. Take an implication and transform it into a query or a set of queries.
2. Pose the queries to the system and show the expert the result.
3. Let the expert decide if the implication holds or not.

The web is large, there are many ways to query it.
→ We focus on: social question answering, crowdsourcing, linked open data, web search.

How to query?
Agenda

1. Motivation

2. Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3. Implementation

4. Conclusion + Discussion
Abstract Query Strategy

**we assume:** most information that we can retrieve from the web is *factual*, i.e., information about objects and their attributes

**however:** implications are *terminological* knowledge, which is more rare and harder to retrieve

**our approach:** querying for counterexamples, i.e., factual knowledge
Abstract Query Strategy

- $\varphi[x]$: query for an instance in a specific query language
- $\text{Ans}(\varphi[x])$: set of objects $d$ for which $\varphi[d]$ is true
- query $q[x]$ for a counterexample of $L \rightarrow R$:

$$q[x] := \bigwedge_{l \in L} \varphi_l[x] \land \bigvee_{r \in R} \neg \varphi_r[x]$$

- alternatively: $Q = \{q_r[x] \mid r \in R\}$ with

$$q_r[x] := \bigwedge_{l \in L} \varphi_l[x] \land \neg \varphi_r[x]$$

then

$$\text{Ans}(q[x]) = \bigcup_{q_r \in Q} \text{Ans}(q_r[x])$$
Agenda

1 Motivation

2 Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3 Implementation

4 Conclusion + Discussion
Social Question Answering/Crowdsourcing

Idea: forward the query to other experts

- social question answering (e.g., StackExchange, Yahoo! Answers)

Challenge for all cases: how to ensure correctness?
Social Question Answering/Crowdsourcing

Idea: forward the query to other experts

- crowdsourcing (e.g., Amazon Mechanical Turk)

Challenge for all cases: how to ensure correctness?
Idea: forward the query to other experts

- other approaches (e.g., e-mail, (micro-)blog)

Challenge for all cases: how to ensure correctness?
Agenda

1 Motivation

2 Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3 Implementation

4 Conclusion + Discussion
Linked Open Data Cloud

- part of an effort to build the *Semantic Web*
- large knowledge bases like YAGO or DBpedia
- typically provide a *SPARQL* interface to query
- *limitation*: available knowledge still very limited
- → details in the paper

[SPARQL Explorer for http://dbpedia.org/sparql]

```
SELECT DISTINCT ?country WHERE {
  ?country dcs:subject dbc:European_countries .
  OPTIONAL {
    ?y dcs:subject dbc:Member_states_of_NATO .
    FILTER (?country = ?y) .
  } FILTER (!BOUND(?y))
} UNION
```

**SPARQL results:**

<table>
<thead>
<tr>
<th>country</th>
</tr>
</thead>
<tbody>
<tr>
<td>:Cyprus</td>
</tr>
<tr>
<td>:Georgia_(country)</td>
</tr>
<tr>
<td>:Turkey</td>
</tr>
</tbody>
</table>
Agenda

1. Motivation

2. Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3. Implementation

4. Conclusion + Discussion
Assumptions:

1. A significant share of web documents describe singular objects.
2. For every attribute there is a search term, the presence of which in a web document is a good indicator for the described object having the attribute.
3. Likewise, the absence of this search term is a good indicator for the object not having the attribute.
Given an implication $L \rightarrow R$, we pose for each $r \in R$ the query

$$q_r[x] := +l_1 \ldots +l_{|L|} \neg r$$

and show the expert the resulting list of web pages.

Extensions:

- add domain $d$ of objects: $q_r[x] := +d +l_1 \ldots +l_{|L|} \neg r$
- restrict to site $s$: $q_r[x] := +\text{site}: s +l_1 \ldots +l_{|L|} \neg r$
Agenda

1 Motivation

2 Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3 Implementation

4 Conclusion + Discussion
Implementation

- Java-based prototype web application
- using FCAlib implementation of Next Closure algorithm
- posing queries to the Microsoft Bing API

prototype
http://greymane.l3s.uni-hannover.de:8888/

source code
https://github.com/rjoberon/web-attribute-exploration
Implementation: Example

<table>
<thead>
<tr>
<th></th>
<th>NATO</th>
<th>EU</th>
<th>Euro</th>
<th>Schengen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech Republic</td>
<td>×</td>
<td>×</td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>Norway</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>Germany</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
</tbody>
</table>

first implication is $\emptyset \rightarrow \{\text{NATO, Schengen}\}$

- raw queries:
  - $\neg\text{NATO}$
  - $\neg\text{Schengen}$
- but: plain negation not supported in web search
- hence, we add the object domain “Countries in Europe”:
  - $"\text{Countries in Europe}" \neg\text{NATO}$
  - $"\text{Countries in Europe}" \neg\text{Schengen}$
Implementation: Example

1. Countries of Europe - Geography Home Page - Geography at About.com
   http://geography.about.com/library/maps/bireurope.htm
   The countries of Europe.: city geography europe maps serbia and montenegro bosnia and herzegovina countries of europe

2. European Countries
   http://www.aneki.com/europe.html
   List of Countries in Europe with maps, statistics, and country comparisons of all the European Nations

3. Countries in Europe - YouTube
   http://www.youtube.com/watch?v=k2AMN0L-rfk
   This video includes all the countries in Europe except Vatican, Monaco, San Marino, Andorra, Liechtenstein, Malta. I hope it will be helpful. The song is aria ...

4. ESL - English Exercises: Countries in Europe
   http://www.englishexercises.org/exercise.asp?id=8524
   Fullscreen: Countries in Europe by robirimini. Date: 30 - Sep - 2012 Level: elementary Age: 8-11. Description: Three exercises about Italy, Germany, France, Spain ...

5. List of countries in EUROPE, list of countries of EUROPE, European ...
   Europe is unique continent which is not surrounded by water from all directions, and has overland border with the neighbouring continent Asia. Definition of correct ...

6. Pages Europe, Europe Yellow Pages, Europe Business Directory ...
   http://pageseurope.com/
   ... Europe Hotels, All Inclusive Europe, Europe Real Estate, Travel to Europe, European Vacations, jobs in Europe, Map of Europe, European Football, Countries in Europe

7. Countries in EUROPE - Cost of Living City to City
   http://www.earthcosts.com/countries-europe/
   New posts: Hot thread with new posts: No new posts: Hot thread with no new posts: Thread is closed

8. The Largest Countries in Europe | Infoplease.com
   http://www.infoplease.com/askeds/argest-countries-europe.html
   The Question: According to your info Sweden is the fourth largest country in Europe. Which are the three bigger ones?

9. How Many Countries are there in Europe? - Buzzle
   Europe is one of the most beautiful continents in the world, with a variety of countries and cultures in its fold. Wondering how many countries are there ...

10. Richest Countries In Europe
    http://www.aneki.com/europe_richest.html
    List of the richest countries in Europe ... Plutocrats: The Rise of the New Global Super-Rich and the Fall of Everyone Else

→ add restriction to site en.wikipedia.org:
+"Countries in Europe" +site:en.wikipedia.org -NATO
+"Countries in Europe" +site:en.wikipedia.org -Schengen
Web-Based Attribute Exploration

Formal Context

<table>
<thead>
<tr>
<th>Countries in Europe</th>
<th>NATO</th>
<th>EU</th>
<th>Euro</th>
<th>Schengen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech Republic</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Norway</td>
<td>x</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

change context

Attribute Exploration

The current implication is: \([\emptyset] \Rightarrow [\text{Schengen, NATO}]\).

You can either accept it or provide a counterexample:

Can you find a counterexample within the following web search results?

1. +"Countries in Europe" -"Schengen" site:en.wikipedia.org
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org

custom search

1. Former countries in Europe after 1815 - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Former_countries_in_Europe_after_1815
   This article gives a detailed listing of all the countries, (including puppet states), that have existed in Europe since the Congress of Vienna in 1815 to the present ...
2. Category:Former countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Former_countries_in_Europe
   Former countries in Europe after 1815; List of early East Slavic states; List of East Slavic duchies; List of historic states of Germany; List of historic states of Italy
3. Category:Countries in Europe - Wikipedia, the free encyclopedia
   http://en.wikipedia.org/wiki/Category:Countries_in_Europe
   Subcategories. This category has the following 58 subcategories, out of 58 total.
4. List of national capitals of countries in Europe by area ...
   This list includes the capitals of European countries for area. The list is below; Bank City Country...
Web-Based Attribute Exploration

Formal Context

<table>
<thead>
<tr>
<th>Countries in Europe</th>
<th>NATO</th>
<th>EU</th>
<th>Euro</th>
<th>Schengen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Czech Republic</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Norway</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Attribute Exploration

The current implication is: \( \emptyset \Rightarrow \text{[Schengen, NATO]} \).

You can either accept it or provide a counterexample:

<table>
<thead>
<tr>
<th>NATO</th>
<th>EU</th>
<th>Euro</th>
<th>Schengen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can you find a counterexample within the following web search results?

1. +"Countries in Europe" -"Schengen" site:en.wikipedia.org
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org
Implementation: Example

Can you find a counterexample within the following web search results?

1. +"Countries in Europe" -"Schengen" site:en.wikipedia.org
2. +"Countries in Europe" -"NATO" site:en.wikipedia.org

"Countries in Europe" -"Schengen" site:en.wikipedia.org

custom search

1. Former countries in Europe after 1815 - Wikipedia, the free ...  
   http://en.wikipedia.org/wiki/Former_countries_in_Europe_after_1815  
   This article gives a detailed listing of all the countries, (including puppet states), that have existed in Europe since the Congress of Vienna in 1815 to the present ...
2. Category:Former countries in Europe - Wikipedia, the free encyclopedia  
   http://en.wikipedia.org/wiki/Category:Former_countries_in_Europe  
   Former countries in Europe after 1815; List of early East Slavic states; List of East Slavic duchies; List of historic states of Germany; List of historic states of Italy
3. Category:Countries in Europe - Wikipedia, the free encyclopedia  
   http://en.wikipedia.org/wiki/Category:Countries_in_Europe  
   Subcategories. This category has the following 58 subcategories, out of 58 total.
4. List of national capitals of countries in Europe by area ...  
   This list includes the capitals of European countries by area. The chart is below. Rank City Country Area (km 2) 1. Ankara Turkey 7003251600000000000 2,516 ...
5. List of sovereign states and dependent territories by population ...  
   http://en.wikipedia.org/wiki/Largest_countries_in_Europe  
   This list includes the largest countries in Europe by population and area. The countries are listed in descending order of population. Rank Country ...
Why are no countries among the top results?

- matching only against *textual content*
  → domain restriction useless
- increased importance if search terms occur in page title
- *ranking by PageRank* prefers pages with many incoming links

Why are countries missing that do constitute a counterexample

Why are countries returned that do not constitute a counterexample?

→ limited validity of assumptions 2 and 3:

- web pages *do not* always contain search terms corresponding to attributes that the objects they describe *do* have
- web pages *do* sometimes mention terms corresponding to attributes that the objects *do not* have
Agenda

1 Motivation

2 Approach
   - Abstract Query Strategy
   - Social Question Answering/Crowdsourcing
   - Linked Open Data Cloud
   - Web Search Engines

3 Implementation

4 Conclusion + Discussion
Conclusion

- finding counterexamples is difficult, often even for a human expert
- presented approach has limited applicability for web search
- could be useful for exploring linked open data (or finding errors)
- general problem of the open world assumption
Conclusion

Some of the problems could be solved by

- providing more information about objects + attributes
- combining sources (1. linked open data, 2. web search, 3. crowdsourcing/social question answering)
- deeper analysis of retrieved information
- waiting for better knowledge bases ;-)}
Discussion

1 Motivation
2 Approach
   1 Abstract Query Strategy
   2 Social Question Answering/Crowdsourcing
   3 Linked Open Data Cloud
   4 Web Search Engines
3 Implementation
4 Conclusion + Discussion

Thank you for your attention!
Now it is time for discussion.

On the Web

Sebastian http://www.inf.tu-dresden.de/?node_id=3383
Robert http://www.kbs.uni-hannover.de/~jaeschke/
prototype http://greymane.l3s.uni-hannover.de:8888/
source code http://github.com/rjoberon/web-attribute-exploration