Scientific Profiling based on Semantic Analysis in Social Networks

Laurens De Vocht

Supervisors
Gonzalo Parra
Selver Softic

Promotors
Erik Duval
Martin Ebner

April 11, 2011
Agenda

- Problem Statement
- Solution
- Architecture
- Implementation
- Evaluation
- Conclusion
- Discussion
Problem Statement

Web users generate a massive unstructured data flow

Social Media Landscape

Who has scientific information relevant for me?
Solution

Annotate Data from Social Networks

Community approved ontologies: FOAF, SIOC

Linked Open Data

Connect People and Resources that share Scientific Affinities
Architecture

- Overview
- Grabeeter
- Framework
- API
Architecture: Overview
Architecture: Overview

Social Networks

Linked Open Data Cloud

Output Format

Framework

Aggregate

Interlink

Publish

Linked Data

Annotate

Analyse

Scientific Information

Archived/Cached Data

Twitter

DBPedia
Colinda
GeoNames

Grabeeter

Semantic Profiling Network

JSON
RDF (XML)

Scientific Profiling API

Annotate

Interlink

Publish
Architecture: Grabeeter
= Twitter aggregation & archiving tool
(developed at TUGraz)
Architecture: Grabeeter

= Twitter aggregation & archiving tool
(developed at TUGraz)
Architecture: Overview

Social Networks

Linked Open Data Cloud

Output Format

Annotate

Interlink

Aggregated/Cached Data

Linked Data

Analyse

Publish

Framework

Twitter

DBPedia

Colinda

GeoNames

Grabeeter

Semantic Profiling Network

JSON

RDF (XML)

Scientific Profiling API

Scientific Information

Aggregate

Interlink

Annotate

Analyse

Publish
Architecture: Framework

Applications

Programming Interface

Analysis

High Level Queries

Extraction

Interlinking

SQL Queries

Triplification

SPARQL Queries

Grabeeter

RDF Store
Architecture: API

- **User Profile**
  
  http://api.semanticprofiling.net/profile.php?user=screen_name

- **Discover people, events...**
  
  http://api.semanticprofiling.net/discovery.php?
  
  find=persons|events|popular_friends|popular_mentions|popular_events
  
  user=screen_name

- **Register new Twitter user**
  
  http://api.semanticprofiling.net/register.php?user=twitter_user

- **Event Details**
  
  http://api.semanticprofiling.net/event.php?name=event_name
Architecture: API

<table>
<thead>
<tr>
<th>Viewer</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="http://api.twitter.com/1/users/show.json?user_id=14301066" alt="JSON" /></td>
<td></td>
</tr>
</tbody>
</table>
- name: "Selver Sofic"
- screen_name: "selvers"
- location: ![http://ws.geonames.org/searchJSON?q=Austria](http://ws.geonames.org/searchJSON?q=Austria)
- image: ![http://a0.twimg.com/profile_images/905118560/f9e4b6eba.13070201_3_normal.jpg](http://a0.twimg.com/profile_images/905118560/f9e4b6eba.13070201_3_normal.jpg)
- description: "Semantic Web, Linked Data, Social Media, Reputation Systems enthusiast and researcher. Phd Student @ Graz University of Technology & a very...
- friends
- mentions
- scientific_events
  - SWAP 2010 : 6th Workshop on Semantic Web Applications and Perspectives : 2
  - iKnow 2010 : 10th International Conference on Knowledge Management and Knowledge Technologies : 1
  - WWWJ-QDW 2010 : WWWJ SI Querying the Data Web : 1
  - ESWC Inductive & Probabilistic 2011 : CFP for ESWC'2011 special track on Inductive and Probabilistic Approaches : 1
- interests
  - linkeddata : 10
  - reputationsystem : 6
  - readcast : 6
  - RDF : 2
  - SemWeb : 2
  - Il3 : 2
  - LinkedData : 2
  - drupalcampcph : 2
  - HTTP : 2
  - semanticweb : 2
  - web3 : 2
  - twitteraddict : 2
  - semweb : 2
  - google2010 : 2
Architecture: API
Implementation

› Hashtags as Identifiers
  › not always strong or consistent enough
  › properties of good hashtags formalized
  › helpful in assessment of valuable identifiers

› Expert Search/Profiling with Linked Data
  › aggregate and analyze certain types of data
  › need to surpass limits of closed data sets
  › LOD delivers multi-purpose data

References
Evaluation

‣ Approach
‣ Usability Case Study
‣ Search Quality
‣ Results so far
Evaluation: Approach

- Test **usability** & **search quality**
- Web application: “Scientific Affinity Browser” as case study.
- Development in 3 phases:
  - 1. User Interface only
  - 2. Fixed Data Set
  - 3. Integrated with Framework
Evaluation: Usability Case Study

Connecting researchers based on shared scientific events (conferences)
Evaluation: Search quality

- **Satisfaction** questionnaire
  Users rate their search results
- **Measure** **Relevance**
  - Recall
  - Precision
Evaluation: Results so far

- Definitely useful application
- Use of the map view makes sense
- People - Event split confusing
- View of own profile
  - not a suitable starting point
  - only useful in comparison
  - shouldn’t be always visible
- Person-specific affinities
  - too much hidden
Evaluation: Results so far
Challenges

› **Rank tags**

by importance, not just frequency of use

› **Visualization**

of similarities and links between users and entities

› **Detect communities**

using a graph-based algorithm on the linked data network
## Schedule

<table>
<thead>
<tr>
<th>Iteration 8</th>
<th>feb 21 - mar 2</th>
<th>Implement framework feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration 9</td>
<td>mar 3 - mar 16</td>
<td>Literature study 2</td>
</tr>
<tr>
<td>Iteration 10</td>
<td>mar 17 - apr 8</td>
<td>Evaluation phase 1 &amp; Report/presentation</td>
</tr>
<tr>
<td>Iteration 11</td>
<td>apr 9 - apr 28</td>
<td>Evaluation phase 2</td>
</tr>
<tr>
<td>Iteration 12</td>
<td>apr 29 - may 15</td>
<td>Evaluation phase 3</td>
</tr>
</tbody>
</table>
Conclusion

- Framework designed to support a broad range of social scientific semantic-based applications
- Realized with current state of the art technologies
- Interlinking with Linked Open Data Cloud
- Evaluation case study: scientific affinity browser web app
- Follow project blog http://blog.semanticprofiling.net
- All References: see reports