User Interests Driven Web Personalization based on Multiple Social Networks

Yi Zeng, Ning Zhong, Xu Ren, Yan Wang

International WIC Institute, Beijing University of Technology
P.R. China
Semantic Data at Web Scale

From large scale Web pages to large scale linked open semantic data

March, 2010: 13 Billion RDF Triples
October, 2011: 31.6 Billion RDF Triples
June, 2011: 12 Billion RDF Triples from the Web

The Large Knowledge Collider (LarKC) Project

11 Countries、13 Research Institutions and Universities

AstraZeneca
SIEMENS
cycorp
HLRlS
ontotex
Saltlux
CEFRIEL

Personalization for Large scale and Web Enabled Semantic Data Processing (cont.)

An illustration of the basic idea:

For more details:


Personalization for Large scale and Web Enabled Semantic Data Processing (cont.)

A Comparative Study of Query Time and Efficiency for Different Strategies

SwetoDBLP dataset: $1.49 \times 10^7$ RDF Triples

Participants 7 DBLP authors:

- Preference order 100%: List 2, List 3 $\approx$ List 1
- Preference order 100%: List 2 $\approx$ List 3
- Preference order 83.3%: List 2 $>\text{List 3} \not\approx \text{List 1}$
- Preference order 16.7%: List 3 $>\text{List 2} \not\approx \text{List 1}$

See references in the previous page
Massive Semantic Data from the Social Web

- The social Web platforms and the microblog platforms adopt and benefit from semantic techniques.
- The semantic Web gets huge data from these Social Web platforms.

Cyber-Social Sensors

845 million active users
http://en.wikipedia.org/wiki/Facebook

LinkedIn

350 million users
- 300 million tweets per day
- 1.6 billion queries per day
http://en.wikipedia.org/wiki/Twitter

flickr

60 million users

- Interesting Places
- Interesting Events

Facebook

845 million active users
http://en.wikipedia.org/wiki/Facebook

- Friends
- Personal Notes
- Likes

- Following, Followers
- Real time personal information
- Interesting news

- From Web of Contents to Web of People
- Users play more and more important roles
Personal Interests Data Fusion Strategies

**Weighted Fusion Strategy:** \( I(i) = \sum_{n=1}^{m} w_n \times I(i)_n \)

- Average fusion strategy
  \[ w_n = 1 / n \]
  \[ w_1 + w_2 + \ldots + w_n = 1 \]

- Time-sensitive fusion strategy
  \[ w_1 : w_2 : \ldots : w_n = f_1 : f_2 : \ldots : f_n \]
  \[ w_1 + w_2 + \ldots + w_n = 1 \]

Slides 7-10 are from our following paper:
An Illustration of Multi-source Personal Interests Fusion

Evolution of Scientific Information Sharing
“Open Science” Challenges Journal Tradition with Web Collaboration

User: Frank van Harmelen

Data Source:

- Twitter
- Facebook
- LinkedIn

Top-K interests from different sources
- Some of the interests have overlaps among each other.
- Diversities among these Top-K interests are even more obvious.

A comparative study of interests from three single sources
An Illustration of Multi-source Personal Interests Fusion

Update frequency:
Twitter: $f_1=2.5$, Facebook: $f_2=0.2$, LinkedIn: $f_3=0.0004$ (per day)

Weighted Interests Fusion Function:
$$I(i) = 0.9258 \times I(i)_1 + 0.0741 \times I(i)_2 + 0.0001 \times I(i)_3$$

A comparative study of interests from a single source and multiple interests sources

**Average Fusion**: Twitter(7), Facebook(7), LinkedIn(2)

**Time Sensitive Fusion**:
1. Top-10 overlaps with Twitter;
2. Values are very close to the ones from Twitter, but entirely different;
3. No interests from Facebook and LinkedIn.
Interests Representation and Reasoning about Interests

Interests Representation using e-FOAF:interest

Frank van Harmelen is interested in RDF in a certain degree

A Fragment of AI Ontology

RDF representation of AI Ontology

Appeared on Frank van Harmelen's homepage, but not elsewhere.
Collaboration network is already too complex, but... Academic collaboration candidates not only appear on publication data, but also on many other social networking environment such as Twitter.

A Snapshot from Semantic Web Dog Food Affiliation Map

Data Sources:
Twitter Data, Semantic Web Dog Food data, Google Maps API
Twitter data acquisition to:

- Locate the end user;
- Find agents that the end user follows;
- User real time interests analysis;
- Locating followings and their interests
AAVRA: Data Acquisition from SWDF

Real time acquisition by SPARQL end point

```
SELECT DISTINCT $person $person_name $affiliation $affiliation_name
WHERE {
  $person a foaf:Person.
  $person foaf:name $person_name.
  $person foaf:made $InProceedings.
  $InProceedings foaf:maker $person_url.
  $person_url foaf:name "Frank van Harmelen".
  $person swrc:affiliation $affiliation.
  $affiliation foaf:name $affiliation_name
}
```

In previous work we have shown that the MapReduce framework for distributed computation can be deployed for highly scalable inference over RDF graphs under the RDF Schema semantics. Unfortunately, several key optimizations that enabled the scalable RDFs inference do not generalize to the richer OWL semantics. In this paper we analyze these problems, and we propose solutions to overcome them. Our solutions allow...
# AAVRA: Generating Levels of Recommendation

## Interpretations on different groups of data from SWDF and Twitter

<table>
<thead>
<tr>
<th>Interest Levels</th>
<th>Formula</th>
<th>Result Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Coauthor_{SWDF}(p,u) \land TFing(u,p) )</td>
<td>( T_1 )</td>
</tr>
<tr>
<td>2</td>
<td>( Coauthor_{SWDF}(p,u) \land \neg TFing(u,p) )</td>
<td>( T_2 )</td>
</tr>
<tr>
<td>3</td>
<td>( TFing(u,p) \land PCoauthor_{SWDF}(p,u) )</td>
<td>( T_3 )</td>
</tr>
<tr>
<td>4</td>
<td>( TFing(u,p) \land SIT(p,u,K) \land \neg SWDF(p) )</td>
<td>( T_4 )</td>
</tr>
<tr>
<td>5</td>
<td>( TFing(u,p) \land \neg SIT(p,u,K) \land \neg SWDF(p) )</td>
<td>( T_5 )</td>
</tr>
</tbody>
</table>
## AAVRA: Recommendation Results Analysis

<table>
<thead>
<tr>
<th>Interest Level</th>
<th>Recommendation Ratio(%)</th>
<th>Results Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.014</td>
<td>Paul Groth</td>
</tr>
<tr>
<td>2</td>
<td>0.210</td>
<td>Spyros Kotoulas(3), Jacopo Urbani(3), Eyal Oren(2), Henri Bal(2), Zharko Aleksovski(2), Zhisheng Huang(1),...</td>
</tr>
<tr>
<td>3</td>
<td>0.154</td>
<td>Kalina Bontcheva, Lynda Hardman, Peter Mika, Steffen Staab, Denny Vrandecic, Ivan Herman, Michael Hausenblas, ...</td>
</tr>
<tr>
<td>4</td>
<td>0.505</td>
<td>Stefano Bertolo, Dan Brickley, DERI Galway, Web Foundation, Ontotext AD...</td>
</tr>
</tbody>
</table>

Recommendation Ratio = Recommended Results / Candidate Space
Candidate Space: 7131 persons (SWDF+Twitter)
Calculation of $SIT(p,u,K)$, Top-10 interests, $K=1$

0.8835% candidates are recommended overall.
Active Academic Visit Recommendation: A Snapshot

The 3rd level of recommendation: \( TFing(u, p) \land PCoauthors_{SWDF}(p, u) \)

- University of Sheffield (Kalina Bontcheva)
- University of the West of England (Richard McClatchey)
A conservative estimate would be that it would take 10,000 triples just to describe each human, which gives us 100 trillion ($10^{14}$).
Thank You!