Generating and Summarizing Explanations for Linked Data

Rakebul Hasan
INRIA Sophia Antipolis, France
ESWC 2014
Context

• Linked Data is constantly created, copied, moved around, and integrated
  – History of data lifecycle to help users judge the trustworthiness of data

• Explanations
  – How data is obtained
  – Flow of information in data derivations
  – Understanding and trust

• “Oh, yeah?” button to support users in assessing the reliability of information encountered on the Web
  – “so how do I know I can trust this information?” - Tim Berners-Lee, Consistent User Interface, W3C Design Issues, 1997
Explanation and Linked Data

• How to explain Linked Data?
  – Data derivations are inherently distributed.

• Publish explanation metadata as Linked Data
  – Enables explanation for distributed data
    • Follow the links and retrieve explanation metadata for source data recursively
Ratio4TA

- How to represent explanation metadata?
- Ratio4TA ontology
  - An extension of W3C PROV-O
  - [http://ns.inria.fr/ratio4ta](http://ns.inria.fr/ratio4ta)
Explanación

Overwhelming
Summarizing Explanations

• Summarization
  – Salient, abstract, and coherent information
  – Filtering: a set of classes used in the reasoning

• Entry point to the full explanation

• Provide a means to filter information in large explanations

- Xiang Zhang, Gong Cheng, and Yuzhong Qu. 2007. Ontology summarization based on RDF sentence graph, WWW ’07
Ranking Measures

• **Salient RDF Statements**

\[ S_{SL}(i) = \theta_1 \times C_{DN}(subjectOf(i)) + \theta_2 \times C_{DN}(objectOf(i)) \]

• **Similar RDF Statements**

\[ S_{SM}(i, FL) = \theta_1 \times similarity_{node}(subjectOf(i), FL) \\
+ \theta_2 \times similarity_{node}(predicateOf(i), FL) \\
+ \theta_3 \times similarity_{node}(objectOf(i), FL) \]

Ranking Measures

- Abstract Statements

\[ S_{AB}(i) = \frac{1}{\text{level}(i)} \]
Re-Ranking Measures

- Subtree Weight in Proof Tree

\[ score_{ST}(i) = \frac{\sum_{j \in \text{subtree}(i)} \text{score}(j)}{\| \text{subtree}(i) \|} \]
Re-Ranking Measures

• Coherence

\[ i = \arg \max_{j \in RL \setminus S} (\lambda_1 \times \text{score}(j) + \lambda_2 \times \text{reward}(j, S)) \]

\[ \text{reward}(j, S) = 1 - \frac{\text{coherent}(S)}{\text{coherent}(S \cup j)} \]
Evaluation

• Query result and inferences from selected subsets of DBPedia and Geonames data.
  – RDFS type propagation
  – owl:sameAs
  – transitivity of gn:parentFeature

• Explanations in natural language
Evaluation

• A query
  – Scientists born in United Kingdom

• Query result with explanation
  – Bob because
    – Bob is a Computer Scientist
    – Computer Scientists are Scientists
    – Bob was born in London
    – London is part of England
    – England is part of United Kingdom
  – Rating the necessity of each explanation statements from scale of 1 to 5
Evaluation

Knowledge of RDF

- Yes
- No

Gender

- Female
- Male

Age

- 20-24
- 25-29
- 30-34
- 35-39
- 40-44
- 45-49
- 50-54
- 55-59

People
Analysis of ground truths

Cosine similarity

<table>
<thead>
<tr>
<th></th>
<th>Avg.</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without Filtering</td>
<td>0.836</td>
<td>0.048</td>
</tr>
<tr>
<td>With Filtering</td>
<td>0.835</td>
<td>0.065</td>
</tr>
</tbody>
</table>
Measure Combinations

• Ranking

\[ score_{SL}(i) = S_{SL}(i) \]
\[ score_{SL+AB}(i) = \lambda_1 \times S_{SL}(i) + \lambda_2 \times S_{AB}(i) \]
\[ score_{SL+SM}(i) = \lambda_1 \times S_{SL}(i) + \lambda_2 \times S_{SM}(i, FL) \]
\[ score_{SL+AB+SM}(i) = \lambda_1 \times S_{SL}(i) + \lambda_2 \times S_{AB}(i) + \lambda_3 \times S_{SM}(i, FL) \]

• Re-Ranking
  – Subtree weight: \( S_{ST} \)
  – Coherence: \( S_{CO} \)

• Sentence graph summarization: \( S_{SG} \)

• Parameters: weighted average with equal weights
Evaluating Rankings
Evaluating Summaries
Summary

• Explanation for Linked Data
  – Vocabulary for explanation metadata

• Summarizing explanations
  – High accuracy
  – outperforms the sentence graph based ontology summarization approach
Thank you

Data, code, survey data

http://ns.inria.fr/ratio4ta/sm/