An FCA Framework for Knowledge Discovery in SPARQL Query Answers

Melisachew Wudage Chekol
Amedeo Napoli

LORIA (INRIA, CNRS, and Université de Lorraine), France

October 21 – 23, 2013

Introduction

- SPARQL
  - a W3C recommended query language for RDF graphs.
  - query answers can be provided in different formats: TEXT, JSON, HTML, XML, RDF, CSV.
- Formal Concept Analysis (FCA)
  - used for knowledge discovery within data represented by means of objects and their attributes.
  - concept lattices can reveal hidden relations within data and can be used for organizing, classifying, and even mining data.

Formal Concept Analysis (FCA)

FCA can be used for:
- classification and organization of data, knowledge discovery,
- ontology completion, supporting bottom-up construction of ontologies,
- role assertion analysis,
- computing subsumption hierarchy of least common subsumers,
- exploring finite models,
- discovering formal concepts in the Semantic Web data,
- providing an entry point to a dataset using questions in a way that can be navigated.

FCA – Formal Context

A formal context is \((G, M, I)\) where:
- \(G\) = set of objects
- \(M\) = set of attributes
- \(I \subseteq G \times M\)

<table>
<thead>
<tr>
<th>(G)</th>
<th>(M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Table: A formal context.

FCA – Concept Lattice

<table>
<thead>
<tr>
<th>Concept Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Figure: A concept lattice.

Objective

- organization and classification of SPARQL answers using FCA.
- visualization of SPARQL answers using concept lattices.

Method

Given \(A \subseteq G\) and \(B \subseteq M\) of a formal context \((G, M, I)\)
- with a derivation operator \(\subseteq \):
  - \(A' = \{m \in M | \forall g \in A \exists (g, m) \in I\}\)
  - \(B' = \{g \in G | \forall m \in B \exists (g, m) \in I\}\)
- the pair \((A, B)\) is a formal concept if
  - \(A' = B'\) and \(B' \subseteq A\)
- a set of concepts ordered with the set inclusion relation form a concept lattice.

Example

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

Conclusion

- we provided an organization SPARQL query answers based on a concept lattice, that can be navigated for mining or retrieving specific patterns in query results w.r.t. user constraints.
- this work shows some of the benefits that FCA provides to the semantic web.

Future Directions

- Experimentation: compare of answer format generations (logarithmic scale)
- To investigate how well this approach scales, given the size of SPARQL query answers over large datasets.

References

Franz Baader and Fabio Delgrande.
A finite basis for the set of entailments holding in a finite model.

Matthew A. Parnell and Enrico Motta.
Extracting relevant questions to an RDF dataset using formal concept analysis.

Melissa Korfhage, Ernest Lecue, Yu Shang Tan, Sebastian Lien, Byeon H.K. Koo, and Jeou Fun Lee.
Formal concept discovery in semantic web data.

Basu Sircar.
A survey on how description logic ontologies benefit from FCA.
In CLA, volume 672, pages 2–21, 2010.