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 – 25, 2013

ISWC
Sydney, Australia

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
- $I \subseteq G \times M$

	m_1	m_2	m_3	m_4
g_1	×			×
g_2			×	×
g_3		×		
g_4			×	×

Table: A formal context.

FCA – Formal Concept

Given $A \subseteq G$ and $B \subseteq M$ of a formal context (G, M, I)

- with a derivation operator $()'$,

$$A' = \{m \in M \mid \forall g \in A, (g, m) \in I\}$$
$$B' = \{g \in G \mid \forall m \in B, (g, m) \in I\}$$

- the pair (A, B) is a formal concept if

$$A' = B \text{ and } B' = A.$$

- a set of concepts ordered with the set inclusion relation form a concept lattice.

FCA – Concept Lattice

	m_1	m_2	m_3	m_4
g_1	×			×
g_2			×	×
g_3		×		
g_4			×	×

Table: A formal context.

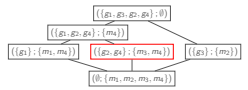


Figure: A concept lattice.

Objective

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

Method

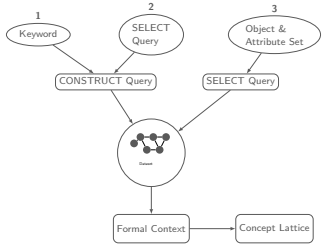
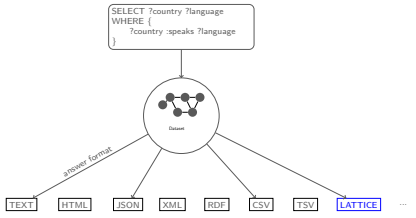
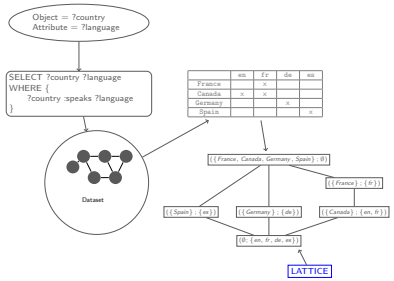


Figure: Architecture of SPARQL query answers organization.

Method



Example

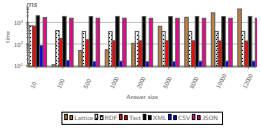


Conclusion

- 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: comparison 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 Felix Distel.
A finite basis for the set of el-implications holding in a finite model.
In ICFCA, pages 46–61. Springer, 2008.

Mathieu d'Aquin and Enrico Motta.
Extracting relevant questions to an RDF dataset using formal concept analysis.
In Proceedings of the sixth international conference on Knowledge capture, pages 121–128. ACM, 2011.

Markus Kirchberg, Erwin Leonardi, Yu Shyang Tan, Sebastian Link, Ryan KL Ko, and Bu Sung Lee.
Formal concept discovery in semantic web data.
In ICFCA, pages 164–179. Springer-Verlag, 2012.

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