Mining the Web of Data with Metaqueries

Francesca A. Lisi

University of Bari “Aldo Moro”
Department of Computer Science
Lab of Knowledge Acquisition and Machine Learning (LACAM)

FrancescaAlessandra.Lisi@uniba.it

ILP 2018
The Web of Data

- Feature: builds upon the WWW infrastructure to represent and interrelate data (aka Linked Data),
- Aim: transforming the Web from a distributed file system into a distributed database system.
- The foundational standards of the Web of Data include:
  - URI used to identify resources
  - RDF used to relate resources
RDF as a data model

- In RDF\(^a\) data is represented in the form of triples \(\langle \text{subject predicate object} \rangle\).
- The resulting collection of triples is a directed, labeled graph which can be accessed by posing SPARQL\(^b\) queries.
- The link between RDF and Description Logics (DLs) allows several entailment regimes for query answering in SPARQL.

\(^a\)https://www.w3.org/RDF/
\(^b\)https://www.w3.org/TR/rdf-sparql-query/
Knowledge graphs (KGs)

- Huge RDF graphs, see, e.g., DBpedia ([http://wiki.dbpedia.org/](http://wiki.dbpedia.org/))
- Automatically constructed by applying information extraction techniques

An example of KG [Tran et al., 2017]
The curation of KGs

- KGs are inherently *incomplete*.
- KGs particularly need to be curated by performing the task of *completion* (aka *link prediction*)
- Data mining algorithms can be exploited to automatically build rules able to make predictions on missing links.
New facts, e.g., $\text{livesIn}(alice, \text{berlin})$, $\text{livesIn}(dave, \text{chicago})$ and $\text{livesIn}(lucy, \text{amsterdam})$, can be derived from the following mined rule:

$$r_1 : \text{isMarriedTo}(x, y), \text{livesIn}(x, z) \Rightarrow \text{livesIn}(y, z)$$ (1)

and used to complete the KG.
Challenges of WoD Mining

- Size of KGs
- Open and distributed environment

Suggested solution (already sketched in [Lisi, 2017])

- Exploiting some useful meta-information about the KG in hand.
  - e.g., domains, ranges and confidence values of relations inside the KG (i.e., its schema)
- Adapting well-known data mining techniques that work at the meta-level
  - e.g., metaquerying [Ben-Eliyahu-Zohary and Gudes, 1999]
Metaquerying

- Technique for mining *frequent patterns* in relational databases
- A *metaquery* is a template that describes the type of pattern to be discovered in relational databases [Shen et al., 1996].
- Metaqueries are naturally expressed by means of a *second-order logic language*. 
Contribution of the paper

1 Proposal of a metaquerying approach to WoD mining
2 Definition of a metaquery language for WoD mining
   - based on second-order DLs, but
   - implementable with SPARQL.
3 Preliminary analysis of mechanisms for metaquery answering

An example of metaquery for WoD mining

\[ P(X, Y), Q(X, Z) \Rightarrow Q(Y, Z) \]
References I

Towards efficient metaquerying.

Towards a metaquery language for mining the web of data.

Metaqueries for data mining.

Towards nonmonotonic relational learning from knowledge graphs.