Discovering Subsumption Axioms with Concept Annotation

Pierre Monnin[1,✉], Amedeo Napoli[1] and Adrien Coulet[1,2]

[1] LORIA (CNRS, Inria Nancy-Grand Est, Université de Lorraine), Vandoeuvre-Lès-Nancy, France [2] BMIR (Stanford University), Stanford, USA

✉ pierre.monnin@loria.fr

Linked Open Data (LOD)

- Collection of inter-domain data sets
- Represented using Semantic Web Standards (RDF, URIs...)
- LOD resources represent entities of the real world (persons, organizations, places...)
- Triple: basic element of an RDF graph

$$\langle subject, predicate, object \rangle$$

- Predicates link resources to other resources, literals or classes of an ontology (**class instantiation**)

Ontology

- Formal representation of a particular domain
- Consists of classes and relationships between them
  $\mathcal{C}_\mathcal{O}$: set of classes of the ontology
- **Subsumption relationship**
  - Transitive relationship denoted by $\sqsubseteq$
  - $c \sqsubseteq d \Rightarrow$ every instance of c is an instance of d
- LOD resources can instantiate several classes from multiple ontologies
- **Type** of a resource $\mathtt{r}$: classes that $\mathtt{r}$ instantiates
  $\mathrm{type}(\mathtt{r}) = \{\mathtt{c} \in \mathcal{C}_\mathcal{O} \mid \langle \mathtt{r}, \mathtt{type}, \mathtt{c} \rangle\}$
- **Extended type** of a resource $\mathtt{r}$: takes into account the transitivity of $\sqsubseteq$
  $\mathrm{extdtype}(\mathtt{r}) = \mathrm{type}(\mathtt{r}) \cup \{\mathtt{d} \in \mathcal{C}_\mathcal{O} \mid \exists \mathtt{c} \in \mathrm{type}(\mathtt{r}),\ \mathtt{c} \sqsubseteq \mathtt{d}\}$

Motivation

- Ontologies and LOD data sets are of various quality and completeness
- Regularities in associated LOD data sets could be leveraged to complete ontologies
- Here, we aim at **discovering subsumption axioms** between classes of an ontology based on the **regularities** of associated LOD data sets between **LOD resources** and the **predicates** they are subject of

Method Summary

Table 1: An example of RDF triples

$\langle \mathtt{r}_1, \mathtt{type}, \mathtt{k}_1 \rangle$	$\langle \mathtt{r}_2, \mathtt{pred}_3, \mathtt{o}_5 \rangle$
$\langle \mathtt{r}_1, \mathtt{type}, \mathtt{k}_2 \rangle$	$\langle \mathtt{r}_3, \mathtt{type}, \mathtt{k}_1 \rangle$
$\langle \mathtt{r}_1, \mathtt{pred}_1, \mathtt{o}_1 \rangle$	$\langle \mathtt{r}_3, \mathtt{type}, \mathtt{k}_2 \rangle$
$\langle \mathtt{r}_1, \mathtt{pred}_2, \mathtt{o}_2 \rangle$	$\langle \mathtt{r}_3, \mathtt{pred}_1, \mathtt{o}_6 \rangle$
$\langle \mathtt{r}_2, \mathtt{type}, \mathtt{k}_1 \rangle$	$\langle \mathtt{r}_4, \mathtt{type}, \mathtt{k}_1 \rangle$
$\langle \mathtt{r}_2, \mathtt{type}, \mathtt{k}_2 \rangle$	$\langle \mathtt{r}_4, \mathtt{type}, \mathtt{k}_2 \rangle$
$\langle \mathtt{r}_2, \mathtt{type}, \mathtt{k}_4 \rangle$	$\langle \mathtt{r}_4, \mathtt{type}, \mathtt{k}_5 \rangle$
$\langle \mathtt{r}_2, \mathtt{type}, \mathtt{k}_5 \rangle$	$\langle \mathtt{r}_4, \mathtt{pred}_2, \mathtt{o}_7 \rangle$
$\langle \mathtt{r}_2, \mathtt{pred}_1, \mathtt{o}_3 \rangle$	$\langle \mathtt{r}_4, \mathtt{pred}_3, \mathtt{o}_8 \rangle$
$\langle \mathtt{r}_2, \mathtt{pred}_2, \mathtt{o}_4 \rangle$	

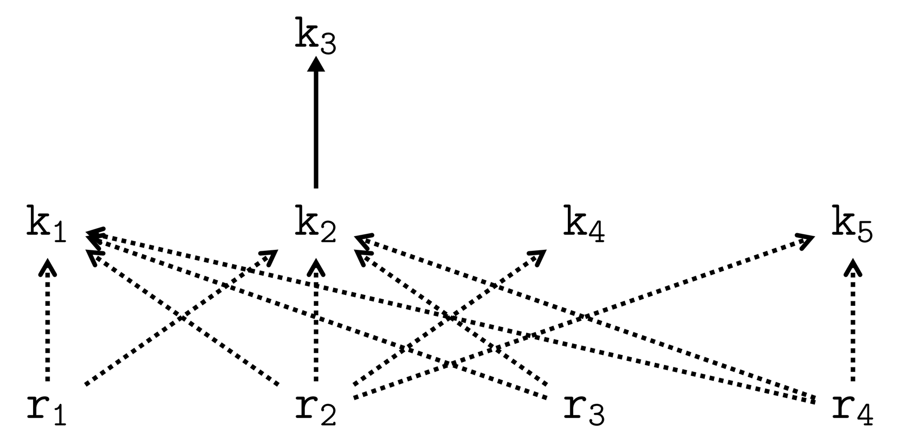

Figure 1: Example of ontology classes ($\mathtt{k}_1$, $\mathtt{k}_2$, $\mathtt{k}_3$, $\mathtt{k}_4$ and $\mathtt{k}_5$) being instantiated by resources ($\mathtt{r}_1$, $\mathtt{r}_2$, $\mathtt{r}_3$ and $\mathtt{r}_4$).

Table 2: Formal context generated from the RDF triples. A cross in the table indicates that there exists at least one triple where the subject (in row) and the predicate (in column) appear simultaneously.

	type	$\mathtt{pred}_1$	$\mathtt{pred}_2$	$\mathtt{pred}_3$
$\mathtt{r}_1$	×	×	×	
$\mathtt{r}_2$	×	×	×	×
$\mathtt{r}_3$	×	×		
$\mathtt{r}_4$	×		×	×

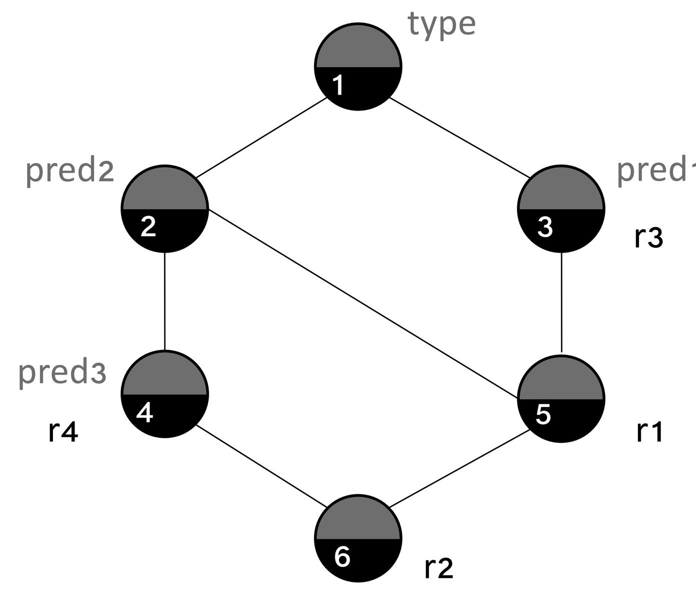

Figure 2: Line diagram representing the concept lattice built from the formal context. Concepts are displayed using the reduced labeling.

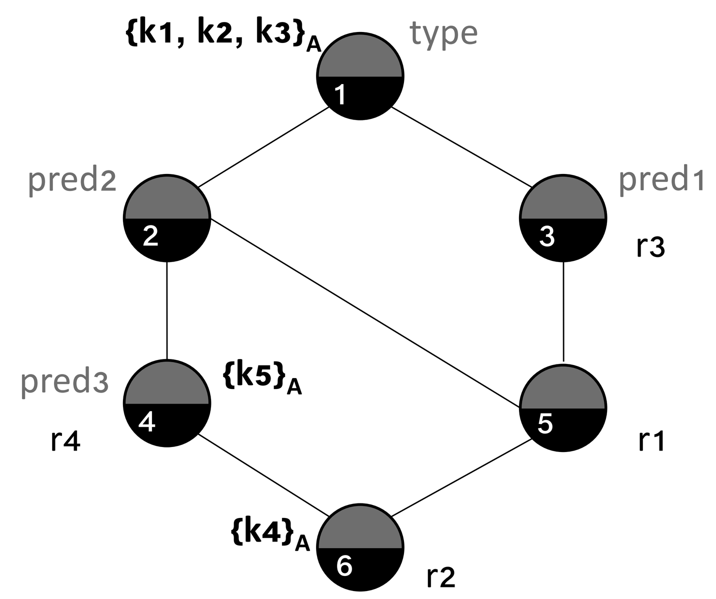

Figure 3: Line diagram representing the annotated lattice. Concepts are displayed using the reduced labeling.

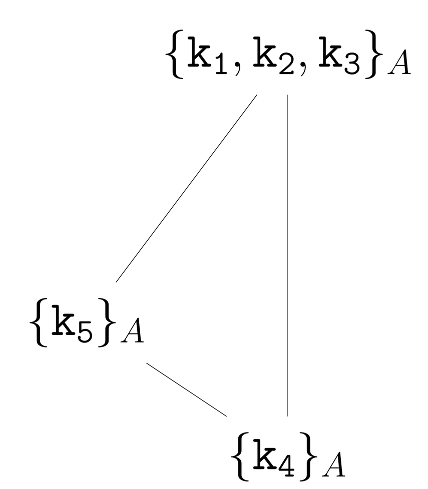

Figure 4: Line diagram representing the induced order on proper annotations from the annotated lattice.

Table 3: Discovered axioms from the induced order in Figure 4.

$\mathtt{k}_4 \sqsubseteq \mathtt{k}_1$	$\mathtt{k}_4 \sqsubseteq \mathtt{k}_3$	$\mathtt{k}_5 \sqsubseteq \mathtt{k}_1$	$\mathtt{k}_5 \sqsubseteq \mathtt{k}_3$
$\mathtt{k}_4 \sqsubseteq \mathtt{k}_2$	$\mathtt{k}_4 \sqsubseteq \mathtt{k}_5$	$\mathtt{k}_5 \sqsubseteq \mathtt{k}_2$	

Formal Concept Analysis (FCA)

- A **concept lattice** is generated from a **formal context**
- Each concept (A, B) of the lattice is formed by
  - *An extent* A containing LOD resources
  - *An intent* B containing the common predicates the resources in A are subject of
- The concept lattice highlights regularities between LOD resources and their predicates
- Ontology classes need to be associated with this structure to discover subsumption axioms

Concept Annotation

- Associates each formal concept (A, B) with an annotation $A^\diamond = \cap_{\mathtt{r} \in A} \mathrm{extdtype}(\mathtt{r})$
- $A^\diamond$ represents the common extended type of LOD resources in the extent
- Subsumption axioms can be directly discovered from the induced order on proper annotations
- Discovered axioms can be:
  - Already explicitly stated in the ontology
  - Not explicitly stated but inferable from the ontology
  - New axioms neither explicitly stated in nor inferable from the ontology

Results on DBpedia

- Experimentation on the DBpedia 2016-04 data set
- Ontologies: DBpedia Ontology and YAGO

Examples of discovered axioms:

- `yago:WikicatRedWineGrapeVarieties` $\sqsubseteq$ `yago:WikicatGrapeVarieties`
- `yago:WikicatTreesOfEcuador` $\sqsubseteq$ `yago:WikicatMedicinalPlants`

Acknowledgements

This work is supported by:

- the *PractiKPharma* project, founded by the French National Research Agency (ANR), Grant No: ANR-15-CE23-0028 `http://practikpharma.loria.fr`
- the *Snowball* Inria Associate Team `http://snowflake.loria.fr`