ECAI 2014
T. Schaub et al. (Eds.)

doi:10.3233/978-1-61499-419-0-979

Combining Reasoning on Semantic Web Metadata

Loris Bozzato and **Luciano Serafini**[1]

Abstract. As the amount of available linked data expand and the number of related applications increases, the management of aspects such as provenance and access control of such data begin to become an issue. Current approaches do not provide sufficient support for automatic reasoning over different metadata types and their possible interdependencies. MetaReasons is a framework that supports representation and automated reasoning over metadata in a single logical formalism. Different types of metadata, like data-provenance and accessibility-restrictions, are represented as distinct meta-theories and dependencies between metadata types are represented by rules between different meta-theories. In this paper we present the definition of the MetaReasons framework and two examples meta-theories for provenance and access control. Moreover, we propose a materialization calculus for forward reasoning on the two aspects.

1 INTRODUCTION

As the use of Semantic Web (SW) / LOD data continuously expands, a flexible management for different aspects of such data become more and more urgent: among these, we find aspects of provenance, access control, privacy and trust relative to the combinations of multiple datasets and sources of information. We note that such aspects consist of *meta-information* qualifying a dataset or relating each other different datasets: as such, data-sources can be read as individuals in a *meta-level* in the domain of different *meta-theories*.

Meta-reasoning takes clearly a major role in the management of such aspects: it allows to infer implicit relations among datasets or classify newly added datasets. The problem of representing and reasoning about single metadata information of datasets has already been considered in the area of SW / LOD data by multiple alternative frameworks (e.g. [3, 5]). However, we observe that meta-data about different aspects are dependent one with the other: e.g. accessibility and trust associated to a particular dataset might depend on its provenance. Meta-data, therefore, should not be considered as separate aspects in reasoning. We are thus interested in a *flexible combination of reasoning and querying* between the metalevel (i.e. about datasets) and the knowledge level (i.e. inside datasets). For this purpose we have to consider the relations between reasoning (and querying) *about* datasets in the metalevel and reasoning *inside* each dataset. Indeed, reasoning and query answering in single datasets can now be influenced by all of the reasoning at the meta level.

In this paper we provide a definition of a framework, called *MetaReasons (MR)*, for the representation and reasoning over metadata in an integrated metalevel structure[2]. MetaReasons is a two-layer framework, composed of an *object layer* containing the data-

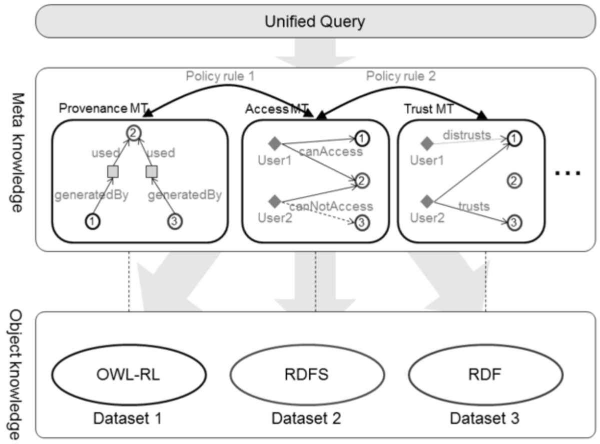

Figure 1. MetaReasons framework architecture

sources and a *meta-layer* containing knowledge about meta-data (Figure 1). This structure is similar to our previous works on the CKR contextual framework [6, 1]. We thus have a clear separation between the level representing the metainformation (*Metaknowledge*) associated to a set of information sources (viewed as atomic components) and the level (*Object knowledge*) respective to the actual dataset contents. Each of the *metatheories* in the upper level separately encodes one of the metadata aspects to be represented. Metatheories can use their own schema and possibly their local reasoning formalism, but their final representation should be reconduced to a single formalism, possibly encoded in OWL2. This will enable to adapt different existing models and readily available ontologies for different aspects under a single framework. This architecture implies that we can decouple reasoning on metaknowledge from the one on the object knowledge, thus allowing to use different formalisms on the two parts. Dependency axioms (i.e. *policies*) connecting metatheories can be defined in the part of the metaknowledge covering all of the single theories. On this architecture, *unified queries* over both the meta and the object knowledge can be expressed in standard SPARQL (possibly extended with the primitives defined by the architecture) exploiting the features of named graphs.

Complete formal definitions for MR, proofs and a detailed example are provided in the accompanying Technical Report [2].

2 MR ARCHITECTURE AND METATHEORIES

Unified Knowledge Bases. In order to uniformly represent and reason over meta- and object-level information, we want to model knowledge bases as two layered structures $\mathfrak{K} = \langle \mathfrak{M}, \mathfrak{D} \rangle$: we call such structures *Unified Knowledge Bases (UKB)*, since they offer an unified model to reason on both levels. Given a fixed set of datasets, with names in $\mathbf{N}$, the upper layer of an UKB is composed by a DL knowl-

[1] Fondazione Bruno Kessler, Via Sommarive 18, 38123 Trento, Italy. E-mail: {bozzato,serafini}@fbk.eu
[2] MR has been proposed inside PlanetData NoE (www.planet-data.eu) for reasoning with provenance, access control, privacy and trust of SW data.

edge base $\mathfrak{M}$ containing a set of *metatheories* MT_i, each one encoding a specific type of meta-information, and a set P of axioms, that we call *policies*, spanning over and linking elements of any metatheory. The lower layer of the UKB contains a set $\mathfrak{D}$ of *datasets* (i.e. DL knowledge bases) DS_n for each dataset name $n \in \mathbf{N}$ (with their local vocabulary and language). We do not fix the DL language(s) used in the UKB: however, in the following we consider $\mathcal{SROIQ}$-RL (a restriction of $\mathcal{SROIQ}$ to the form of OWL RL axioms) for the definition of our metatheories and calculus.

Interpretations for an UKB follow its two layered structure: they are structures $\mathfrak{I} = \langle \mathcal{M}, \mathcal{I} \rangle$ where, $\mathcal{M}$ is an interpretation for the whole metalevel KB (thus giving a single DL interpretation for all metatheories and policy KB) and each $\mathcal{I}(n)$ with $n \in \mathbf{N}$ is a DL interpretation for dataset DS_n. The notion of *model* and *entailment* w.r.t a UKB can be naturally extended to these structures.

Metatheories. We propose two metatheories that can be "imported" in the UKB architecture to encode models for provenance [3] and access control [5] of datasets.

These two models use a third metatheory for representing the notion of derivability across datasets. The idea is that the derivation of a fact from a set of facts having some meta-properties induces other meta-properties on the derived fact. For instance, the provenance of a fact α derived from facts β and γ depends on the provenance of β and γ. Thus, we maintain information on how facts are derived from other facts in the meta-theory for derivability MT_d. The only predicate of MT_d is $\mathsf{derivedFrom}$, a transitive role linking a dataset d_i with the set of datasets $d_1, \ldots, d_k$ containing the facts used to infer the facts in d_i. We impose the following semantic condition: all of the consequences derivable (at instance level) from the contents of $\mathrm{DS}_1, \ldots, \mathrm{DS}_k$ are contained in the interpretation of DS_i.

We encode the model for provenance from [3] in the meta-theory MT_{prov}. We adapt the original representation to extend the granularity from triples to datasets. By doing this we have to adapt the notion of composition of provenance information by inference: we do so by explicitly adding information about composition of datasets using MT_d. Intuitively, in the encoding, initial datasets are explicitly assigned to a Color (representing their information source) using a functional property $\mathsf{hasColor}$. On the other hand, if a dataset d_i is $\mathsf{derivedFrom}$ datasets d_1 and d_2, it is assigned a color that is the combination of colors of the two sources: the combined color is represented by an $\mathsf{hasDefiningColor}$ relation from d_i to d_1 and d_2 (corresponding to the combination operator $+$ in [3]).

The model we encode as our metatheory MT_{ac} for access control is presented in [5]. The approach is similar to previous model: the idea is to add abstract access control tokens to datasets in order to identify their accessibility properties. As above, the goal is to describe the propagation of access control information from explicit to inferred triples. Thus, initial datasets are assigned access control tokens in $\mathsf{AccessToken}$ (possibly using access control authorizations realized as SPARQL queries [5]) connected by a functional role $\mathsf{hasLabel}$. If a dataset is obtained as derivation of other datasets, then its label is the composition of the labels of defining sources: thus, if d_i is $\mathsf{derivedFrom}$ d_1 and d_2, we require that d_i $\mathsf{isComposedFrom}$ d_1 and d_2. That is, the role $\mathsf{isComposedFrom}$ is used to represent the combination operator $\odot$ of [5].

3 MATERIALIZATION CALCULUS

We propose a formalization for an inference method in MetaReasons: the method is a datalog-based calculus for instance checking in $\mathcal{SROIQ}$-RL based UKBs, corresponding to an extension to MR of the materialization calculus presented in [4] (following the extension to the CKR framework in [1]).

The calculus works by compiling the whole UKB to a single datalog program. The translation has three components: (i) *input translations* I_{rln}, I_{meta}, where given an axiom or signature symbol α and $n \in \mathbf{N}$, each $I(\alpha, n)$ is a set of datalog facts; (ii) *deduction rules* P_{rln}, $P_{der\text{-}obj}$, $P_{prov\text{-}meta}$, $P_{ac\text{-}meta}$, which are sets of datalog rules; (iii) *output translation* O, where given an axiom α and $n \in \mathbf{N}$, $O(\alpha, n)$ is a single datalog fact. Intuitively, $\mathcal{SROIQ}$-RL input I_{rln} and deduction P_{rln} rules provide the translation and interpretation of $\mathcal{SROIQ}$-RL axioms. I_{meta} encode conditions specifically needed for the translation of the metalevel layer. Rules in $P_{prov\text{-}meta}$ and $P_{ac\text{-}meta}$ correspond to the semantic conditions imposed on MT_{prov} and MT_{ac}, while $P_{der\text{-}obj}$ define the propagation of knowledge across "derived" datasets that is implied by $\mathsf{derivedFrom}$ assertions in MT_d. Rules in O provide the translation of "output" ABox assertions that can be proved by applying the rules of the final program.

To produce a program that represents the input UKB, we first translate $\mathfrak{M}$ in a *meta program* $PM(\mathfrak{M})$ using input rules I_{rln} and I_{meta} and including P_{rln}, $P_{prov\text{-}meta}$ and $P_{ac\text{-}meta}$. Then, for each dataset DS_n, we define the *local program* $PD(n)$ using I_{rln}, P_{rln} and $P_{der\text{-}obj}$. The *UKB program* $PK(\mathfrak{K})$ is encoded as the union of meta program and all local programs. We say that $\mathfrak{K}$ *entails* an axiom α in a dataset DS_n if $PK(\mathfrak{K}) \models O(\alpha, n)$. We can show that the translation provides a sound and complete calculus for instance checking in $\mathcal{SROIQ}$-RL UKBs.

4 CONCLUSIONS: FUTURE WORKS

In this paper we presented the first steps in the definition of the MetaReasons framework, an approach for reasoning with multiple metadata aspects over SW / LOD datasets.

One of the next steps in the framework development will consist of its implementation over tools for management of SW data. The presented materialization calculus will constitute the formal base for the implementation of a forward reasoning procedure over RDF data, similarly to our work for the CKR framework [1]. Another parallel activity concerns the definition of new metatheories covering different metalevel aspects. Such metatheories can be obtained by encoding known models or may be already available as OWL ontologies that can be readily "plugged in" our framework.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no.257641 (PlanetData NoE).

References

[1] L. Bozzato and L. Serafini, 'Materialization Calculus for Contexts in the Semantic Web', in *DL2013*, CEUR-WP. CEUR-WS.org, (2013).

[2] L. Bozzato and L. Serafini, 'Combining reasoning on semantic web metadata', Technical Report TR-FBK-DKM-2014-1, FBK, Trento, Italy, (2014). `http://dkm.fbk.eu/index.php/Resources`.

[3] G. Flouris, I. Fundulaki, P. Pediaditis, Y. Theoharis, and V. Christophides, 'Coloring RDF Triples to Capture Provenance', in *ISWC 2009*, volume 5823 of *LNCS*, pp. 196–212. Springer, (2009).

[4] M. Krötzsch, 'Efficient Inferencing for OWL EL', in *JELIA 2010*, volume 6341 of *LNCS*, pp. 234–246. Springer, (2010).

[5] V. Papakonstantinou, M. Michou, I. Fundulaki, G. Flouris, and G. Antoniou, 'Access control for RDF graphs using abstract models', in *SACMAT 2012*, pp. 103–112. ACM, (2012).

[6] L. Serafini and M. Homola, 'Contextualized knowledge repositories for the semantic web', *J. of Web Semantics*, **12**, (2012).