Discovering Expressive Rules for Complex Ontology Matching and Data Interlinking

Manuel Atencia[1], Jérôme David[1], Jérôme Euzenat[1], Liliana Ibanescu[2], Nathalie Pernelle[3], Fatiha Saïs[3], Élodie Thiéblin[4], and Cassia Trojahn[4]

[1] Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble, France
firstname.lastname@inria.fr
[2] UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France
firstname.lastname@agroparistech.fr
[3] LRI, Paris Sud University, CNRS 8623, Paris Saclay University, Orsay F-91405, France
firstname.lastname@lri.fr
[4] IRIT, UMR 5505, 1118 Route de Narbonne, F-31062 Toulouse, France
firstname.lastname@irit.fr

1 Introduction

Ontology matching and data interlinking as distinct tasks aim at facilitating the interoperability between different knowledge bases. Although the field has fully developed in the last years, most ontology matching works still focus on generating simple correspondences (e.g., $Author \equiv Writer$). These correspondences are however insufficient to fully cover the different types of heterogeneity between knowledge bases and complex correspondences are required (e.g., $LRIMember \equiv Researcher \sqcap \exists belongsToLab.\{LRI\}$). Few approaches have been proposed for generating complex alignments, focusing on correspondence patterns or exploiting common instances between the ontologies. Similarly, unsupervised data interlinking approaches (which do not require labelled samples) have recently been developed. One approach consists in discovering linking rules on unlabelled data, such as simple keys [2] (e.g., $\{lastName, lab\}$) or conditional keys [3] (e.g., $\{lastName\}$ under the condition $c = Researcher \sqcap \exists lab.\{LRI\}$). Results have shown that the more expressive the rules are, the higher the recall is. However naive approaches cannot be applied on large datasets. Existing approaches presuppose either that the data conform to the same ontology [2] or that all possible pairs of properties be examined [1]. Complementary, link keys are a set of pairs of properties that identify the instances of two classes of two RDF datasets [1] (e.g., $\{\langle creator, auteur\rangle, \langle title, titre\rangle\}$ linkkey $\langle Book, Livre\rangle$, expresses that instances of the $Book$ class which have the same values for properties $creator$ and $title$ as an instance of the $Livre$ class has for $auteur$ and $titre$ are the same). Such, link keys may be directly extracted without the need for an alignment.

2 Proposed approach

We introduce here an approach that aims at evaluating the impact of complex correspondences in the task of data interlinking established from the application of keys (Figure 1).Given two populated ontologies O_1 and O_2, we first apply the CANARD system [4]

for establishing complex correspondences (1). Then, the key discovery tools VICKEY [3] and LinkEx are applied for the discovery of simple keys, conditional keys, and link keys from the instances of O_1 and O_2, exploiting the complex correspondences as input (as a way of reducing the key search space) (2). The keys are then applied in the data interlinking task, which can also benefit from the complex correspondences (as a way of extending the sets of instances to be compared) (3). Finally, as CANARD considers shared instances, the matching is iterated by considering the detected identity links.

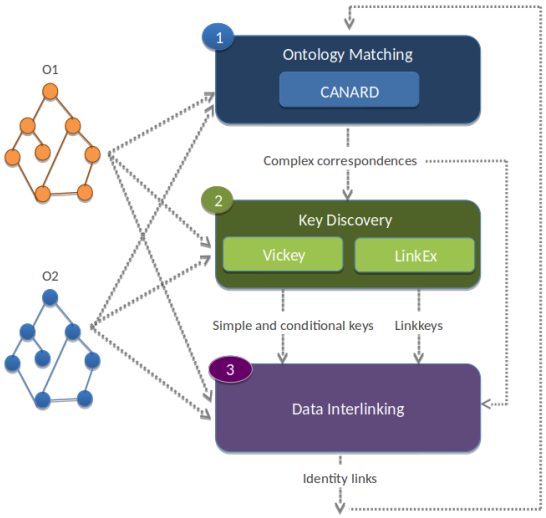

Fig. 1. Workflow of ontology matching and data interlinking enhanced by key discovery.

We plan to evaluate the approach to verify, on the one hand, whether the use of complex correspondences allows to improve the results of data interlinking. On the other hand, thanks to the use of the detected identity links, it would also be reasonable to expect improvements in ontology matching results. Experiments will be run on DBpedia and YAGO, covering different domains such as people, organizations, and locations, as there exists reference entity links or these datasets.

Acknowledgement. This work is supported by the CNRS Blanc project RegleX-LD.

References

1. M. Atencia, J. David, and J. Euzenat. Data interlinking through robust linkkey extraction. In *ECAI*, pages 15–20, 2014.
2. D. Symeonidou, V. Armant, N. Pernelle, and F. Saïs. Sakey: Scalable almost key discovery in rdf data. In *ISWC*, pages 33–49, 2014.
3. D. Symeonidou, L. Galárraga, N. Pernelle, F. Saïs, and F. M. Suchanek. VICKEY: mining conditional keys on knowledge bases. In *ISWC*, pages 661–677, 2017.
4. É. Thiéblin, O. Haemmerlé, and C. Trojahn. CANARD complex matching system: results of the 2018 OAEI evaluation campaign. In *OM@ISWC*, pages 138–143, 2018.