Special issue on ontology and linked data matching

Michelle Cheatham, Isabel Cruz, Jérôme Euzenat, Catia Pesquita

Special Issue on Ontology and Linked Data Matching

Editorial [1]

Michelle Cheatham [a], Isabel Cruz [b], Jérôme Euzenat [c] and Catia Pesquita [d,**]

[a] *DaSe Laboratory, Wright State University, 3640 Colonel Glenn Hwy., Dayton, OH 45435, USA*
E-mail: michelle.cheatham@wright.edu

[b] *Department of Computer Science, University of Illinois at Chicago, 851 S Morgan St., Chicago, IL 60607, USA*
E-mail: ifc@cs.uic.edu

[c] *Univ. Grenoble Alpes, INRIA, CNRS, LIG, F-38000 Grenoble, France*
E-mail: Jerome.Euzenat@inria.fr

[d] *LaSIGE, Faculdade de Ciências, Universidade de Lisboa, Campo Grande 1749-016 Lisboa, Portugal*
E-mail: clpesquita@fc.ul.pt

1. Introduction

Semantic web technologies break down many of the barriers to leveraging the large amount of data and information that has been collected or created. The use of unique identifiers, transport protocols like HTTP, and uniform data description languages like RDF go a considerable way towards providing seamless access to this data. Consequently, the semantic web has grown with the continual creation of new ontologies and linked data covering a wide variety of domains, and applications and analytical techniques using this data have been created. However, while physical data silos have waned, the lack of semantic links between ontologies and linked datasets, supports, in effect, invisible virtual silos preventing these resources from being queried, browsed, or leveraged in a truly uniform way. If such links could be generated in a reliable and scalable way, the network effect would greatly increase the utility of these resources. It is for this reason that the topic of ontology and linked data matching is both important and timely.

Ontology and linked data matching has been an active area of research for over a decade now [4], and related fields such as database schema alignment [2] and coreference resolution in structured and semi-structured text [3,5] have received significant attention for even longer. This work has generated many successes. In the annual Ontology Alignment Evaluation Initiative, the average and best performance of ontology alignment systems on these benchmarks has generally increased from year to year [1]. Top alignment systems such as AgreementMakerLight and LogMap are able to generate coherent and consistent alignments with an F-measure as high as 0.94 on some tasks.

More work remains to be done to address all of the challenges inherent in semantic data integration, however. For example, current matching systems tend to perform well on finding one-to-one equivalence relations between classes, but they are often less effective at finding other types of relationships between other types of entities. Succeeding in these areas may require different similarity metrics, filtering techniques, or other additions to current methods. Problems can also arise when the data to be aligned lacks a significant T-box or A-box, or when ontologies or linked data sets are very large. Further complicating things, desirable matches may depend on the context in which they

[**]Corresponding author. E-mail: cpesquita@di.fc.ul.pt.

will be used. Ontology and linked data matching systems can also be difficult to use – how to set parameters required by a matching algorithm is often not clear, the reasons behind the matches they generate (or the implications of those matches) is sometimes not immediately evident to users, and matching algorithms can not always incorporate user feedback to improve the quality of the matches they generate. These pending research issues, among many others, were the impetus for this special issue of the Semantic Web Journal.

The three papers in this special issue address various aspects of the matching problem. For instance, "On the Efficient Execution of Bounded Jaro-Winkler Distances" by Dreßler and Ngonga Ngomo points out that most ontology and linked data matching approaches involve a syntactic comparison of entity labels at some point in the process. If the datasets to be linked are very large, this creates a bottleneck. They therefore present several different approaches to quickly filter out target strings that cannot possibly have a Jaro-Winkler similarity greater than a given threshold to a given string. This lossless approach pushes the state-of-the-art in this area, even on small datasets.

The paper "An Unsupervised Data-driven Method to Discover Equivalent Relations in Large Linked Datasets" by Zhang, Gentile, Blomqvist, Augenstein, and Ciravegna takes on the problem of finding equivalent relations (i.e. properties) either within a single dataset or, potentially, across two or more distinct datasets. Their extensional matching approach is entirely unsupervised. Interestingly, it does not require the user to specify a threshold value – the appropriate threshold is automatically determined by the algorithm on a concept by concept basis using an unsupervised clustering algorithm. The performance of the proposed approach was shown to result in a significantly higher F-measure than even supervised baseline models.

The focus of "A Session-based Ontology Alignment Approach enabling User Involvement" by Lambrix and Kaliyaperumal is on effectively and efficiently involving the user in the matching process. The paper describes a "session-based" ontology alignment system that allows a user to provide feedback on a suggested partial mapping. This feedback is used immediately to improve the configuration of the matching algorithm, including the weights of the different similarity metrics used by the matcher and the associated thresholds. The system allows the user to spread their work on the matching task across several different sessions, making the approach particularly well suited for matching large ontologies.

This special issue would not have been possible without support from many different people. Seven papers were submitted, of which these three were accepted. We are very grateful to the authors of all seven papers for their response to the call. Additionally, we were very fortunate to benefit from the diligent reviews submitted in response to each paper. Thanks go to all those who participated in the review process. Finally, we would like to thank the editorial team of the Semantic Web Journal for giving us the opportunity to expose some of the current work related to ontology and linked data matching.

References

[1] Ontology alignment evaluation initiative. http://oaei.ontologymatching.org. Accessed: 2016-11-21.

[2] C. Batini, M. Lenzerini, and S. B. Navathe. A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys*, 18(4):323–364, 1986.

[3] C. F. Dorneles, R. Gonçalves, and R. dos Santos Mello. Approximate data instance matching: a survey. *Knowledge and Information Systems*, 27(1):1–21, 2011.

[4] J. Euzenat and P. Shvaiko. *Ontology Matching*. Springer, 2007.

[5] S. Soderland. Learning information extraction rules for semi-structured and free text. *Machine learning*, 34(1-3):233–272, 1999.