

# An Ontology-based Data Matching Framework: Use Case Competency-based HRM

Peter De Baer<sup>1</sup>, Yan Tang<sup>1</sup>, Pieter De Leenheer<sup>2</sup>

<sup>1</sup> VUB STARLab, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

<sup>2</sup> Collibra nv/sa, Ransbeekstraat 230, 12 Brussels, Belgium

{Peter.De.Baer, Yan.Tang}@vub.ac.be; Pieter@Collibra.com

**Abstract.** As part of the European PROLIX (Process Oriented Learning and Information eXchange) project, VUB STARLab designed a generic ontology-based data matching framework (ODMF). Within the project, the ODMF is used to calculate the similarity between data elements, e.g. competency, function, person, task, and qualification, based on competency-information. Several ontology-based data matching strategies were implemented and evaluated as part of the ODMF. In this article we describe the ODMF and discuss the implemented matching strategies.

**Keywords:** data matching, competency management, matchmaking, ontology

## 1 ODMF

Semantic data matching plays an important role in many modern ICT systems. Examples are data mining [6], electronic markets [1], HRM [2], service discovery [5], etc. Many existing solutions, for example [2], make use of description logics and are often tightly linked to certain ontology engineering platforms and/or domains of data matching. This often leads to a knowledge bottleneck because many potential domain users and domain experts may not be familiar with description logics or the specific platform at hand. To avoid such potential technical barrier we designed the ODMF so that it is independent of a particular ontology engineering platform, and does not require the use of description logics. Instead, we make use of the combination of an ontologically structured terminological database [3] and a DOGMA ontology [4] to describe data. Both the DOGMA ontology and the terminological database make use of natural language to describe meaning. On top of this semantic data model we developed an interpreter module and a comparison module. Both the interpreter and the comparator make use of a library of matching algorithms. The matching algorithms have access to the data model via an API, and may be written in any programming language that can access this Java API. Via the terminology base, data can be described and interpreted in different natural languages. We believe that this multilingualism will improve the usefulness of the framework within an international setting.

The ODMF is designed to support data matching in general. Currently, the ODMF has been, however, only implemented and evaluated as part of the European

integrated PROLIX project<sup>1</sup>. Within the PROLIX platform<sup>2</sup>, the ODMF supports semantic matching of competency-based data elements, e.g. competency, function, person, task, and qualification.

## 2 Matching strategies

We implemented and evaluated several ontology-based data matching algorithms within the ODMF. These algorithms relate to three major groups: (1) string matching, (2) lexical matching, and (3) graph matching. However, most matching algorithms make use of a combination of these techniques.

1. *String matching techniques* are useful to identify data objects, e.g. competences and qualifications, using a (partial) lexical representation of the object. We selected two matching tools for this type of data matching: (a) regular expressions and (b) the SecondString<sup>3</sup> library.
2. *Lexical matching techniques* are useful to identify data objects, e.g. competences and qualifications, using a (partial) lexical representation of the object. In addition to plain string matching techniques, linguistic information is used to improve the matching. We selected two techniques to improve the matching: (a) tokenization and lemmatization and (b) the use of an ontologically structured terminological database.
3. *Graph matching techniques* are useful (a) to calculate the similarity between two given objects and (b) to find related objects for a given object.

## References

1. Agarwal, S., Lamparter, S.: SMART - A Semantic Matchmaking Portal for Electronic Markets. In: Proceedings of the Seventh IEEE International Conference on E-Commerce Technology, pp. 405 – 408, (2005)
2. Biesalski, E., Breiter, M., Abecker, A.: Towards Integrated, Intelligent Human Resource Management. In: 1st workshop "FOMI 2005", Formal Ontologies Meet Industry (2005)
3. De Baer, P., Kerremans, K., and Temmerman, R.: Constructing Ontology-underpinned Terminological Resources. A Categorisation Framework API. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, Copenhagen (2008)
4. Jarrar, M., Meersman, R.: Formal Ontology Engineering in the DOGMA Approach. In: On the Move to Meaningful Internet Systems: CoopIS, DOA, and ODBASE, LNCS, Springer Verlag, pp. 1238-1254 (2002)
5. Shu, G., Rana, O. F., Avis, N. J., Dingfang, C.: Ontology-based semantic matchmaking approach. In: Advances in Engineering Software 38, pp. 59–67, ScienceDirect (2007)
6. Stamou, S., Ntoulas, A., and Christodoulakis, D.: TODE- an ontology based model for the dynamic population of web directories. In: "Data Mining with Ontologies: Implementations, Findings and Frameworks", edited by Nigro, H., O., Cisaró, S., E., G., Xodo, D., H., IGI Global (2007)

---

<sup>1</sup> <http://www.prolixproject.org/>

<sup>2</sup> <http://prolixportal.prolix-dev.de/>

<sup>3</sup> <http://secondstring.sourceforge.net/>