

On Discovering Semantic Correspondences Between Ontologies

Jorge Martinez-Gil

Abstract

This is a summary of the main ideas of the doctoral thesis *On Discovering Semantic Correspondences Between Ontologies*, defended at the University of Málaga in 2010. The thesis studies methods for discovering correspondences between independently developed ontologies with limited human supervision. Its main proposal treats ontology matching as an automatic configuration problem: several elementary similarity functions are selected and combined to obtain a matching function suited to a given task. This idea leads to ontology meta-matching, where heuristic search is used to determine useful matchers, weights, aggregation choices, and thresholds. The thesis also introduces reverse ontology matching, which seeks a matching function capable of reproducing an existing alignment. Publications from 2007–2010 document the empirical preparation, optimization methods, related applications, and the reverse-matching proposal.

Keywords: ontology matching, ontology alignment, meta-matching, genetic algorithms, reverse ontology matching, Semantic Web

1 Introduction

Ontologies encode concepts, relations, and constraints in a machine-readable form. Two ontologies may describe the same subject yet use different names, structures, modeling choices, or levels of detail. Ontology matching seeks correspondences between their entities, and the resulting set of correspondences is commonly called an alignment. Reliable alignments support semantic data integration, service discovery, information exchange, and knowledge reuse.

The thesis addresses a recurring difficulty: no single similarity function performs well for every pair of ontologies. A lexical method may work when entity names are informative, a structural method may work when taxonomies are comparable, and a semantic method may be preferable when terminology differs. The proposed solution is to combine elementary methods and to configure that combination automatically. This research direction belongs to a wider effort to make Web information amenable to formal computation and machine reasoning.

Several preceding studies supplied technical and empirical foundations. Work on versioned and temporal XML documents dealt with branched changes in structured Web data. Research on alternative textual renderings examined how ontology representation affects the evidence available to a matcher [4]. These studies helped establish two assumptions used throughout the thesis: representations affect matching quality, and semantic resources must be processed under heterogeneous and changing conditions.

2 From Elementary Matchers to Meta-Matching

Let O_1 and O_2 be two ontologies. An elementary matcher s_i assigns a similarity value to an entity pair (x, y) , where $x \in O_1$ and $y \in O_2$. A composite matching function can be written in a simple weighted form:

$$M(x, y; \mathbf{w}) = \sum_{i=1}^k w_i s_i(x, y), \quad (1)$$

where w_i denotes the contribution of matcher s_i . A correspondence is accepted when $M(x, y; \mathbf{w})$ exceeds a threshold τ . More elaborate versions may select only a subset of matchers or use another aggregation rule. The design problem is to determine the configuration

$$\theta = (S, \mathbf{w}, \tau, A), \quad (2)$$

where S is the selected matcher set and A is the aggregation operation.

Ontology meta-matching searches for a configuration θ that optimizes a chosen quality criterion. The target may favor precision, recall, F -measure, or a low false-positive rate. This separates two tasks that are

often mixed together: elementary matchers produce evidence about entity pairs, and the meta-matching method decides how that evidence should be combined.

The genetic-algorithm study showed how a population-based search can optimize composite ontology matchers without exhaustively enumerating the parameter space [5]. Candidate solutions encode matcher weights and related parameters; their fitness is measured from the alignment quality obtained on training cases. The associated JISBD study evaluated two heuristic approaches to the same meta-matching problem, contrasting a lower-cost search with a genetic search over a larger set of configurations. These methods pursue the thesis objective of reducing manual tuning and allowing the matching process to adapt to the selected evaluation goal.

3 Empirical Preparation and Similarity Design

Automatic configuration should reflect the properties of the ontologies that will be matched. A statistical study of a sizable OWL sample examined features relevant to alignment, including ontology language, size, and the number and type of represented entities [2]. This empirical work motivated parameterized methods that do not assume a single typical ontology shape.

Representation is another source of variation. The comparison of textual renderings studied whether alternative textual descriptions of ontology entities can improve alignment quality [4]. The underlying lesson is that a matcher receives a derived view of an ontology rather than an entirely neutral input. Labels, identifiers, comments, and serialized descriptions expose different signals, and their usefulness varies across data sets.

The MaSiMe study developed a configurable similarity measure for tag-cloud refactoring [3]. Its application differs from ontology alignment, yet it supports the same methodological idea: similarity should be adjustable to task requirements rather than fixed in advance. MaSiMe combines user considerations, removes redundant tags, and uses a parameter study to identify a useful configuration. It provides an application-level demonstration of customized similarity aggregation during the period in which the thesis was developed.

4 Reverse Ontology Matching

Standard ontology matching applies a matching function to two ontologies and produces an alignment. Reverse ontology matching starts from an existing alignment and seeks a function that can reproduce it [1]. Let R be a reference alignment. The reverse problem can be expressed as

$$\theta^* = \arg \max_{\theta} Q(M_{\theta}(O_1, O_2), R), \quad (3)$$

where Q measures agreement between the alignment generated with configuration θ and the observed alignment R .

The aim is usually an equivalent matching function, not a proof that the exact historical procedure has been recovered. Several configurations may yield the same alignment, especially when the example set is small. The recovered function is still useful when it captures the decisions encoded in the alignment.

Reverse ontology matching offers several forms of reuse. It can model the behavior of a human expert, support matching from examples, imitate an existing tool, or represent the information stored in a collection of alignments [1]. This changes the role of an alignment: it becomes training evidence for a reusable procedure rather than only a static output.

5 Related Semantic-Web Applications

The thesis research was connected to applications in which semantic correspondences affect the use of Web resources. Work on SAWSDL services studied automatic grounding and the improvement of annotated service descriptions. Ontology matching is relevant in this setting because service annotations may refer to distinct vocabularies, making correspondence discovery necessary for service invocation or composition.

The conceptual discussion of formal reasoning on the Web, the work on textual ontology views [4], and the SAWSDL application place ontology alignment within a broader program of machine-processable Web knowledge. The thesis contributes an optimization-oriented component to that program: it studies how correspondence-producing functions can be assembled, learned, and recovered.

6 Contributions and Limitations

The thesis makes three principal contributions. It formulates ontology matching as a configurable combination of elementary matchers. It applies heuristic optimization to choose effective configurations with reduced user intervention. It introduces reverse ontology matching as a means of learning from previously created alignments.

The approach also has limits. Optimization requires evaluation data, and errors in a reference alignment can shape the learned function. A configuration learned for one family of ontologies may transfer poorly to another family. Reverse matching may return several equally plausible functions, so the recovered model should be interpreted as behaviorally equivalent on the available cases. Quality scores based on correspondence sets also omit some properties relevant to an application, including logical consistency and the cost of human correction.

These limits do not weaken the main result. They identify where empirical sampling, validation sets, and task-specific objectives are needed. The OWL study [2] and the heuristic evaluations [5] provide an initial basis for such choices.

7 Closing Remarks

On Discovering Semantic Correspondences Between Ontologies presents ontology alignment as a search over matching functions rather than as the execution of one fixed algorithm. Meta-matching selects and combines elementary evidence according to an explicit quality goal. Reverse ontology matching extracts a reusable function from an existing alignment. The publications produced through 2010 trace the route from Web data representation and similarity design to genetic optimization, empirical OWL analysis, semantic-service applications, and reverse engineering of alignment behavior. Together, they form a coherent account of how ontology matching can become less dependent on repeated manual configuration.

References

- [1] Jorge Martinez-Gil and José Francisco Aldana-Montes. Reverse ontology matching. *SIGMOD Record*, 39(4):5–11, 2010.
- [2] Jorge Martinez-Gil, Enrique Alba, and José Francisco Aldana-Montes. Statistical study about existing OWL ontologies from a significant sample as previous step for their alignment. In *Proceedings of the Fourth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, pages 980–985. IEEE Computer Society, 2010.
- [3] David Urdiales-Nieto, Jorge Martinez-Gil, and José Francisco Aldana-Montes. MaSiMe: A customized similarity measure and its application for tag cloud refactoring. In *On the Move to Meaningful Internet Systems: OTM 2009 Workshops*, volume 5872 of *Lecture Notes in Computer Science*, pages 937–946. Springer, 2009.
- [4] Jorge Martinez-Gil, Ismael Navas-Delgado, Antonio Polo Márquez, and José Francisco Aldana-Montes. Comparison of textual renderings of ontologies for improving their alignment. In *Proceedings of the Second International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, pages 871–876. IEEE Computer Society, 2008.
- [5] Jorge Martinez-Gil, Enrique Alba, and José Francisco Aldana-Montes. Optimizing ontology alignments by using genetic algorithms. In *Proceedings of the First International Workshop on Nature Inspired Reasoning for the Semantic Web (NatuReS)*, volume 419 of *CEUR Workshop Proceedings*, pages 31–45, 2008.