Chapter 9
Semantic Matching with S-Match

Pavel Shvaiko, Fausto Giunchiglia, and Mikalai Yatskevich

Abstract We view matching as an operation that takes two graph-like structures (e.g., lightweight ontologies) and produces an alignment between the nodes of these graphs that correspond semantically to each other. Semantic matching is based on two ideas: (i) we discover an alignment by computing semantic relations (e.g., equivalence, more general); (ii) we determine semantic relations by analyzing the meaning (concepts, not labels) which is codified in the entities and the structures of ontologies. In this chapter, we first overview the state of the art in the ontology matching field. Then we present basic and optimized algorithms for semantic matching as well as their implementation within the S-Match system. Finally, we evaluate S-Match against state of the art systems, thereby justifying empirically the strength of the approach.

9.1 Introduction

Matching is an important operation in many applications, such as ontology integration, data warehouses, peer-to-peer data sharing, etc. The matching operation takes two graph-like structures (e.g., lightweight ontologies [14]) and produces an alignment, that is a set of mapping elements (or correspondences), between the nodes of the graphs that correspond semantically to each other.

There exist various solutions of matching, see [7, 37, 40, 41] for recent surveys. In turn, some recent examples of individual matching approaches can be found in [2, 3, 6, 8, 10–12, 23, 27, 32, 38].1 We concentrate on a schema-based solution, i.e.,

---

1See www.OntologyMatching.org for a complete information on the topic.
a matching system that exploits only the schema information and does not consider the instance information. We follow an approach called semantic matching [15]. This approach is based on two key ideas. The first is that we calculate correspondences between ontology entities by computing semantic relations (e.g., equivalence, more general, disjointness), instead of computing coefficients rating match quality in the $[0, 1]$ range, as it is the case in most previous approaches, see, for example, [6, 10, 30, 33]. The second idea is that we determine semantic relations by analyzing the meaning (concepts, not labels) which is codified in the entities and the structures of ontologies. In particular, labels at nodes, written in natural language, are automatically translated into propositional formulas which explicitly codify the labels’ intended meaning. This allows us to translate the matching problem into a propositional validity problem, which can then be efficiently resolved using (sound and complete) state of the art propositional satisfiability (SAT) deciders, e.g., [28].

A vision for the semantic matching approach and some of its implementation within S-Match were reported in [15–17, 19, 22]. In turn, the works in [18, 21, 23, 43] focused on the following aspects of S-Match: (i) algorithms and implementation, (ii) discovering missing background knowledge in matching tasks, (iii) explanation of matching results and (iv) large-scale evaluation. This chapter builds on top of the above mentioned works and provides a summative account for the semantic matching approach (hence, the key algorithms are identical to those in the above mentioned works).

The rest of the chapter is organized as follows. Section 9.2 overviews the state of the art in the ontology matching field. Section 9.3 presents the semantic matching approach. Section 9.4 introduces the optimizations that allow improving efficiency of the basic version of Sect. 9.3. Section 9.5 outlines the S-Match architecture. The evaluation results are presented in Sect. 9.6. Finally, Sect. 9.7 provides conclusions and discusses future work.

9.2 State of the Art

A good survey and a classification of ontology matching approaches up to 2001 was provided in [40], a semantics driven extension of its schema-based part and a user-centric classification of matching systems was provided in [41], while the work in [9] considered both [40, 41] as well as some other classifications.

In particular, for individual matchers, [41] introduced the following criteria which allow for detailing further the element and structure level of matching: syntactic techniques (these interpret their input as a function of their sole structures following some clearly stated algorithms, e.g., iterative fix point computation for