Title	Semantic Enrichment in Ontologies for Analysis and Matching
Author(s)	Nwe, Ni Tun
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3567
Rights	
Description	Supervisor:Satoshi Tojo, 情報科学研究科, 博士

Semantic Enrichment in Ontologies for Analysis and Matching

Ontology---a formal, explicit, shared conceptualization of a domain---is intended to facilitate semantic interoperability among distributed and intelligent information systems where diverse software components, computing devices, knowledge, and data, are involved. Since a single global ontology is no longer sufficient to support a variety of tasks performed on differently conceptualized knowledge, ontologies have proliferated in multiple forms of heterogeneity even for the same domain, and such ontologies are called *heterogeneous ontologies*. For interoperating among information systems through heterogeneous ontologies, mapping mechanisms need to bridge their knowledge gaps. *Ontology matching* (or mapping) is a process of finding correspondences between semantically related entities in heterogeneous ontologies.

The main aim of this research is *to deal with wide-scale semantic heterogeneity in ontology matching*. Although several efforts in ontology mapping have already been contributed, they have different focuses, assumptions, and limitations. A common idea among existing methods is that possible correspondences between two ontologies are determined by the similarity of entity names; this is known as *name-based matching*. In order to decide semantic correspondences between concepts, the methods need to analyze the similarities between all related properties and instances; this is known as *content-based matching*. In the case of wide-scale semantic heterogeneity, content-based matching becomes complex, and user's approval or expert-interaction needs to verify mapping results.

In my research, I focus on two issues. The first issue is that the chance of correspondence between two terminologically quite different concepts is very less or not obtainable through name-based matching, because the name of a concept cannot express the precise semantics of the concept. Thus, what is an alternative approach besides name-based matching, to determine the possible correspondences between terminologically heterogeneous ontologies? The second issue is how to reduce complexity, concerning wide-scale semantic heterogeneity in content-based matching. To accomplish the major aim and focuses, my underlying assumption is *the more explicit semantics is specified in ontologies, the feasibility of matching will be greater*. Therefore, I proposed a semantically-enriched model of ontologies (called EnOntoModel) in which every domain concept is treated as a sort---an entity type that carries a criteria for determining the individuation, persistence, and identity of its instances---regarding every individual defined in a universe of discourse is countable and identifiable. In the philosophical literature, ontological concepts can be classified into four disjoint sort categories: type, quasi-type, role, and phase. I set up a logic-based formal system to classify domain concepts into these sort categories, using three philosophical notions: identity, existential rigidity, and external dependency. In my research, this classification knowledge is intuitively represented as concept-level properties that are different from individual-level properties which are used to specify individuals. Then, I defined EnOntoModel in which the semantics of domain concepts are described by using individual-level properties, as well as concept-level properties; this is my approach of *semantic enrichment*. The innovation behind EnOntoModel is to supply an identifiable link between two heterogeneous descriptions of a concept, regarding that if two concepts are semantically equivalent, then they must be classified within the same sort category. For the usability of EnOntoModel, I implemented sortal meta-class ontology as an open source interface in enrichment process as well as conceptual analysis of enriched ontologies. By the aim of the thesis, I designed a matching method between enriched ontologies.

A novel idea of EnOntoModel-based Ontology Matching (EOM) method is that direct concept matching is driven between the same categories of sorts instead of exhaustive search among all sorts, because domain concepts are systematically classified into four disjoint sort categories. Moreover, it is examined that EnOntoModel can support not only determining the scope of possible correspondences, but also determining the most relevant properties which can certainly indicate a correspondence between two similar concepts. This means that semantic correspondences between highly heterogeneous concepts can be achieved without taking an exhaustive search in taxonomies and an analysis among all related properties.

The method is implemented in Java for matching between OWL ontologies by utilizing Jena OWL API and Protégé OWL API. The efficiency of EOM method is evaluated in terms of mathematical complexity and proved that this method could reduce the complexity of the matching process by comparing it with other methods, particularly GLUE's content learners. Moreover, an experiment is done in two real data sets, and the effectiveness of EOM is shown in terms of precision and recall.