

DIMITRIOS A. KOUTSOMITROPOULOS

**«Methods and Techniques for Semantic Web Knowledge Discovery: Deductive Knowledge Acquisition from Ontology Documents and the Semantic Profiling Technique»**

DOCTORATE THESIS

ABSTRACT

Semantic Web is a combination of technologies and standards in order to give Web information strictly defined semantic structure and meaning. Its aim is to enable Web users and automated agents to process, manage and utilize properly described information in intelligent and efficient ways.

Nevertheless, despite the various techniques that have been proposed, there is no clear method such that, by taking advantage of Semantic Web technologies, to be able to retrieve information deductively, i.e. to infer new and implicit information based on explicitly expressed facts.

In order to address this situation, the problem of *Semantic Web Knowledge Discovery* (SWKD) is first specified and introduced. SWKD takes advantage of the semantic underpinnings and semantic descriptions of information, organized in a logic theory (i.e. ontologies expressed in OWL). Through the use of appropriate automated reasoning mechanisms, SWKD makes then possible to deduce new and unexpressed information that is only implied among explicit facts.

The question as to whether and to what extent do Semantic Web technologies and logic theory contribute *efficiently* and *expressively* enough to the SWKD problem is evaluated through the establishment of a *SWKD methodology*, which builds upon recent theoretical results, as well as on the qualitative and experimental comparison of some popular inference engines, based on Description Logics. It is shown that the efficiency and expressivity of this method depends on specific theoretical, organizational and technical limitations.

The experimental verification of this methodology is achieved through the development and demonstration of the *Knowledge Discovery Interface* (KDI), a web-distributed service that has been successfully applied on experimental data. The results taken through the KDI confirm, to a certain extent, the assumptions made mostly about expressivity and motivate the examination and investigation of the newly proposed extensions to the Semantic Web logic theory, namely the OWL 1.1 language.

In order to strengthen the expressivity of knowledge discovery in the case of particular knowledge domains a new technique is introduced, known as *Semantic Profiling*. This technique evolves traditional *Metadata Application Profiling* from a flat aggregation and mixing of schemata and metadata elements to the substantial extension and semantic enhancement and enrichment of the model on which it is applied. Thus, semantic profiling actually profiles an ontological model for a particular application, not only by bringing together vocabularies from disparate schemata, but also through the semantic intension and semantic refinement of the initial model. This technique and its results are experimentally verified through its application on the CIDOC-CRM cultural heritage information model and it

is shown that, through appropriate methods, the general applicability of the model can be preserved.

However, for SWKD to be of much value, it requires the availability of rich and detailed resource descriptions. Even though information compatible with the Semantic Web logic theory are not always readily available, there are plenty of data organized in flat metadata schemata. To this end, it is investigated whether SWKD can be *efficiently* and *expressively* applied on such *semi-structured knowledge models*, as is the case for example with the Dublin Core metadata schema. It is shown that this problem can be partially reduced to applying semantic profiling on such models and, in order to retain interoperability and resolve potential ambiguities, the OWL 1.1 *punning* feature is investigated, based on which a name definition may have variable semantic interpretation depending on the ontological context.

In conclusion, these newly proposed methods can improve the SWKD problem in terms of expressive strength, while keeping complexity as low as possible. They also contribute to the creation of expressive descriptions from existing metadata, suggesting a solution to the Semantic Web *bootstrapping* problem. Finally, they can be utilized as the basis for implementing more efficient techniques that involve distributed and incremental reasoning.