

Abstract

Availability of controlled vocabularies, ontologies, and so on is enabling feature to provide some added values in terms of knowledge management. Nevertheless, the design, maintenance and construction of domain ontologies are a human intensive and time consuming task. The Knowledge Extraction consists of automatic techniques aimed to identify and to define relevant concepts and relations of the domain of interest by analyzing structured (relational databases, XML) and unstructured (text, documents, images) sources.

Specifically, methodology for knowledge extraction defined in this research work is aimed at enabling automatic ontology/taxonomy construction from existing resources in order to obtain useful information. For instance, the experimental results take into account data produced with Web 2.0 tools (e.g., RSS-Feed, Enterprise Wiki, Corporate Blog, etc.), text documents, and so on. Final results of Knowledge Extraction methodology are taxonomies or ontologies represented in a machine oriented manner by means of semantic web technologies, such as: RDFS, OWL and SKOS.

The resulting knowledge models have been applied to different goals.

On the one hand, the methodology has been applied in order to extract ontologies and taxonomies and to semantically annotate text. On the other hand, the resulting ontologies and taxonomies are exploited in order to enhance information retrieval performance and to categorize incoming data and to provide an easy way to find interesting resources (such as faceted browsing). Specifically, following objectives have been addressed in this research work:

- *Ontology/Taxonomy Extraction*: that concerns to automatic extraction of hierarchical conceptualizations (i.e., taxonomies) and relations expressed by means typical description logic constructs (i.e., ontologies).
- *Information Retrieval*: definition of a technique to perform concept-based the retrieval of information according to the user queries.
- *Faceted Browsing*: in order to automatically provide faceted browsing capabilities according to the categorization of the extracted contents.
- *Semantic Annotation*: definition of a text analysis process, aimed to automatically annotate subjects and predicates identified.

The experimental results have been obtained in some application domains: e-learning, enterprise human resource management, clinical decision support system.

Future challenges go in the following directions: investigate approaches to support ontology alignment and merging applied to knowledge management.

Keywords

Knowledge Extraction, Ontology Extraction, Ontology, OWL, Fuzzy Relational Concept Analysis, Fuzzy Formal Concept Analysis, Semantic Web, Clinical Decision Support System, Faceted-Browsing, Information Retrieval, Semantic Annotation.