Knowledge Discovery in the semantic wiki for ontology enrichment

Context and motivation

Semantic wikis are collaborative sites that allow users in a community to create and share knowledge. The wiki pages are semantically annotated by a set of categories previously defined as concepts in a domain ontology and tags (keywords) can be freely associated with them. The ontology contains all terms, relationships and constraints on expert knowledge in the field. It is maintained and used by them to annotate the wiki pages by categories. The categories are created by the experts and make it possible to organize the links between the pages. The tags are provided freely by the users and associate to a page a particular knowledge or an original point of view of the user. The wiki thus allows to store an enormous amount of semantic data on its pages which hides knowledge that can be exploited for a better organization and use. In order to improve navigation and semantic search in the wiki, the hierarchy of categories must be updated periodically.

In order to solve this problem, a solution consists in using data mining techniques for ontology enrichment. In [1], the authors first extract association rules in a medical database. Then, they compare these rules with those defined in the ontology. Only the new and unexpected rules are retained and used for the enrichment of the ontology. In [2], an approach based on the frequent patterns of useful tags makes it possible to find new categories of the semantic wiki that are used for ontology enrichment. In this approach, the authors have introduced in the frequent patterns discovery process a semantic pruning phase in order to remove some patterns already founded in the ontology. However, it is difficult to formalize all such kinds of patterns. Similarly, in these different approaches, the relations between the elements of the database are not taken into account. More precisely, in the case of the semantic wiki, the semantic relations between pages (properties) have not been taken into account. This is the same for semantic relations between categories. However, we can say for instance that two pages in the same category are closer than two pages of different categories. To support semantic relations, [3] proposes the use of a logical formalism for the extraction of multi-relational association rules from the knowledge base of the ontology for its enrichment. To support semantic relations, [3] proposes the use of a logical formalism for the extraction of multi-relational association rules from the knowledge base of the ontology for its enrichment.

Objectives of the thesis

The aim is to take stock of the use of data mining techniques, in particular association rules for the ontology enrichment from different sources, to propose new methods for extracting knowledge from Semantic wiki for ontology enrichment. These techniques should take into
account the semantic relations between pages and between categories. They should as far as possible integrate ontology knowledge into the extraction process to avoid costly extraction. The main expected contributions are:

- State of the art on the use of data mining techniques, in particular association rules for ontology enrichment from different sources
- Contribution to ontology enrichment in a semantic wiki by the use of data mining techniques that take into account the semantic relations between pages and categories
- Contribution to data mining techniques by taking into account ontology knowledge during the extraction process

Contact:
Cheikh Talibouya Diop: cheikh-talibouya.diop@ugb.edu.sn

