Chapter 13

Adaptive Ontology-Based Web Information Retrieval: The TARGET Framework

Cédric Pruski
Centre de Recherche Public Henri Tudor, Luxembourg

Nicolas Guelfi
University of Luxembourg, Luxembourg

Chantal Reynaud
Laboratory of Computer Science (LRI), University of Paris-Sud, France

ABSTRACT

Finding relevant information on the Web is difficult for most users. Although Web search applications are improving, they must be more “intelligent” to adapt to the search domains targeted by queries, the evolution of these domains, and users’ characteristics. In this paper, the authors present the TARGET framework for Web Information Retrieval. The proposed approach relies on the use of ontologies of a particular nature, called adaptive ontologies, for representing both the search domain and a user’s profile. Unlike existing approaches on ontologies, the authors make adaptive ontologies adapt semi-automatically to the evolution of the modeled domain. The ontologies and their properties are exploited for domain specific Web search purposes. The authors propose graph-based data structures for enriching Web data in semantics, as well as define an automatic query expansion technique to adapt a query to users’ real needs. The enriched query is evaluated on the previously defined graph-based data structures representing a set of Web pages returned by a usual search engine in order to extract the most relevant information according to user needs. The overall TARGET framework is formalized using first-order logic and fully tool supported.

DOI: 10.4018/978-1-4666-2779-6.ch013
INTRODUCTION

Information retrieval is a topic that has been under investigation for years. Human beings, who are curious by nature, are always looking for improving their knowledge on a given subject. This is all the more so true since the advent and the popularity of the WWW (Berners-Lee, Cailliau, Groff, & Pollermann, 1992) which has become the largest and the most dynamic accessible source of information ever. Nevertheless, because of the aforementioned characteristics and the heterogeneity of its content (from the structural to the semantic point of view), it is often hard for common Web users to find the information they are really interested in.

One reason for this failure is the difficulty for users to understand the way search applications interpret the submitted queries. In fact, the dynamic aspect of knowledge in general and of the Web in particular made that the selected keywords are often outdated and search engines are not able to adapt queries to this evolution. Another reason lies in the difficulty for users to clearly characterize, at query level, the search domain as well as their view on this domain. Actually, the former is usually huge and most of the time fuzzy in users’ mind. This is why common queries are made up of two or three keywords, usually ambiguous which give poor results at interpretation time. Therefore, it will be a huge advantage for users if they are assisted for the characterization of the targeted search domain as well as for expressing their view on the domain they are interested in. To this end, technologies of the Semantic Web (Berners-Lee, Hendler, & Lassila, 2001) can be the key to success. Actually, ontologies (Gruber, 1993) have this modeling ability to represent a given domain and to offer vocabularies to express queries. However, existing work on ontology evolution is not mature enough to provide a technique for making ontology adapt automatically to changes in the modeled domain.

In addition, ontologies can be helpful in the structuring of the Web. At the beginning, the World Wide Web was made of documents containing only textual information and its structure was based on hyperlinks pointing from a page to another which was the success story for many Web search engines (Page & Brin, 1998). Then, because of its ever increasing popularity, the content of the Web has evolved in quantity but also in quality (use of multimedia, definition of languages for structuring the content, etc). In particular the introduction of a tinge of semantics extracted from ontologies improves Web structuring which, in turn, facilitates Web search and increases the relevance of the returned pages.

The relevance of the results of a search depends on the choice of the keywords of a query and of the interpretation of thereof by search engines. Queries have to be built in accordance with the targeted search domain and the knowledge that characterizes users best. Both domains can be modeled using ontologies. Ontologies provide users with vocabularies to express queries according to well defined rules. But they enabled also user queries to be enriched in order to integrate the targeted domain, users’ characteristics and the evolution of the search domain. Verified on appropriate web data structures, such queries can lead to more relevant results than results provided by usual search engines.

Following this approach, we introduce the TARGET framework for improving the relevance of a domain specific Web search. Its foundations rely on the use of adaptive ontologies, a new model of ontology based on the ideas developed by psychologists (Piaget, 1974), that aims at facilitating and automating the adaptation of ontology according to ongoing changes occurring in the real world. We propose a mechanism for adaptation which relies on the definition of rules that make it possible for the ontology to evolve semi-automatically. These ontologies are central in our approach. First, they offer a vocabulary to express queries. Second, we propose to use