

Using Ontologies To Improve Search On WWW.HR Web Directory

Damir Jurić, Maja Matijašević, Gordan Gledec

Department of Telecommunications

Faculty of Electrical Engineering and Computing, Zagreb, Croatia

E-mail: damo_damr@yahoo.co.uk

1. Introduction

Ontology is a hierarchical structure of knowledge about entities which subcategorizing entities according to their essential (or at least relevant and/or cognitive) qualities. Entities can have subcategories (subclasses or up classes), or different names (alternative words) and restrictions (entities with specific characteristics). This paper proposes using ontologies for the purposes of searching and viewing information in the Web directory, and describes the application that is capable of using ontologies to get conclusions from the input data on the more conceptual level

2. Motivation

Today, search engines mostly use search indices to find the occurrences of the words the users enter in the query, but the meaning of such words is usually unattended. Results differ from users' wishes, and can often be meaningless. The goal of this work is to create ontology and an application that is able to search that ontology and view results for WWW.HR Web directory users. Ontology is created using the Ontology Web Language (OWL) [1].

3. Creating the ontology

Creating the ontology may be a complex job because there is no automated process to do it. Final results depend on the creator's knowledge. However, the process of creation could be split into three simpler subprocesses: consideration about domain, planning the domain and finally, writing the ontology.

Consideration about the domain is the first step in ontology creation, involving decisions like what will be part of ontology and what should be omitted. Already existing categorization of the WWW.HR Web directory was used in order to complete this step. Planning the ontology is the most difficult and demanding task. This step requires consultation with relevant sources of information (like current subcategorization of the Web directory or analysis of the most popular words users use to search the directory). Since all required knowledge about the domain cannot be extracted from such sources, parts of the knowledge must come from the creator of the ontology, whose knowledge is used to synthesize knowledge from many different sources. After the process, the ontology can be defined in OWL.

4. Application creation

The application was coded in Perl (*Practical Extraction and Report Language*) [2]. If the regular search yielded insufficient number of relevant results, the application may consult ontology and extract knowledge about user's query. That knowledge is contained in the subclasses of the queried word, alternative words or synonyms, up classes and possible restrictions. Subclasses are words that have the meaning (albeit partially) similar to the query, so searching with subclasses is still similar like searching with the original class. Alternative words are usually synonyms or words that come from other languages. Restrictions are used if queries have specific requirements.

5. Results and discussion

Three types of tests were conducted. In the first test, the directory was searched along with alternative words for user's query. In the second test all subclasses were included. In the third test, we used restrictions. First two tests showed more results than standard search, and the third test showed the opposite, just as it was expected. It is important to search along with alternative words because in most languages one word usually has several other forms. Using the ontologies, less common words can be included in input, such as:

- words that have synonyms that are used almost with the same frequency as the original word,
- words that come from foreign language but are used in Croatian language almost as frequently as the original words,
- one word may be used for many other words although they are just similar words and not synonyms,
- words may have synonyms used much more frequently than the original words.

Searching with all subclasses included yielded more results than standard search. Searching for query "*car transport*" with subclasses like "*bus transport*" or "*stations*" resulted in 56 matches, as opposed to 3, when searching only using query "*car transport*". Searching for query „*historical monuments*“ yielded only 1 result, but with subclasses like „*church*“ and „*castle*“ included, it yielded 50 results.

Searching with restrictions resulted in less matches because of existing specific requirements (like location or part), so only pages matching those requirements were taken into account. Example would be for the query „*airport*“ which yielded 9 results, but ontology found restriction for location (cities Zagreb, Zadar, Split), so the number of results could be much smaller.

6. Conclusion

In this paper we propose using ontologies for the purpose of searching Web directory like www.hr.hr. Ontology in the domain of “tourism” was devised, and application capable of handling it created and tested. The results show important improvements in regards to the number of matching results returned to user queries.

References:

- [1] World Wide Web Consortium: URL: <http://www.w3c.org/2001/sw/WebOnt/>
- [2] Comprehensive Perl Archive Network, URL: <http://www.cpan.org/>
- [3] Roger L. Costello, David B. Jacobs, *OWL Web Ontology Language*, tutorial, The MITRE Corporation, 2003. URL :<http://www.xfront.com/>
- [4] Đurđica Težak, *Pretraživanje informacija na Internetu: priručnik s vježbama*, Hrvatska sveučilišna naklada, Zagreb. 2002.
- [5] Andrijana Prskalo, *WWW tražilica prilagođena hrvatskom jeziku*, Diplomski rad br. 2186, Fakultet elektrotehnike i računarstva, Zagreb, 1997.