

Web Semantics for Intelligent and Dynamic Information Retrieval Illustrated Within the Mental Health Domain

Maja Hadzic, Elizabeth Chang

Digital Ecosystems and Business Intelligence Institute, Curtin University of Technology
GPO Box U1987, Perth 6845, Australia

{m.hadzic, e.change}@curtin.edu.au

Abstract. Much of the available information covering various knowledge domains is distributed across various information resources. The issues of distributed and heterogeneous information that often come without semantics, lack an underlying knowledge base or autonomy, and are dynamic in nature of are hindering efficient and effective information retrieval. Current search engines that are based on syntactic keywords are fundamental for retrieving information often leaving the search results meaningless. Increasing semantics of web content (web semantics) would solve this problem and enable the web to reach its full potential. This would allow machines to access web content, understand it, retrieve and process the data automatically rather than only display it. In this paper we illustrate how the ontology and multi-agent system technologies which underpin this vision can be effectively implemented within the mental health domain. The full realisation of this vision is still to come.

Keywords: web semantics, ontology, multi-agent systems, ontology-based multi-agent system, mental health, intelligent information retrieval, dynamic information retrieval.

1 Introduction

In 2007, the internet is frequently used as an information source for a multiplicity of knowledge domains. The general public uses Google predominately to obtain information covering these domains. The users inevitably have different access and understanding of the results they obtained from their search. As Google is not built to separate authoritative from dubious information sources, the users may have to use specialised search engines.

The accumulation of published information is an additional problem that complicates the search. For example, biomedical researchers would use PubMed which is a service of the U.S. National Library of Medicine that includes over 16 million citations from life science journals for biomedical articles back to the 1950s. Using the PubMed search engine, the user receives a list of journals related to the specified keyword. It is now left to the user to read each journal individually and try to establish links within this information. This would be easy if the journal list consisted of only a few journals. However, the journal list usually consists of thousands of

journals and the medical researchers usually do not have time to go through these results thoroughly. There is a high chance that some important information will be overlooked.

Specific and targeted searches are very difficult with current search engines. For example, a search for “genetic causes of bipolar disorder” using Google provides 960,000 hits consisting of a large assortment of well meaning general information sites with few interspersed evidence-based resources. The information provided by the government sites is not necessarily returned on the first page of a ‘google’ search. A similar search on Medline Plus retrieves all information about bipolar disorder plus information on other mental illnesses. The main problem of the current search engines is that they match specific strings of letters within the text rather than searching by meaning.

There is a need to design an intelligent search engine that would perform a search based not on keywords but on the meaning of the information. The search engine would go through the available information, understand this information and select highly relevant information as well as link this information and present it in a format meaningful to the user.

In this paper, we will briefly introduce the technologies underpinning such a search engine. We will address current issues related to information access and retrieval from the web. We will introduce the meaning of web semantics and the role of ontologies and agent technologies in the creation of a semantically rich environment. We will also illustrate this on an example from the mental health domain.

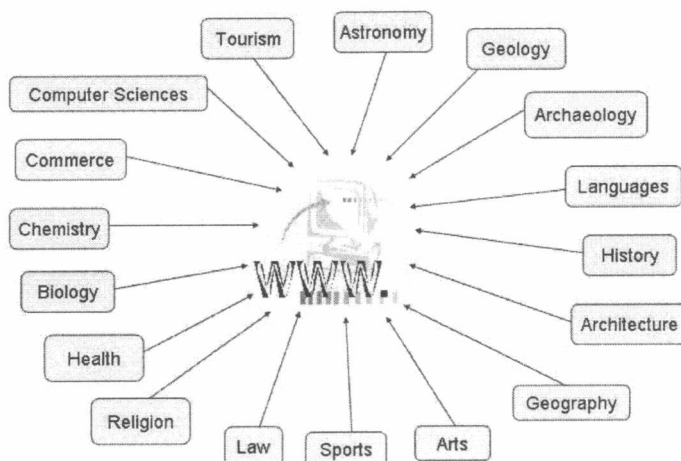


Figure 1 Information on the web covers various knowledge domains

2 Information Variety

The number of active users and information resources are increasing each day. Adding content to the web is quite an easy task and access to the web is uncomplicated and fast. The available information covers various knowledge domains and various disciplines such as astronomy, geology, biology, computer sciences, history, arts, religion and so forth (Figure 1).

Finding relevant information among all the information resources is a difficult task. The sheer volume of currently available information limits access and retrieval of required information. Current use of web content is based on browsing and syntactic key-searches.

