White Paper

SEMANTIC WEB AS THE WEB FULL OF MEANING

THE NETWORK OF DOCUMENTS VERSUS THE NETWORK OF MEANING

The Internet is an extraordinarily vast network of data, the amount of which is impossible to be stored or organized by any other means available to humans. It grows and reforms every second, spread across millions of computers around the world. It contains an immense amount of meaningful information which has to be structured in order to be accessible and comprehensible. Traditional methods of organization of data online applied most commonly from the beginning of the network's inception until today are in principle based on documents or files. This means that the stored data, although meaningful when accessed by human users, remains neutral in terms of content for the computers. This implies that any structuring of meaning has to be done manually and, considering the exceptional volume of data, it is a process that is difficult, cumbersome and often providing little effect. Automatic structuring on the other hand can only reach a certain level – for instance, when searching through documents conducted by means of search engines.

Although recent months have seen new features implemented in major search sites, the fundamental principle of their operation remains the same: search engine software scans the content of the Internet across hyperlinks connecting different pages for simple strings of characters. This process provides search results and, on the surface level, is in fact very quick and efficient, considering how vast the number of items to be analyzed and returned is. Search engines like Google use extremely sophisticated and ingenious algorithms that aim to provide as precise and relevant results as possible. There is however one extremely important drawback which only recently has begun to surface – the Web, with all its abundance of information, lacks semantics. Google will look for sequences of letters and words, but in essence this is it. All that is achieved is a list of files, meaningless semantically for the software. No extent of serious organization or interpretation of data by the computers themselves is possible. Moreover, documents have unique formal structure and the meaning that is understandable for humans stored in them cannot be effectively extracted – the most that can be done, as mentioned before, is basic detection of strings of characters.

MakoLab S.A., ul. Demokratyczna 46, 93-430, Łódź, t: 48 42 683 74 60, f: 48 42 683 74 99, www.makolab.pl info@makolab.pl

WHAT COMPUTERS CAN UNDERSTAND FROM SEMANTICS

Before investigating the description of the Semantic Web and the possibilities it offers as opposed to the dominant paradigm today, an important step will be to describe what is actually meant by meaning for a computer.

At the beginning of development of the concept and technology of the Semantic Web, there was some confusion in the media that actually hindered the movement by generating false expectations. There was talk of new possibilities for strong AI (Artificial Intelligence), for machines becoming able to understand natural text just as humans do. The research is strongly connected with developing AI, but this kind of revolution is not to be expected from the project.

What is meant by the verb “to understand” in the context of semantic software is to extract and be able to process meaningful content from documents. This involves performing logical operations, ascribing classes and properties, making inferences etc. As a result, we obtain tremendously powerful tools for cataloguing and deriving new meaning from vast amounts of data. On the other hand, if and how the process is unlike human comprehension of natural text is a matter of philosophical investigation. For practical purposes however, “understanding” in perspective of the Semantic Web means that a computer may perform operations on meaningful data itself, not just on basis of bare syntax.

ONTOLOGIES

This is entirely unlike what the conventional methods used today offer. Semantic Web enables to establish references of terms and corresponding beings of any given kind. This is achieved by creating ontologies. What the originally metaphysical term means here is a structure that specifies kinds of entities, relationships and rules that govern these relationships. An ontology lays a framework for entities to be ascribed to, and provides the semantic software with principles for categorizing and reasoning on them.

An ontology can pertain to any set of entities. It can describe terms used in a bookstore, a biological research project, or a commercial offer of products. Theoretically, any portion of human knowledge could be described this way.

PRACTICAL APPLICATION

Let's focus on possible practical implications. One example that vividly presents what benefits spring from introduction of semantic mechanisms is as follows. Imagine you are looking for a pharmacy that is open on Sunday until 8 o'clock and is located within the perimeter of 2 kilometers of your location. You have different options, but at present, none of them offers a quick way to achieve this goal. You may access map sites and try to find pharmacies manually. This is however a cumbersome process, because you need to search the map's database and access each returned result to see if it matches the attributes you are looking for. Moreover, every pharmacy has a different website, with different structure and different scope of information, demanding to scan each one for the same pieces of information. The other option would be to use a search engine. However, typing "pharmacy open on Sunday until 8pm within 2 kilometers from my location" will give mostly irrelevant results, precisely because what search engines do, is looking for strings of characters. For these mechanisms, the only relation between the terms "open" or "pharmacy" is the relation of strings of characters in the query and those appearing across the Internet.

In a semantic search engine however, the sought after entity is not merely a group of letters arranged in a specific way. A "pharmacy" will have specific properties. It is a class of entities which have several attributes such as: opening hours, location, days of operation, brand and many more. The search engine will then "understand" that "open on" relates to opening time and that "within 2 kilometers" is a distance of the provided location calculated on basis of the location of the pharmacy. All these things are inferred in real time and are not based on fixed, prewritten information. The engine no longer looks for mere letters, but scans across the class of pharmacies and checks which entities belonging to it satisfy the provided conditions. This provides precise, transparent results for a question which is actually based in natural language we use every day. Computer and internet technologies provide incredible capabilities; yet providing answer to such a simple sentence is still impossible. Semantic Web will change that.

The above mentioned example is obviously a very simple one, but we may easily imagine what extent of opportunities would spring from real-time reasoning on much greater scale. Portals such as Wikipedia would be able to provide direct answers to complex questions, instead of referring to entire sites, leaving the visitor with the task. Various branches of

business, such as the automobile industry, where millions of different objects and combinations of objects are dealt with could obtain a tool that creates syntheses of any

portion of data, however scattered or belonging to different sources. Scientific knowledge could be organized and analyzed on basis of any logical operations. The opportunities are immense and already today a growing number of industries, research operations and commercial enterprises are implementing semantic technologies. It is in fact the first real shift in formal organization of the World Wide Web since its creation.