

PREREQUISITES FOR CREATION AND MAIN IDEAS OF THE SEMANTIC WEB

Федосенко Никита, Груздо Ирина

Харьковский Национальный Университет Радиоэлектроники

Abstract

In this article we will discover how the semantic analysis works, types of semantic search and why we need Semantic Web. Also we will discover the main ideas and building blocks of the Web 3.0 standard, such as Resource Description Framework and Web Ontology Language.

Анотація

У даній публікації ми розглянемо основні моменти роботи семантичного аналізу, види семантичного пошуку і навіщо ж нам потрібна Семантична Павутина. Також ми розглянемо основні ідеї та компоненти стандарту Web 3.0, такі як RDF (середовище опису ресурсів) і OWL (мова опису онтологій).

Introduction

With the increasing amount of textual information, automatic text processing systems are becoming increasingly relevant. The main goal of such systems is the semantic analysis of the text. Semantic analysis of the text evaluates the number of words or phrases that determine the meaning of the text, its semantic core and statistical indicators. The statistical indicators consist of a variety of criteria, such as symbols, grammatical errors, unique and meaningful words.

Semantic search is one of the methods of information retrieval and is a process of searching documents by the semantic content. There are two main types of semantic search:

- Full-text search - search through the entire contents of the document using pre-built indexes.
- Searching by metadata is a search for certain document attributes that describe certain objects supported by the system. For example, the author, address, organization name, etc.

Using full-text search is a time-consuming and not optimized solution. The concept of using of metadata search is becoming popular nowadays. Although, this solution has some obstacles on its way. Most of today's Web content is suitable for human consumption. Typical uses of the Web today involve humans seeking and consuming information, searching and getting in touch with other humans. The software tools to support these activities are not particularly well developed; in fact, the main ones remain search engines. Moreover, the technology of these tools remains roughly the same, and Web content outgrows technological progress. In particular, information retrieval is not very well supported. The major obstacle is that, at present, the meaning of the Web content is not machine accessible [1], in the sense that computers cannot interpret words, sentences, and the relationships between them.

The concept of creating a network of websites' metadata was introduced by Tim Berners-Lee, one of the founders of the World Wide Web in 2001. His idea was promoted by the W3C consortium then. This idea was called Semantic Web or Web 3.0. Its main goal was to add some logic to the current Web, express the meaning of the data by a series of formal rules, which would make information accessible to machines. It would give machines the ability to query the data, independent of the form where this data was presented. By the way, the Semantic Web is just an extra layer over the current Web. The semantic web improves web technologies in order to generate, share and connect content through search and analysis based on the ability to understand the meaning of words, rather than on keywords or numbers.

Such solution introduces an extra layer on the top of XML, adding more meaning to encoded information. This layer is represented by Resource Description Framework (RDF), RDF Schema and Web Ontology Language (OWL).

RDF is a description language, which holds the semantics of data represented on the Web. Though, RDF is based on XML, it does not depend on XML syntax, that's a different layer of abstraction. RDF is based on three main concepts: resources, classes and properties. Resources are the describe things e.g. articles, books, authors. All of the resources have unique identifier called Uniform Resource Identifier (URI). Classes are the collections of objects with same properties. And properties describe the relation between classes or resources.

Web Ontology Language (OWL) is a knowledge representation language for authoring ontologies. Ontology itself is a formal naming and definition of the types, properties that really exist in a particular domain of discourse [2]. RDF and Web Ontology Language work together to provide the most precise description of web resources.

Combining this capability with natural language processing, in Web 3.0, computers can understand information like humans in order to provide faster and more relevant results. They become more intelligent to satisfy the needs of users. Also With Web 3.0, information is getting more connected thanks to semantic metadata. As a result, the user experience evolves to another level of connectivity that leverages all the available information.

From a mathematical point of view, the semantic web represents a graph with resources in vertices, and relations between resources in arcs. This global semantic network is a set of subsystems that reflects to some resource, idea or term.

In this publication, we cover the most fundamental concepts and abstracts of semantic analysis, semantic search and have a look on the Semantic Web concept. Also, we have discovered the main techniques of semantic search and building blocks of the Web 3.0 standard. We have learnt about Resource Description Framework language and Web Ontology Language presented by the Semantic Web network. On one hand, the Semantic Web will allow users to talk to machines like to real people, what can make searching of information a lot easier. But on the other hand, this solution requires all of the websites add metadata to be indexed in the Semantic Web. All in all, the Semantic Web seems to be a promising solution for process automation using artificial intelligence.

References:

1. Antoniou G, Van Harmelen F. A Semantic Web primer. Cambridge, MA: MIT Press, Apr 2004. (Preliminary version available at: <<http://www.ics.forth.gr/isl/swprimer/>>. [rev. 3 Jul 2004; cited 25 Apr 2005].).

2. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pontiac Grove, CA: Brooks Cole Publishing Co, 2000. Quoted in: Jacob EK. Ontologies and the Semantic Web. Bull Am Soc Inform Sci Technol. 2003 Apr/May; 29(4):19–22