Abstract

Semantic web is a kind of webs that is able to describe things to be understood by computers. Automatically answering any query without human interactions is one of the key challenges in computer science area. Semantics can help in answering such queries. Consequently, extracting information from unstructured documents and transforming them into semantic web
form is an important trend. Semantic web mining is a combination of two trends; semantic web and web mining. Our extracting and structuring system clarify the meaning of the web mining. The obtained data converted to the semantic web format. And so, the semantic web mining trend was illustrated. This paper concentrates on extracting data from the web page tables. Data on the Web in the HTML tables are mostly structured. However; we usually do not know the structure in advance. Thus, data of interest cannot be directly queried. Data extraction and structuring system is proposed to put data extracted into the semantic web form. After putting extracted data in the semantic web format, it can be queried using semantic web query language. Experimental results show that the data of interest can be located and build its new structure using semantic web.

Reference

Index Terms

Computer Science

Information Retrieval

Key words

- semantic web
- information extraction
- information structuring
- natural language processing
- natural language processing
- wrapper generation
- semantic web mining
- extracting and structuring data