
Abstract

With the rapid development of Internet, Web has been becoming a main information source through which we can obtain the useful information. Nowadays there are millions of Websites and billions of homepages in Internet. This explosive growth of information on the internet has greatly increased the need for Information Retrieval System such as Search engine.

Nowadays most popular search engines such as Google, Alta Vista and yahoo are all based on HTML documents. Despite the success of HTML-based keyword search engines shortcoming emerge inside them such as lack of semantics retrieval. These search engines have HTML file based web server model it possesses certain limitations. Extensible Markup Language (XML) has recently emerged as the document standard for representing and exchanging data on the Web. Now XML turns Web into a database. The database is database of xml websites. To help Web users to retrieve the useful information in XML documents rapidly has been becoming a hot topic.

The goal of the thesis is to develop the Search Engine for searching the websites in XML/XSL. It provides two level searches in comparison to the existing search engines. The two level searches comprise of basic search and refine search. The basic search is similar to the conventional HTML search engine. But due to website made in XML it also provides the semantic information of keyword to user.

One more functionality comes in refine level search where the user can refine his search according to DTD/Tags information given to him. In addition an efficient *Compressed Tries* data structure used to implement the indexer. It also frees the user from remembering the structure of XML document and writing the sophisticated queries for searching from XML documents.