Abstract

The plentiful content of the World-Wide Web is useful to millions. Information seekers use a search engine such as Google, Yahoo etc to begin their Web activity. Our aim is to make a search tool that is cost-effective, efficient, fast and user friendly. In response to a query, it should retrieve the most relevant information which has been stored into the database. It should
also be portable, so that it can easily be deployed at any platform without any cost and inconvenience. Our goal is to make a Web Search Engine that will retrieve the best matched WebPages in the shortest possible time. This paper proposes an algorithm for crawler in which crawler crawls the WebPages recursively and stores the relevant data in the database. The algorithm uses the basic principles of tree structure while maintaining the crawled data by the crawler to be used by the search engine. The proposed work makes the searching on the web more efficient. It uses the tree/node structure in the database which filters the searched word more efficiently and gives faster results to the user. The paper has also implemented the crawler indexing with tree structure using HTML based Update File at Web Server while making the crawling and searching more efficient.

Reference

- Changshang Zhou, Wei Ding, Na Yang, Double Indexing Mechanism of Search Engine based on Campus Net, Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC’06)

Index Terms

Computer Science
Information Retrieval
**Key words**

<table>
<thead>
<tr>
<th>Crawler Indexing</th>
<th>Tree Structure</th>
<th>World-Wide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>