Abstract

Now a day, the web search engine becomes a very important gateway for common people looking for useful information on the internet but due to the dynamic nature of the web, it is difficult to find the relevant documents to fulfill user requirements. For this purpose, search engine maintains the index of documents stored in the repository. When a user enters a query search engine searches the index in order to find the relevant documents to fulfill user requirements. The query topic relevance depends on the information stored in the index. The performance of search engine is depends on the powerful structure of the index. Generally, inverted index are based on the frequency of keywords present in number of documents. An improved indexing mechanism to index the web documents is being proposed to improve the efficiency of the search engine that keeps the topic id along with the document id and also inverted index for ancestors in the (term, d, t) form. The structure is implemented using Trie structure. The proposed method will efficiently store the documents and will make the search fast.


References


Index Terms

Computer Science  Information Sciences
Keywords

Search engine, Indexing, Web document, Repository, Inverted index