Abstract

In this era of digital tsunami of information on the web, everyone is completely dependent on the WWW for information retrieval. Most of the information is hidden behind the query interface. In the query interface the user types the keyword to access the web pages. These pages are known as the Hidden web, Invisible Web or Dark Web. Such kind of web pages cannot be indexed by the Search Engines. As these are not indexed by the search engines these pages cannot be returned and displayed to the users. This paper discusses the various reasons due of which they are not indexed by the search engines and the possible solutions for these reasons.

References

- Sriram Raghavan Hector Garcia-Molina Computer Science Department Stanford University Stanford, CA 94305, USA, "Crawling the HiddenWeb";
- Ping Wu Ji-Rong Wen, Huan Liu, Wei-Ying Ma "Query Selection Techniques for Efficient Crawling of Structured Web Sources";
Timestamp based Recrawling Technique (TSBCT)

- Jian Qiu, Feng Shao, Misha Zatsman, Jayavel Index Structures for Querying the Deep Web, Workshop on the Web and Databases (WebDB), 2003, 79-86

Index Terms

Computer Science

Algorithms

Keywords

Hidden Web Search Engine Surface Webquery Interface