Abstract

The majority of the websites encapsulating online information are dynamic and hence too sophisticated for many traditional search engines to index. With the ever growing quantity of such hidden web pages, this issue continues to raise diverse opinions between the research and practitioner among the web mining communities. Several aspects enriching these dynamic web pages are bringing more challenges day-by-day to index them. By explaining these aspects and challenges, in this paper we have presented a framework for dynamic web indexing. With the implementation of this framework and the results which we have found from it, all the necessary experimental setup and the developmental processes are explained. We have concluded by exposing a possible future scope through the integration of Hadoop-Mapreduce with this framework to update and maintain the index.
Crawling the Hidden Web: An Approach to Dynamic Web Indexing

References

- King-Ip Lin, Hui Chen. &quot;Automatic Information Discovery from the &quot;Invisible
Web”, Information Technology: Coding and Computing (ITCC’02), IEEE, 2002.

Index Terms

Computer Science
Information Retrieval

Keywords

Dynamic web pages  crawler  hidden web  index  hadoop