Abstract

Web crawling is the foremost step to perform the effective and efficient web content search so that the user will get the specific web pages initially in an indexed form. Web crawling is not only used for searching a webpage over the web but also to order them according to user interest. There are number of available search engines and the crawlers that accept the user query and provide the page search. But, there is still the requirement and scope to improve the search mechanism. In this present work, dynamic and user interest evolution based parametric approach is defined to perform the web crawling and to arrange the web pages in more definite way. In this work a layered approach is defined, in which the initial indexing will be performed based on the keyword oriented content match and later on the indexing will be modified based on user recommendation. The presented work will provide an recommendation based web page indexing so that effective web crawling will be performed.

References

- Hussein Issa Rutgers Business School, Rutgers University "Application of Duplicate Records detection Techniques to Duplicate Payments in a Real Business Environment"
- Hani Khoshdel Nikkhoo "The Impact of Near Duplicate Documents on Information Retrieval Evaluation" by A thesis presented to the University of Waterloo in fulfillment of the thesis requirement for the degree of Master of Mathematics in Computer Science Waterloo, Ontario, Canada, 2010

Index Terms

Computer Science Information Sciences

Keywords
Crawling Indexing Recommender system