Abstract

With the explosive growth of information in the Web, users face difficulties while finding their desired information. Search engine helps the user by retrieving useful information from this huge collection based on his/her search query and presents a list of relevant web pages as a search result. However, without proper ranking of pages in the result through the relevancy of pages to the search query, the user may need to explore the whole list for discovering the appropriate page(s), thereby involving huge search time. Although a number of ranking algorithms such as HITS, PageRank, Weighted PageRank and etc., are developed to assist the search engine, but none of them provides page ranking with high accuracy. In this paper, we propose a score-based web page ranking algorithm involving web content mining and usage information of the pages. Our algorithm considers both syntactical and semantic matches of the search query to the pages. For a web page, syntactical score is calculated based on the total number of exact matches of the search words in the page. Besides, semantic score is measured using synonym matches of the search words. Moreover, we incorporate the usage information of the pages as page popularity in order to comprise the user interest in the ranking order. The total relevant score of each page is calculated using the summation of the
A Score based Web Page Ranking Algorithm

syntactical and semantic scores of the page and its page popularity. Finally, the pages are
ranked according to their total relevant score. Based on several performance evaluation
measures, experimental results show considerable improvement in the page ranking using our
algorithm as compared to other known approaches.

References

- Han, J. , Kamber, M. (2006). Data Mining, Southeast Asia Edition: Concepts and
  Techniques. Morgan kaufmann.
  Surveys (CSUR), 32(2), 144-173.
  TREC-8. In TREC.
  Computer Networks and ISDN Systems, 30(1), 161-172.
  content-based metasearch engine for images. In Photonics West'98 Electronic Imaging
  (pp. 118-128). International Society for Optics and Photonics.
  patterns. In Proceedings of the 11th International Conference on Information Integration and
  Web-based Applications & Services (pp. 465-469). ACM. Cooley, R. , Mobasher, B. ,
  Conference on (pp. 558-567). IEEE.
- Ramulu, V. S. , Kumar, C. N. S. , Reddy, K. S. A Study of Semantic Web Mining:
  Integrating Domain Knowledge into Web Mining. International Journal of Soft Computing and
  Engineering (IJSCE), 2(3).
  the ACM (JACM), 46(5), 604-632.
  in a hyperlinked environment. In Proceedings of the 21st annual international ACM SIGIR
  conference on Research and development in information retrieval (pp. 104-111). ACM.
  HITS algorithm for detecting Web communities. Systems and Computers in Japan, 35(13),
  32-42.
A Score based Web Page Ranking Algorithm

- Karpeles M. 2009. Modeling and optimizing hyper textual search engines, based on the research of Larry page and sergey brin. Yunfei Zhao Department of Computer Science, University of Vermont Slide from spring.

Index Terms

Computer Science Algorithms

Keywords

HITS  PageRank  Score Based Page Rank (SBPR).