Abstract

Rating a page to be a best one, based only on Page Ranking algorithm of Brin and Page would be insufficient. This method relied totally on Link information alone. However, due to application of Soft Computing in Data Mining and Knowledge Discovery, machines were made more effective, additional features of a Page involving its indexing, terms used, capitalizations, anchor texts, hit information, etc. were considered. The classification problem helped to induce this to a great extent. The complexity of dealing with a large number of web pages on the net made researchers to think of solutions dealing with sampling pages randomly and then making an analysis of the features of these pages. Soft Computing techniques were used for analysis of the features of the page. These techniques involved Genetic Algorithms, Neural Networks, Fuzzy Logic and Rough sets. User’s profiles of pages were created from the retrieved ones. Good and bad Pages were categorised on the basis of the terms they contained and these profiles were preserved for further reference. Pages were compared with each other for their similarity using Jaccard score and Best First search algorithm with developed software agents. Adaptive methods were used. Such methods were close to the concept of Genetic algorithm applications. The frequency at which a user visited web pages was also considered as a parameter of interest. Techniques to generate features of pages using co-occurrence analysis were developed and web pages were classified based on machine learning. A good
method of rating a page provided benefits like relevance, efficiency and indirectly on a crawl priority of a search engine which was more preferred. The web content designed as on date is for human reading and not typically tractable for machines. The semantic web had to provide structured content by adding annotations. Tools were made available to do these conversions. User-generated metadata that expresses a user taste and interest was used to personalize information to an individual user. Specifically, a machine learning method that analyzed a corpus of tagged content was to be used to find hidden topics. It then used these learned topics to select content that matched a users' interest, thus returning best relevant information pages. Even though Google scholar does not use synonyms and is strict to article text for searching a document, the use of synonyms reduce irrelevant search, causes intent drifting but synonym discovery is context sensitive these features motivate the use of synonyms to expediate the search and to rank relevant documents at a higher position. Google and Wordnet use synonyms but no documentation mentions using combination of synonyms for a term to generate a better relevant search, The present paper will concentrate on presenting a developed search technique to find a best page based on synonyms. The technique is based on the concept of adaptive search using synonyms of a search keyword extracted from a dictionary. These synonyms are then combined in different sets and given to a search engine which will return most relevant documents required by the user at a higher ranking.

References
Finding the Best page using Synonyms

- M. J. Carman, M. Baillie, and F. Crestani, "Tag data and personalized information retrieval", in SSM, 2008, pp. 27–34.
- Yanhong Li, "Towards a Qualitative Search Engine", 1089-7801/98, IEEE INTERNET COMPUTING JULY • AUGUST 1998
- Hang Cui, Ji-Rong Wen, Jian-Yun Nie, and Wei-Ying Ma, Member, "Query Expansion by Mining User Logs", in IEEE Transactions On Knowledge And Data
Finding the Best page using Synonyms

Engineering, Vol. 15, No. 4, July/August 2003

Index Terms

Computer Science  Information Sciences

Keywords

Best Page  relevance  users' interest  synonyms  metadata