Abstract

The Web, containing a large amount of useful information and resources, is expanding rapidly. Web crawlers are one of the most crucial components in search engines and their optimization would have a great effect on improving the searching efficiency. Focused Crawlers can selectively retrieve Web documents relevant to a specific domain to build collections for domain-specific search engines. In this paper, we use a genetic algorithm with focused crawling for improving its crawling performance. Expands initial keywords by using a genetic algorithm for focused crawling. The results showed that our approach could build domain-specific collections with higher quality than traditional focused crawling techniques.

References

- Soumen Chakrabarti, Kunal Punera, Mallela Subramanyam, "Accelerated Focused Crawling through Online Relevance Feedback," in WWW2002, May 7-11, Honolulu, Hawaii,
USA 2002.
- Yuxin Chen, Edward A. Fox et. al "A Novel Hybrid Focused Crawling Algorithm to Build Domain-Specific Collections"; Virginia Polytechnic Institute & State University Blacksburg, VA, USA pp- 85, 2007

Index Terms

Computer Science
Algorithms
Keywords
  Crawling  focused crawling  Genetic Algorithm  web crawler