Abstract

Search Engine spam is a poison for the search engine. It is created by the search engine spammers for commercial benefits. It affects quality of search engine. Already there are many algorithms available for filtering the search engine spam. But the spammers are often changing the strategy for creating the search engine spam. So there is a need to detect it in efficient way. The proposed system detects the search engine spam using an integrated hybrid genetic algorithm based decision tree. The proposed system is compared with different criteria and is shown the best performance than other methods.

References


29. Vijay Krishnan and Rashmi Raj, Web spam detection with Anti-Trust rank, In AIRWeb'06, August 2006.
35. https://moz.com/top500
40. Larry Page, Sergey Brin, The PageRank citation ranking: bringing order to the web. 1999
43. Payam Refaeilzadeh, Lei Tang, Huan Liu, "Cross Validation" Arizona State University, 2008
45. Cristina Petri, "Decision Trees", Lecture notes 2010
48. Robert P.W. Duin,” Learned from Neural Networks”, Pattern Recognition Group, Netherlands, 2000
51. Gary Stein, Bing Chen, Annie S. Wu, Kien A. Hua, ”Decision Tree Classifier For Network Intrusion Detection With GA-based Feature Selection”, 2005

Index Terms

Computer Science  Algorithms

Keywords

Search Engine Spam, Decision Tree, Genetic Algorithm, Tabu Search, Spamdexing, Feature Selection, Metaheuristic Approach