Abstract

Web mining means searching the Web for find specific information. Web mining operation should be done in a way to give the best results to the user. Two of the best methods in this area are clustering and ranking Web pages. The hereby-proposed method is a new approach which is a combination of the above-mentioned methods. In the proposed method, first, the Web graph is clustered in two phases, based on structural equivalences; next, each cluster is scored according to its value; then, ranking is done on all present pages in the clusters; and, finally, the final rank of each Web page would be the result of multiplying these two values. In the end, Web pages will be presented to the user based on their final rank. The results obtained from the comparison of the proposed algorithm (GCRM) with other methods indicate a good performance of this algorithm in finding high quality Web pages. Since quality is the main parameter in Web mining, main effort in GCRM algorithm is on increasing the quality of found pages, where, according to the results in this area, GCRM has been successful.
- Douglas R. W. and Karl P. R., &quot;Graph and Semigroup Homomorphisms on Networks of Relations,&quot; Social Networks, pp. 193-234, 1983.
- Faust K. and Wasserman S., &quot;Blockmodels: Interpretation and evaluation,&quot; Social Networks, pp. 5-1, 1992.
- Grabmeier J. and Rudolph A., &quot;Techniques of Cluster Algorithms in Data Mining,&quot; Data Mining and Knowledge Discovery, pp. 303-360, 2002.
- Jessop A., &quot;Blockmodels with Maximum Concentration,&quot; European journal of operational research, pp. 56-64, 2008.
- Page L., &quot;The PageRank Citation Ranking: Bringing Order to the Web,&quot; Technical Report, Computer Science Department, Stanford University, 2000.
- Schaeffer S. E., &quot;Graph Clustering,&quot; Computer Science Review, Vol. 1, pp. 27–64, 2007.
- Shaojie Q., Tianrui L., Hong L. and Hongmei C., &quot;A New Blockmodeling based Hierarchical Clustering Algorithm for Web Social Networks,&quot; Engineering Applications of Artificial Intelligence, Vol. 10, No. 16, pp. 1-9, 2012.
A New Approach to Organize the Results of Searching the Web, using a Combination of Ranking and Genetic Structure-based Clustering


Zhang K., October 2007, Visual Cluster Analysis in Data Mining, PHD Thesis, Department of Computing Division of Information and Communication Sciences of Macquarie University.


Index Terms

Computer Science

Artificial Intelligence

Keywords

Web mining, search engines, clustering and ranking