Abstract

Web Documents are of vital importance as they provide information from basic to the core requirements of users. The web documents have similarity based on the content and information represented. The key information in the web documents can be analyzed and can be represented as features. This leads to the importance of using similarity measures for categorizing and representing the web documents back to the users. This is a vital challenge as the documents could be clustered and represented in multiple dimensions for retrieval. This paper performs a complete survey of different techniques that are available for retrieving the web documents. The paper also presents investigations of the different performance
parameters available for measuring the performance of results. The paper also presents research directions for using soft computing techniques for web document retrieval.

References

- http://docs.intersystems.com/cache41/wlk/wl_intro.html
- Zhang, Yanchun; Xu, Guan-Dong, "Using Web Clustering for Web Communities Mining and Analysis," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 20-31, 2008
- Agavriloaei Ioan, Alexandrescu Adrian, Craus Mitică, "Improving web clustering through a new modeling for web documents," International Conference on System Theory, Control, and Computing (ICSTCC), PP. 1 – 6, 2011
- Li Tao-Ying, Chen Yan, "Web Page Clustering Based on Searching Keywords," International Conference on Intelligent Computation Technology and Automation (ICICTA), Vol. 3, PP. 1163 – 1166, 2010
- Takahashi Kou, Miura Takao, Shioya Isamu, "Clustering Web Documents Based on Correlation of Hyperlinks," International Conference on Data Engineering, PP. 1225, 2005
- Yang Christopher C., Ng Tobun Dorbin, "Analyzing and Visualizing Web Opinion"
Certain Investigations for Retrieving Web Documents using Soft Computing Techniques

- Rongfei Jia, Maozhong Jin, Xiaobo Wang, ”Web Objects Clustering Using
Transaction Log" Third International Conference on Knowledge Discovery and Data Mining, PP. 182 - 186, 2010.


**Index Terms**

Computer Science, Computing Techniques

**Keywords**

Clustering, Feature Vector, Neural Network, Index, Dimension, Filters, Probability