S DIGITAL LIBRARY

Libraries &

Institutions

Special trial offer

Learn More

Search Advanced Search

Home Join Publications Conferences **Digital Library Build Your Career Certification & Training** Communities e-Learning Store Login

Ī

Digital Library Home Subscribe Search 🔻 Resources 🔻 Institutional and Library Resources Newsletter About the Digital Library



Information Technology and Applications, International Conference on 2005

Third International Conference on Information Technology and Applications (ICITA'05) Volume 1

Using Topic Keyword Clusters for Automatic Document Clustering

Third International Conference on Information Technology and Applications (ICITA'05) Volume 1

Using Topic Keyword Clusters for Automatic Document Clustering

Sydney, Australia July 04-July 07 ISBN: 0-7695-2316-1

Hsi-Cheng Chang, Hwa Hsia Institute of Technology Chiun-Chieh Hsu, Taiwan University of Science and Technology

DOI Bookmark: http://doi.ieeecomputersociety.org/10.1109/ICITA.2005.303

Data clustering is a technique for grouping similar data items together for convenient understanding. Conventional data clustering methods, including agglomerative hierarchical clustering and partitional clustering algorithms frequently perform unsatisfactorily for large text article collections, as well as the computation complexity of the conventional data clustering methods increase very quick with the number of data items. This paper presents a system for automatic document clustering by identifying topic keyword clusters of the text corpus. The proposed system adopts a multi-stage process. First, an aggressive data cleaning approach is employed to reduce the noise in the free text and further identify the topic keywords within the documents. All extracted keywords are then grouped into topic keyword clusters using the k-nearest neighbor graph approach and the keyword clustering function. Finally, all documents in the corpus are clustered based on the topic keyword clusters. The proposed method was assessed against conventional data clustering methods on a web news collection, indicating that the proposed method is an efficient and effective clustering approach.

This Article

Subscribers, please Login Purchase article: \$19 PDF IEEE Xplore Subscribers

RSS feed

Share

Email this Article to a friend

Bibliographic References

ASCII Text RefWorks Procite/RefMan

Digg Spurl Simpy Del.icio.us Furl Blink GGoogle Y!MyWeb

Search

Similar Articles Articles by Hsi-Cheng Chang Articles by Chiun-Chieh Hsu

Citation:

Hsi-Cheng Chang, Chiun-Chieh Hsu, "Using Topic Keyword Clusters for Automatic Document Clustering," icita, vol. 1, pp.419-424, Third International Conference on Information Technology and Applications (ICITA'05) Volume 1, 2005

Peer Review Notice | Give Us Feedback Usage of this product signifies your acceptance of the Terms of Use

FREE WHITEPAPER

Software quality optimization: balancing business transformation and risk

Sponsored by

GO NOW -

This site and all contents (unless otherwise noted) are Copyright © 2010 IEEE, All rights reserved.

Site Map | Privacy Policy | Nondiscrimination Policy | Contact Us