Abstract

Due to the swift growth of WWW there has been large volume of information is produced and shared by various administrations in nearly every business, industry and other fields. Due to this high explosion it’s really a big challenge to store, manage and access knowledge. Experts estimate that 80 to 90 percent of the data in any organization is unstructured. And the amount of unstructured data in enterprises is growing significantly. Often many times faster than structured databases. Unstructured data files often include text and multimedia content. Examples include e-mail messages, word processing documents, pdfs, videos, photos, audio files, presentations, web pages and many other kinds of business documents. A huge amount of information spread across the web poses a major challenge in identifying relevant information. Existing tools lack analysis and visualization capabilities and traditional result displays long list of documents instead of providing concrete answers. This paper discusses various methods, tools and techniques for mining unstructured data that enables better data analysis and visualization.
1. Data Mining with Big Data Xindong Wu, Fellow, IEEE, Xingquan Zhu, Senior Member, IEEE, Gong-Qing Wu, and Wei Ding, Senior Member, IEEE IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, JANUARY 2014

2. 10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH QIANG YANG Department of Computer Science Hong Kong University of Science and Technology Clearwater Bay, Kowloon, Hong Kong, China XINDONG WU Department of Computer Science University of Vermont 33 Colchester Avenue, Burlington, Vermont 05405, USA xwu@cs.uvm.edu International Journal of Information Technology & Decision Making Vol. 5, No. 4 (2006) 597–604 c_World Scientific Publishing Company

3. Evaluating the Effectiveness of Keyword Search William Webber Computer Science and Software Engineering The University of Melbourne Victoria 3010, Australia wew@csse.unimelb.edu.au

4. Integrating Predictive Analytics and Social Media Yafeng Lu, Robert Krüger, Student Member, IEEE, Dennis Thom, Feng Wang, Steffen Koch, Member, IEEE, Thomas Ertl, Member, IEEE, and Ross Maciejewski, Member.


7. Managing Unstructured Data With Structured Legacy Systems David A. Maluf david.a.maluf@nasa.gov Peter B. Tran peter.b.tran@nasa.gov NASA Ames Research Center Intelligent Systems Division Mail Stop 269-4 Moffett Field, CA 94035


9. F. S. Gharechhopogh Hacettepe University Department Computer Engineering Ankara, Turkey Z. A. Khalifelel IAU Branch of Shabestar Department of Computer Engineering Shabestar, Iran Analysis and Evaluation of Unstructured Data: Text Mining versus Natural Language Processing


12. A Benchmark Suite for Unstructured Data Processing Clinton Wills Smullen, IV Shahrukh Rohinton Tarapore Sudhanva Gurumurthi Department of Computer Science University of Virginia Charlottesville VA 22904 {cws3k,shahrukh,gurumurthi}@cs.virginia.edu

13. Storing of Unstructured data into Mongo DB using Consistent Hashing Algorithm Saranraj Sankarapandi PG Scholar, IIIT-Srirangam, Tiruchirapalli, Tamilnadu, India. Dr. M. Sai Baba Associate Director, RMG, Indira Gandhi Centre for Atomic Research, Kalpakkam, Tamilnadu, India. S. Jayanthi Assistant Professor, Department of Computer Science &
Index Terms

Computer Science

Databases

Keywords

Unstructured data, structured data, data mining, text mining, machine learning, DGE model.