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
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.

5. Handling Unstructured Data for Semantic Web – A Natural Language Processing 
Approach Hemant Kumud Scholars Journal of Engineering and Technology (SJET) ISSN 

6. A Recent Survey on Unstructured Data to Structured Data in Distributed Data Mining 
ISSN:2229-6093

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

8. Approaches for Managing and Analyzing Unstructured Data N. Veeranjaneyulu, M. 
Nirupama Bhat, A. Raghunath School of Computing, Vignan’s University, Guntur, India 
International Journal on Computer Science and Engineering (IJCSE)

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

10. The Case for a Structured Approach to Managing Unstructured Data AnHai Doan, 
Jeffrey F. Naughton, Akanksha Baid, Xiaoyong Chai, Fei Chen, Ting Chen, Eric Chu, Pedro 
DeRose, Byron Gao, Chaitanya Gokhale, Jiasheng Huang, Warren Shen, Ba-Quy Vuong 
University of Wisconsin-Madison

11. The Acumen and Acuity of Data Mining and Text Mining Anjali Jivani, Jait Purohit, 
Kaushal Patel Comp Science & Engineering MSU Baroda, India Volume 4, Issue 11, November 
2014 ISSN: 2277 128X Available online at: www.ijarcsse.com

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 MongoDB 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.