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.

References
1. Data Mining with Big Data Xindong Wu, Fellow, IEEE, Xingquan Zhu, Senior Member, IEEE, Gong-Qing Wu, and Wei Ding, Senior Member, IEEE.

2. 10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH Qiang Yang

3. Evaluating the Effectiveness of Keyword Search William Webber

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

6. A Recent Survey on Unstructured Data to Structured Data in Distributed Data Mining Padmapriya et al

7. Managing Unstructured Data With Structured Legacy Systems David A. Maluf

david.a.maluf@nasa.gov

8. Approaches for Managing and Analyzing Unstructured Data N. Veeranjaneyulu, M. Nirupama Bhat, A. Raghunath School of Computing, Vignan’s University, Guntur, India

9. F. S. Gharechopogh Hacettepe University

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, Jianzheng Huang, Warren Shen, Ba-Quy Vuong


12. A Benchmark Suite for Unstructured Data Processing Clinton Wills Smullen, IV Shahrukh Rhointon 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 &
Survey on Data-intensive Applications, Tools and Techniques for Mining Unstructured Data


Index Terms

Computer Science

Databases

Keywords

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