

{tag}

{/tag}

International Journal of Computer Applications

© 2013 by IJCA Journal

Volume 66 - Number 23

Year of Publication: 2013

Authors:

Rupali Sunil Wagh

10.5120/11258-6501

{bibtex}pxc3886501.bib{/bibtex}

## Abstract

Last few decades have witnessed exponential increase in the use of IT which has resulted into large amount of data being generated, stored and searched. Data may be highly structured stored as records of a DBMS, or may be totally unstructured like blog posts or plain text documents. With the abundance of information being available as text documents, the issue of retrieval of knowledge from such unstructured dataset is posing new challenges to the research community. Legal document analysis is one domain which generates and uses text information in semi structured as well as unstructured form. The process of legal reasoning and decision making is heavily dependent on information stored in text documents. Text Mining (TM) is defined as the process of extracting useful information from text data. Legal text documents are stored using natural languages. For efficient analysis of such documents, text mining, a specialized branch of machine learning can be suitably used. Text mining – which "mines text", is heavily associated with natural language processing and Information Retrieval. TM techniques can be used for extracting relevant knowledge from stored legal documents. The extracted knowledge is used to simplify the preparation of case base, facilitate in decision making and legal reasoning or for automatic identification of legal arguments. Research in the fields of information extraction, natural language processing, artificial intelligence and expert system has augmented text mining process for enhancing the knowledge discovery process in this domain. This paper proposes a study which is aimed at grouping of legal documents based on the contents without taking any external input using unsupervised text mining techniques.

ences

- Kong Yanqing and Guoliang Shi Guoliang, Advances in Theories and Applications of Text Mining. The 1st International Conference on Information Science and Engineering (ICISE2009)
- K. A Vidhya and Aghila G, "Text Mining Process, Techniques and Tools : an Overview", International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 613-622
- Merkl Dieter and Schweighofer Erich "En Route to Data Mining in Legal Text Corpora: Clustering, Neural Computation, and International Treaties", 0-8186-8147-0/97 IEEE 1997
- Cheng Tin Tin, Leonard Cua Jeffrey, Davies Tan Mark, Gerard Yao Kenneth and Edita Roxas Rachel. Information Extraction from Legal Documents, 2009 Eighth International Symposium on Natural Language Processing, 2009 IEEE
- Joshi Sachindra, Deshpande, Prasad M and Hamppp Thomas. Improving the Efficiency of Legal E-Discovery. 2011 Annual SRII Global Conference, DOI 10. 1109/SRII. 2011. 97
- Ismael Hasan, Javier Parapar, Roi Blanco. Segmentation of legislative documents using a domain-specific lexicon. 19th International Conference on Database and Expert Systems Application, DOI 10. 1109/DEXA. 2008. 45
- Palmirani Monica and Brighi Raffaella. Metadata for the Legal Domain. Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03)
- Dozier Christopher and Jackson Peter, "Mining text for expert witnesses", 2005 IEEE
- Roitblat Herbert L. , Kershaw Anne and Oot Patrick. "Document Categorization in Legal Electronic Discovery: Computer Classification vs. Manual Review", JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 61(1):1-11, 2010
- S´anchez D. , Mart´n-Bautista M. J. , Blanco I. , C. Justicia de la Torre. Text Knowledge Mining: An Alternative to Text Data Mining. 2008 IEEE International Conference on Data Mining Workshops
- John Atkinson-Abutridy, Chris Mellish, and Stuart Aitken, "Combining Information Extraction With Genetic Algorithm for Text Mining", IEEE INTELLIGENT SYSTEMS
- Li Yaxiong, Zhang Jianqiang, Dan Hu. Text Clustering Based on Domain Ontology and Latent Semantic Analysis. 2010 International Conference on Asian Language Processing, DOI 10. 1109/IALP. 2010. 55
- Mirco Speretta and Susan Gauch. Using Text Mining to Enrich the Vocabulary of Domain Ontologies. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, DOI 10. 1109/WIIAT. 2008. 288

**Index Terms**

Computer Science

Information Sciences

**Keywords**

Text mining legal document search legal databases unsupervised learning