

Editorial Preface

Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies and Research, Department of Computer Science and Engineering, Delhi, India

Information retrieval is the activity of obtaining information resources relevant to information needed from a collection of information resources. However, the extraction of information should assist the common man to perform crucial decision making in even day to day activities. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links. Keyword searching has been the dominant approach to text retrieval; hypertext has so far been confined largely to personal or corporate information-retrieval applications. Evolving information-retrieval techniques, exemplified by developments with modern Internet search engines, combine natural language, hyperlinks, and keyword searching. Other techniques that seek higher levels of retrieval precision are studied by researchers involved with artificial intelligence.

The aim of the special issue is providing a quality publication with innovative ideas and implementation methodology to upcoming buddy researchers and users in the modern day era.

The unique characteristics of the special issue would be:

1. The proposed work of eminent researchers in the aspect of information system like Industrial systems Evolutionary computation, Autonomic and autonomous systems, Bio-technologies, Knowledge data systems, Mobile and distance education, Intelligent techniques, logics, and systems, Knowledge processing, Information technologies, Internet and web technologies, Digital information processing, Cognitive science and knowledge agent-based systems, Mobility and multimedia systems, Systems performance, Networking and telecommunications, Software development and deployment, Knowledge virtualization, Systems and networks on the chip, Context-aware systems, Networking technologies, Security in network, systems, and applications, Knowledge for global defense, Information Systems [IS], IPv6 Today - Technology and deployment, Modeling, Optimization, Complexity, Control theory and systems, Fault-tolerance and reliability, Data engineering, Enterprise computing and evaluation, Electrical and electronics engineering, Economic decisions and information systems, Intelligent agent technologies, Intelligent and fuzzy information processing, Intelligent computing and knowledge management, Intelligent systems and robotics, Fault-tolerance and reliability, Fuzzy logic & systems and Genetic algorithms which are current topics of research will be part of proposed publication;
2. The proposed publication will be very well targeted towards providing quality, best and latest research by eminent researchers considering the fact that how such researches affect and make significant influences on common people in their everyday life;
3. The area which will be part of published work will be having a significant influence for the business users, common people and has a great impact on the society.

This special issue is a collection of the five papers which are written by eminent professors, researchers and Industry people from different countries. The papers were initially peer reviewed by the Editorial board members, reviewers and industry people who themselves span over many countries.

In the paper, Off-line Versus On-line Sentiment Analysis: Issues with Sentiment Analysis of On-line Micro-texts, this paper argues the non-applicability of the classical time consuming Natural Language Processing (NLP) methods and the affinity of Machine Learning (ML) methods in performing the on-line sentiment analysis by contrasting it with off-line sentiment analysis. Furthermore, it also formalized the on-line sentiment analysis process of on-line micro-texts by raising novel issues and proposing new performance measures for on-line sentiment analysis. This paper has attempted to formalize the real-time process of on-line sentiment analysis, particularly for on-line micro-texts by considering on-line micro-texts as a textual data stream. This paper also provides a description for difficulties in the application of classical NLP methods to the on-line sentiment analysis. This paper also discussed some new performance measures that are needed to consider in on-line sentiment analysis other than the accuracy of classification. Furthermore, a new concept related to on-line sentiment analysis i.e. sentiment drift has been defined and discriminated by concept drift in data stream mining. The significance of window-based approaches for detecting the sentiment change and aggregating the overall result of on-line sentiment analysis process in real-time has been justified in this paper very well.

In the paper, Semantic Analysis Based Approach for Relevant Text Extraction Using Ontology, authors gave a novel technique for retrieving the relevant semantic information represented in the form of ontology for true semantic analysis of given text. The semantic text extraction approach given in this paper helps in providing the relevant and significant text from a document which further gives the prime intention of the author of the document. This is done by extracting the relevant concepts and relationships between these concepts from the text of the document. These semantically extracted interrelated concepts are used to construct the pattern clustered of semantically associated concepts which are further represented in the form of ontology for the document. Finally, the similarity index between any given texts is calculated using the obtained document ontology's. The proposed technique in this paper is used to compute similarity between web documents by extracting the relevant text from the respective web document. However, it can also be used in the information retrieval applications like crawling, indexing, ranking etc. to extract the semantically similar words/text. The proposed technique outperformed on the benchmark dataset as compared to the existing similarity techniques. Further, we will attempt to build our technique more exhaustive by designing a system for automatic construction of domain related base ontology for efficient retrieval of semantic association from each document. Additionally, the document ontology of the semantically extracted texts from each document can also be constructed automatically. We will intend to apply our approach in automatic semantic text extraction environment by designing a ranking algorithm for the semantic search engine to extract the semantically associated web documents with respect to the query given by the search engine.

In the paper, Frequent Itemset Mining in Large Datasets A Survey, in this paper a survey of frequent itemset mining techniques is done which can be used in a parallel environment. Programming models like Map Reduce provides efficient architecture for working with Big Data, paper also provides information about issues and feasibility about technique to be implemented in such environment. Size of data is generating challenges for frequent itemset mining algorithm and association rule generation. Parallel implementation of these algorithms can provide solution for Big Data frequent itemset mining. The challenges in mining frequent itemset are like cardinality estimation, global count management; I/O cost and inter process communication. Combination Map Reduce and HDFS provide an efficient framework for parallel implementation of frequent itemset mining algorithms. There are some issues for implementing FIM algorithms using Map Reduce like mappers task distribution, distribution of data to the mappers, inter-mapper communication (if required), reducer processing of data, iterative calls to the mappers, mappers output mixture and the removal of dependencies between mappers. Selection of solution is based on the problem domain and the nature of the data. The paper provides information about different FIM algorithms and feasibility of their Map Reduce implementation. Analysis of the existing approaches shows that algorithms dependencies on complete structure of data are a critical issue while FIM. Resolving this issue can result in a better and efficient FIM. Future

aspect of FIM on Big Data is the development of fast algorithms which can mining item sets and generation of association rules from them. Distribution of data and distributed processing of data comes with its complexities like, performing mining tasks, proper storage, replication, information sharing, load balancing, inter process communication etc. new algorithm needs to overcome these problems.

In the paper, Geo-Tagging News Stories using Contextual Modeling, author developed a mathematical model of context and geo-tagging with respect to news stories and have exploited that model to geo-tag news stories even in the absence of direct mentions of locations as well as at the granularity of street/locality level. For this, we have incorporated our model with a geo-tagging algorithm and utilised off-the-shelf tools and existing Geocoding APIs. The data set generated after applying our approach has been evaluated with 6 users. In addition, we have evaluated our approach over 10,000 news from the Reuter data set. The results demonstrate the effectiveness of our approach against existing publicly available APIs. The ultimate goal of the approach is to integrate our news data collection with an array of multi-modal data from different social media such as Flickr, Twitter, YouTube as well as from different wearable sensors such as life loggers and GPS trackers and to develop a location-based information fusion system where locations, among other factors, will act as the glue to bind together all data sets. In this regard, the proposed approach will be an essential ingredient of the system. However, we believe that the approach can be adapted for geo-tagging any news stories in any other scenarios with little or no further modifications.

In the paper, Hadoop Map Only Job for Enciphering Patient-Generated Health Data, author make use of only mapper job framework for data encryption All the health care data such as discharge and transfer patient data maintained in Computer based Patient Records (CPR), Personal Health Information (PHI), and Electronic Health Records (EHR). The use of Big Data analytics is becoming increasingly popular at health care centers, in clinical research, and consumer based medical product development. The biggest challenge with implementation of big data is that the nature of information of public health sector is of very sensitive nature and needs to be protected from unauthorized access and release of contents. Due to the sheer volume of data traditional systems have proven to be ineffective in handling and encryption of such huge amounts of data effectively. Therefore, to provide solution to the deidentifying personal health big data in this paper we author made use of the only mapper job framework for data encryption.

Vishal Bhatnagar
Editor-in-Chief
IJIRR