Guest Editorial Preface

Special Issue on Knowledge based Information Retrieval Using Data Analysis

Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

The usage of data analysis to uncover hidden and valuable information has been the key area of research in the modern day era. However, the extraction of information should assist the common man to perform crucial decision making in even day to day activities. The use of the web has already created an immense amount of data which is readily available to be analyzed for nuggets of information. This type of data analysis is done by researchers, scientists, and even data analysts in everyday affairs. However, usage of knowledge which is extracted from the routine activities of users of the web, including the interest or preference of the individual or the qualification/demographic details of the person, helps in understanding the basic mindset of the users who are serving the web for many crucial things. The inclusion of usage of semantics in data analysis has created a boom in the data analysis field. The sheer meaning of a sentence or an opinion of an individual or a group of people has impacted many studies and revealed new information which has proven to be helpful in many areas.

Areas like click stream analysis, which helps to understand the preference of the user, has been analyzed to find information which could be made easily available to the individual quickly. The usage of social network mining helps to create groups which consist of people who are like minded and have common thoughts and agendas.

Knowledge based information retrieval has found applications in various domains and fields, including Sentiment Analysis, Big Data Analysis, Collaboration is Filtering and Crime Investigations, to name a few. The main focus of this special issue is to provide the latest advancements in the problem domain of Knowledge based information retrieval using data analysis tools and techniques which are identified by various researchers in the field of Intelligent Information retrieval by developing improved techniques and algorithms.

In the First article, Sarna and Bhatia discussed that users on the social media can share positive as well as negative information intentionally and unintentionally in the form of multimedia content without knowing its impact on other user, which threatens the security and privacy of social media. Cyberbullying is one of the risks associated with social media. Cyberbullying is an aggressive act carried out intentionally against the victim by posting harmful material on social media to harm his/her reputation. Aggressive act creates depression, anxiety in users which may lead to diversion of attention and sometimes suicidal actions. In this paper we have included a survey on recent algorithms which work on detection of cyberbullying. State-of-the-art studies only focus on the detection of cyberbullying but not on the preventive measures against cyberbullying. In order to tackle this problem, we showed how the severity of the bullying in messages helps to find the real culprit.

In the Second article, Meddah and Khaled presented that process mining provides an important bridge between data mining and business process analysis, the process mining techniques allow for

extracting information from event logs. In general, there are two steps in process mining, correlation definition or discovery and then process inference or composition. Firstly, work consists to mine small patterns from a log traces, those patterns are the representation of the traces execution from a log file of a business process. In this step, author uses existing techniques; The patterns are represented by finite state automaton or their regular expression; The final model is the combination of only two types of small patterns whom are represented by the regular expressions (ab)* and (ab*c)*.In second step authors computed these patterns in parallel, and then combine those small patterns using the composition rules, we have two parties the first is the mine, we discover patterns from execution traces and the second is the combination of these small patterns. Results are general and precise. It minimizes the execution time.

In the third article Jain and Bhatnagar canvassed that Data is continuously snowballing over the years, gradually a huge growth is seen in data to store and tame to yield meticulous result. It gives rise to a concept nowadays, reckoned as big data analytics. With the summer Olympics at Rio de janeiro, Brazil in the year 2016 round the corner, the authors have implemented a mathematical model by implementing efficient map reduce program to predict the number of medals each country might bag at the games. Based on a number of factors such as historical performance of the country in terms of medals won, the performance of athletes, financial scenario in the country, fitness levels and nutrition of athletes along with familiarity to the playing conditions can be used to come up with a reliable estimate.

In the fourth article Walia and Bhatia elaborated that with the advent of web 2.0 and anonymous free Internet services available to almost everyone, social media has gained immense popularity in disseminating information. It has become an effective channel for advertising and viral marketing. People rely on social networks for news, communication and it has become an integral part of our daily lives. But due to the limited accountability of users, it is often misused for the spread of rumours. Such rumour diffusion hampers the credibility of social media and may spread social panic. Analyzing rumours in social media has gained immense attention from the researchers in the past decade. In this paper we provide a survey of work in rumour analysis, which will serve as a stepping-stone for new researchers. We organized the study of rumours into four categories and discussed state of the art papers in each with an in-depth analysis of results of different models used and a comparative analysis between approaches used by different authors.

In the fifth article Goyal and Bhatnagar intricate that the recent growth of e-commerce websites have paved a way for the users to express their opinions on these web portals which, in turn, makes the customers review these comments before buying any product or service. The comprehensive reading of this large number of reviews is cumbersome and tiring. The purpose of this paper is to perform the analysis on the tourism domain reviews to decide whether the document is positive or negative. The traditional methods use a machine learning approach, but we are using an unsupervised dictionary based approach to classify the opinions. The scores of the opinions are extracted using Sentiwordnet, a popular dictionary for calculating the sentiment.

In the sixth article Samuel and Sharma presented that Summary generation is an important process in those conditions where the user needs to obtain the key features of the document without having to go through the whole document itself. The summarization process is of basically two types: 1) Single document Summarization and, 2) Multiple Document Summarization. But here the microblogging environment is taken into account which have a restriction on the number of characters contained within a post. Therefore, single document summarizers are not applicable to this condition. There are many features along which the summarization of the microblog post can be done for example, post's topic, it's posting time, happening of the event, etc. This paper proposes a method that includes the temporal features of the microblog posts to develop an extractive summary of the event from each and every post, which will further increase the quality of the summary created as it includes all the key features in the summary.

As guest editors, we hope that spectrum of research work covered under this special issue will be of value for multitude of readers/researchers. At the same time, we are also grateful to the authors for making their valued research contribution to this issue and their patience during crucial revision stages. The technical standard and quality of published content is based on the strength and expertise of the submitted papers. Our special thanks go to the Editor-in-chief of the International Journal Of Rough Sets and Data Analysis (IJRSDA), Dr. Ahmad Taher Azar, Benha University, Egypt for all his support, efficiency, and competence rendered to this special issue.

Vishal Bhatnagar holds a BTech, MTech and PhD in the engineering field. He has more than 16 years of teaching experience in various technical institutions. He is currently working as an Associate Professor in Computer Science in Engineering Department at Ambedkar Institute of Advanced Communication Technologies and Research (Government of Delhi), GGSIPU, Delhi, India. His research interests include database, advance database, data warehouse, data-mining, social network analysis and big data analytics. He has to his credit more than 80 research papers in various international/national journals and conferences.