ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Text Mining Research: A Survey

R.Janani[1], Dr. S.Vijayarani[2]

PhD Research Scholar, Dept. of Computer Science, School of Computer Science and Engineering, Bharathiar University, Coimbatore, India[1]

Assistant Professor, Dept. of Computer Science, School of Computer Science and Engineering, Bharathiar University, Coimbatore, India[2]

ABSTRACT: Data Mining is the process to extract hidden predictive information from database and transform it into understandable structure for future use. The assorted domains in data mining are Web Mining, Text Mining, Sequence Mining, Graph Mining, Temporal Data Mining, Spatial Data Mining (SDM), Distributed Data Mining (DDM) and Multimedia Mining. Some of the applications of data mining, it is used for financial data analysis, retail and telecommunication industries, science and engineering and intrusion detection and prevention. In this paper we discussed about the text mining techniques and its applications. Text mining is used to extract interesting information or knowledge or pattern from the unstructured texts that are from different sources. It converts the words and phrases in unstructured information into numerical values which may be linked with structured information in database and analyzed with ancient data mining techniques. There are many techniques used in text mining such as information extraction, information retrieval, natural language processing (NLP), query processing, categorization and clustering.

KEYWORDS: Data Mining, Text Mining, Natural Language Processing, Techniques, Issues, Challenges

I.INTRODUCTION

Text Mining is also called as text data mining, which is used to find interesting information from large database. The data mining tools can be designed to handle structured data from database. But the text mining can be work with unstructured data or semi-structured data [1]. Text Mining is to analyze large quantities of natural language text and it detects lexical patterns to extract useful information. Text Mining is useful for organization because most of the information is in text format. The following steps can be included in text mining.

- It converts the unstructured text into structured data
- Identify the patterns from structured data
- Analyze the patterns using Text Mining techniques
- Extract the useful information from the text

The applications of Text Mining are protein interaction, drug discovery, predictive toxicology, identification of recent product potentialities, detection of links between lifestyle and states of health, competitive intelligence and lots of additional [3]. Section II describes the process of text mining. Section III illustrates the techniques in text mining and section IV gives the applications of text mining. Section V explains the issues in text mining and section VI gives the conclusion of this paper.

II. PROCESS OF TEXT MINING

A. Document Gathering

In the beginning step, the documents are collected that are present in numerous formats. The document could be in form of word, html, css, pdf etc. [17].

B. Document Pre Processing

In this process, the given document is processed for eliminating redundancies, inconsistencies, separate words, stemming and documents are ready for next step, and the stages performed as follows[19],

Tokenization: The given document is recognizing as a string and identifying single word in document i.e. the given document string is distributed into one unit or token.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Removal of Stop word: In this stage the removal of common words like a, an, and, of, but etc. has been done.
Stemming: A stem may be a natural group of words with very similar meaning. This method defines the base of the particular word. There are two types of stemming method, Inflectional and Derivational [21].

C. Text Transformation

Text document contains a collection of words and their occurrences. There are two ways for representation of such documents is Bag of words and Vector space model.

D. Attribute Selection

This method leads to giving less database space, minimal search technique by removing irrelevant feature from input document. There are two methods in attribute selection, Filtering and Wrapping methods [13].

E. Pattern Selection

In this stage the standard data mines process combines with the text mining process. Structured database use the classic data mining techniques that resulted from the previous stage. To identify the correctly interesting patterns representing knowledge based on some interestingness procedures [3].

F. Interpretation/ Evaluation

In this stage measures the result, this result can be put away or it will be used for next set of sequence.

III. TECHNIQUES IN TEXT MINING

The techniques in text mining from different areas such as information extraction, information retrieval, natural language processing (NLP), categorization and clustering. All these stages of text mining process can be combined into a single workflow. Figure 1 shows the text mining techniques.

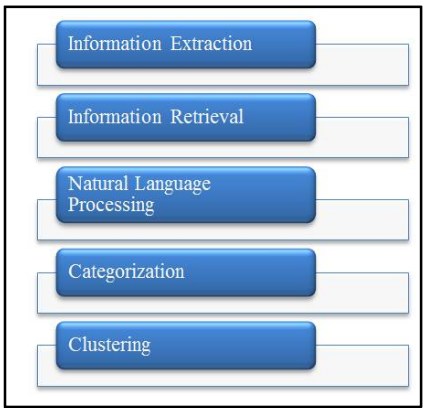

Fig 1: Techniques in Text Mining

A. Information Extraction

Information Extraction is the process of extracting information from unstructured document using the techniques. The main objective of information extraction is to find specific data or information in natural language text. This information is stored in database like patterns and it can be available for further use. It converts a quality of textual documents into more structured database [9]. The major task of information extraction,

- Term Analysis**:** It identifies the term; the term may contain one or more words. It can be helpful for extracting information from documents
- Named Entity Recognition: It identifies the textual information in a document relating the names of person , place , organization or product
- Fact Extraction: It identifies and extracts the complex facts from the documents

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

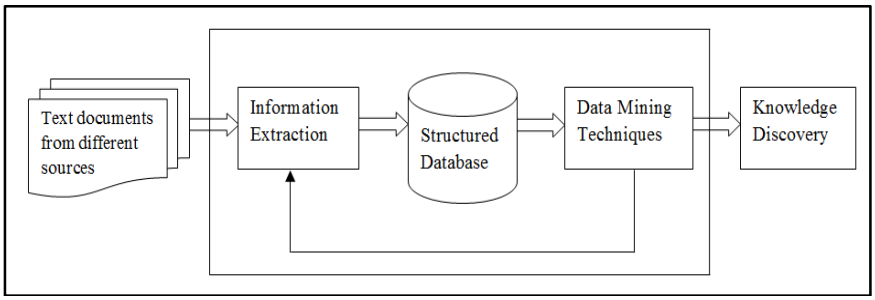

Fig 2: Information Extraction

Table 1: Comparative Analysis of Information Extraction Techniques

S. No	Research Paper Name	Techniques Used	Best Technique	Research Challenges
1	Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization [26]	Stemming, Domain Dictionary, Execution List	Domain Dictionary	It is important to go to the core part of pattern mining algorithms, and analyze the theoretical properties of different solutions.
2	An analysis on text mining- text Retrieval and text extraction [27]	Text Preprocessing, Rule Selection, Rule Application	Rule Selection	In this research paper, they have used three techniques for IE process. Integrating a domain knowledge base with a text mining engine would boost its efficiency, especially in the information retrieval and information extraction phases
3	Information extraction methods and extraction Techniques in the chemical Document's contents: survey [14]	Lexical and Syntactic Analysis, Rule based IE system	Lexical and syntactic analysis	The IE processes are very complicated because they are mainly based on the automatic recognition of human language terms. This challenge is to motivate researchers to work hard in this field to provide appropriate solutions for enhancing the automation process of IE systems.
4	Information Extraction: Techniques and Challenges[21]	Pattern Matching, Lexical Analysis, Name recognition, syntactic structure, coreference analysis, event merging	Lexical Analysis	In this analysis, they have used many techniques for information extraction. In that lexical analysis gives the best result when compared to other techniques. Still the researchers enhance these techniques for future use.
5	Information extraction from Text	Rule based approach, Statistical learning approach	Statistical learning approach	Information extraction is an important text mining problem. With the fast growth of textual data on the Web, it is expected that future work on information extraction will need to deal with even more diverse and noisy text.

B. Information Retrieval

It is used to identify the relevant documents in a document collection which is matching a user's query. The most important application of information retrieval system is search engine like Google, which identify those documents on the World Wide Web that are relevant to user queries or a set of given words [9]. It also refers to the automatic retrieval of documents from document collection.

It deals with crawling, indexing documents and retrieving documents. Information retrieval system used in digital libraries, online document systems and search engine. Information retrieval is deals with entire range of information processing from data retrieval to knowledge retrieval. Figure 3 explains about the information retrieval system.

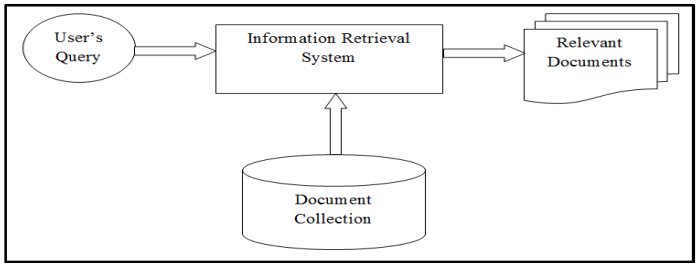

Fig 3: Information Retrieval

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

Table 2: Comparative Analysis of Information Retrieval Techniques

S. No	Research Paper Name	Techniques Used	Best Technique	Research Challenges
1	Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization [26]	Stemming, Domain Dictionary, Execution List	Domain Dictionary	The domain dictionary which defines the set of terms consists of all feature terms is the essence of such mining tools. A lot of work can be done to further improve and extend this implementation.
2	A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques [9]	Retrieval algorithm, filtering algorithm, indexing algorithm	Retrieval Algorithm	In this research work, the variety of algorithms used for information retrieval. Compare to all other algorithms retrieval algorithm is best for retrieving the information from documents. Enhancing domain knowledge with text mining engine would improve the efficiency, especially in the information retrieval phase.
3	Data Mining: a Healthy Tool for Your Information Retrieval and Text Mining [12]	Association rules, statistical analysis, full document text analysis	Association Rules	The main problem in text retrieval was natural language understanding barrier, which proved to be much more challenging in this field. The new generation of information retrieval tools will appear in near future.
4	A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques [9]	Boolean IR model, Vector space model (VSM)	Vector space model	VSM is more recent and advanced than Boolean IR model. The retrieved documents allowing easy deployment of advanced IR techniques. Disadvantages are indistinct relationship between relevance and similarity and unclear query term explication.
5	Information Retrieval Models	Vector Space Model, probabilistic retrieval model, Boolean Model	Boolean Model	The Boolean approach makes it possible to represent structural and contextual information that would be very difficult to represent using the statistical approaches.

C. Natural Language Processing (NLP)

It is concerned with interactions between computer and human (natural) languages. NLP is related to the area of human-computer interaction. NLP is the component of an Artificial Intelligence (AI) [12]. It is used to analyze the human languages so that computers can understand natural languages as humans do. The approaches to NLP is based on machine learning, a type of artificial intelligence that examines and uses the patterns in data to improve a program's own understanding [2]. The role of NLP in Text Mining is to provide the systems in the information extraction phase with linguistic data that they need to perform their task. NLP includes the tasks [11],

- Part Of Speech tagging: It is used to classify the words into categories such as noun, verb or adjective.
- Chunking: It is also called as shallow parsing, used to identify only the main grammatical elements in a sentence such as noun phrases and verb phrases.
- Semantic Role Labeling: It is used to detection of semantic arguments associated with predicate or verb of a sentence.
- Language Model: It assigns a probability of sequence of words by means of probability distribution. It provides context to distinguish between words and phrases that sound similar.
- Semantically Related Words: This is the task of predicting whether two words are semantically related which is measured using the WordNet database.

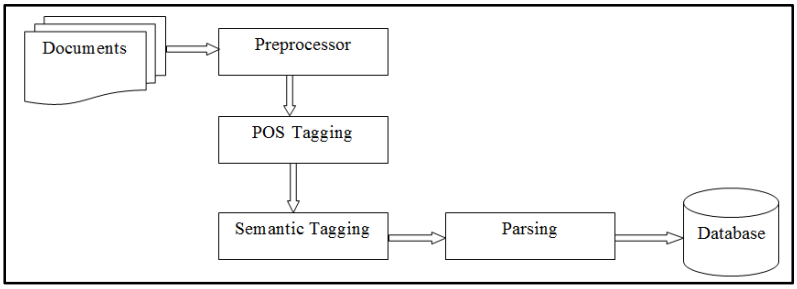

Fig 4: Natural Language Processing

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

D. Categorization

The process of categorization is, recognizing, differentiating and understanding the ideas and objects to group them into classes. A category clarifies the relationship between the subjects and objects of information or knowledge [4]. Categorization is crucial in language, prediction, decision making and in all kinds of environmental interaction. It involves identifying the important themes of a document by placing the document into a predefined set of topics.

Once the document is categorized, a computer program can typically treat the document as a "bag of words" [1]. It doesn't attempt to process the actual information as information extraction does. Relatively, categorization process only counts the words that appear in the document, and from the counts, it identifies the significant topics that the document covers. Categorization typically depends on a thesaurus for which topics are predefined, and relationships are identified by searching for broad terms, narrow terms, synonyms and related terms. Categorization tools commonly have a method for ranking the documents in order which documents have the most content on a particular topic [23].

It can be used in a variety of application domains. Many businesses and industries provide customer support or have to be compelled to answer queries on a number of topics from their customer. The main goal of categorization is to classify a collection of documents into a fixed number of predefined categories. Each document may belong to more than one class [16].

Table 4: Comparative Analysis of Categorization Techniques

S. No	Research Paper Name	Techniques Used	Best Technique	Research Challenges
1	Text Mining Techniques - A Survey [4]	Naïve Bayes Classifier, KNN	Naïve Bayes	Supervised technique having all the input output patterns which are used to train the model, before it can be used to classify the newly arrived document.
2	Text Classification And Classifiers: A Survey [23]	Bayesian classifier, Decision Tree, K-nearest neighbor(KNN), Support Vector Machines(SVMs), Neural Networks,	KNN	The performance of a classification algorithm is greatly affected by the quality of data source. Irrelevant and redundant features of data not only increase the cost of mining process, but also reduce the Quality of the result in some cases.
3	Recent Trends in Text Classification Techniques [3]	KNN,Bayesian Classification,SVM, Association Based Classification,Term Graph Model, Centroid Based Classification, Decision Tree Induction,Neural Network	SVM	One remarkable property of SVMs is that their ability to learn can be independent of the dimensionality of the feature space.
4	A Review of Machine Learning Algorithms for Text-Documents Classification [26]	Rocchio's Algorithm, KNN, Decision Tree, Naïve Bayes, Artificial Neural Network, Fuzzy Correlation, SVM, Genetic Algorithm	KNN	To improve and explore the feature selection methods for better classification process and to reduce the training and testing time of classifier and improve the classification accuracy, precision and re-call.
5	Text Mining Process, Techniques and Tools : an Overview [11]	Naïve Bayes, SVM, Decision tree classifier, Rocchio's Algorithm, KNN	KNN	Training time is relatively expensive and suffers from over fitting by which it is not able to handle continuous variable well.

E. Clustering

Clustering is a process of partitioning a set of data or objects into a set of meaningful sub-classes, is called clusters. This technique is used to group the similar documents. The benefit of clustering is that documents will seem in multiple subtopics, so ensuring that a useful document will not be omitted from search results. A basic clustering algorithm creates a vector of topic for each and every document and measures the weights of how well the document fits into each cluster [24]. Clustering technology can be helpful in the organization of management information systems, which can contain thousands of documents.

There are many clustering methods available and each of them may give a different grouping of a dataset. Clustering methods can be classified into two categories [22],

1. Hierarchical Methods
2. Non-Hierarchical Methods

Hierarchical Methods

Hierarchical clustering constructs a cluster hierarchy or, in other words, a tree form of clusters also called as Dendogram. Each cluster node contains child clusters; relation clusters partitioned the points covered by their common

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

parent. Such associate approach permits exploring data on completely different levels of granularity [20]. These methods are divided into Agglomerative (Bottom-UP) and Divisive (Top-Down) methods.

An **Agglomerative** clustering starts with one point clusters and recursively merges two or more with most appropriate clusters. A **Divisive** clustering starts with one cluster of all the data points and recursively splits the most appropriate clusters. This process continues until the requested number of clusters is achieved.

Non-Hierarchical Methods

The non-hierarchical methods divide a dataset of N objects into M clusters, with or without overlap. These methods are divided into partitioning methods, in which the categories are mutually exclusive and also the less common stamping methods [17]. Each object may be a member of the cluster with most similar.

Table 5: Comparative Analysis of Clustering Techniques

S. No	Research Paper Name	Techniques Used	Best Technique	Research Challenges
1	A tutorial review on Text Mining Algorithms [6]	Hierarchical method, Partitioning method	Partitioning Method	Partitioning Method is an unsupervised learning technique, because there is no predefined set of patterns are available in this method.
2	A Survey of Text Mining Techniques and Applications [3]	K-means, Word Relativity based Clustering	K-means	A K-means algorithm gives the better accuracy when compared with another algorithm. But this methods accuracy based on the dataset.
3	Text Mining Techniques - A Survey [4]	Hierarchical Clustering, K-means	K-means	It is an unsupervised learning technique in which, no pre-defined input-output patterns are there.
4	A Comparison of Document Clustering Techniques [24]	Intra-Cluster Similarity Technique (IST), Centroid Similarity Technique (CST), K-means, bisecting K-means	IST, bisecting k-means	This paper gives the better performance of bisecting K-means compared with regular K-means is due to fact that it produces relatively uniformly sized clusters instead of clusters of widely varying sizes.
5	Survey of Clustering Data Mining Techniques [25]	K-means, K-medoids, Density based algorithms, grid based clustering, Cobweb	Cobweb	This research paper they have used many algorithms for clustering the document. In that, the cobweb algorithm gives more accuracy when compared to other algorithms. But this algorithm has some demerits. Researchers can enhance these difficulties for future use.

IV. APPLICATIONS OF TEXT MINING

Text Mining has a very high commercial value. It is a developing technology for analyzing large amount of unstructured documents for the purpose of extracting interesting pattern or knowledge [1]. There are several domain specific applications of Text Mining, a number of applications had explained here:

A. Security Application

Many text mining packages are marketed for security applications, particularly observation and analysis of online plain text sources like web news, blogs, etc. for national security functions. It also concerned with the study of text encryption and decryption [14] [16].

B. Biomedical Application

Text Mining is used in medical specialty for identification and classification of technical terms within the domain of biological science corresponding to the concepts.

C. Company Resource Planning

Mining Company's reports and correspondences for activities, thus its resource status and problems will be handled properly and future action planned can be design.

D. Market Analysis

With the help of numerous text mining techniques, market analysis is concerned to analyze the competitors within the market and can also be used to monitor customer opinions and searing for new potential customers.

E. Customer Relationship Management

It mainly deals with managing the customer messages. CRM consists of providing applicable service to the customer as per their request and providing fast answers to their queries.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

V. ISSUES IN TEXT MINING

The main challenging issue in text mining arises from the complexity of a natural language itself. The natural language is not free from ambiguity problem. Ambiguity means the capability of being understood in one or more possible ways [4]. In a text document one word can have more than one meanings and one phrase or sentence can be interpreted in many ways which directed to different meanings of statement. Here some of the issues are discussed as follows,

A. Intermediate Form

Intermediate forms with variable degrees of complexity are appropriate for different mining purposes. For a domain-specific knowledge discovery task, it is essential to perform linguistics analysis to derive a sufficiently rich representation to capture the relationship between the objects or concepts defined in the documents [5]. However, this analysis method is computationally expensive and often operates in the order of a few words per second. It remains a challenge to envision how analysis can be made more efficient and scalable for very large amount text.

B. Multilingual Text Refining

Although data mining is basically language independent, text mining comprises a significant language component. It is important to develop text refining algorithms that process multilingual text documents and it produce language-independent intermediate forms [9]. Even though most text mining tools emphasis on processing English documents, mining from documents in other languages permits access to previously unused information and offers a new host of chances [18].

C. Domain Knowledge Integration

Domain knowledge, not provided for by any current text mining tools, can play an important role in text mining. Specifically, domain knowledge can be used as early as within the text processing stage. It is interesting to explore how one can take advantage of domain information to enhance parsing efficiency and derive an extra compact intermediate form [18]. Domain knowledge can also play a part in knowledge distillation. In a classification or predictive modeling task, domain knowledge helps to enhance learning/mining efficiency and quality of mined knowledge.

VI. CONCLUSION

Data Mining is the important as well as active research area helps to extract helpful patterns from the data. These patterns generated facilitate decision making in industries. Text mining is also crucial field that deals with unstructured or semi structured data. In this paper we have delineated the various text mining techniques such as Information Extraction, Information retrieval, Natural Language processing, Categorization and Clustering. And also we have defined text mining processing flow, applications of text mining and issues in text mining. Mining text in different languages may be a major problem, since text mining tools and techniques ought to be able to work with several languages and multilingual languages. Integrating a domain knowledge base with text mining engine would increase its efficiency, especially within the information retrieval and information extraction phase.

REFERENCES

[1].Varsha C. Pande , Dr. A.S. Khandelwal ,A Survey Of Different Text Mining Techniques, IBMRD's Journal of Management and Research, Online ISSN: 2348-5922, Volume-3, Issue-1, March 2014

[2].Gobinda G. Chowdhury ,Natural Language Processing

[3]. Vishal Gupta, Gurpreet S. Lehal,A Survey of Text Mining Techniques and Applications, Journal of Emerging Technologies in Web Intelligence, Vol-1,No-1, August 2009

[4]. DivyaNasa, Text Mining Techniques- A Survey,International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012 ISSN: 2277 128X

[5]. Falguni N. Patel, Neha R. Soni,Text mining: A Brief survey, International Journal of Advanced Computer Research, ISSNprint: 2249-7277, ISSN online: 2277-7970 Volume-2 Number-4 Issue-6 December-2012

[6]. Mrs. SayantaniGhosh,, Mr. Sudipta Roy, Prof. Samir K. Bandyopadhyay, A tutorial review on Text Mining Algorithms, International Journal of Advanced Research in Computer and Communication Engineering ,Vol. 1, Issue 4, June 2012 ISSN : 2278 – 1021

[7]. Andreas Hotho, Andreas Nurnberger, Gerhard Paas, A Brief Survey of Text Mining, May 13,2005

[8]. ShaidahJusoh , Hejab M. Alfawareh, Techniques,Applications and Challenging Issues in Text Mining, International Journal of Computer Science Issues, Vol. 9, Issue 6, No 2, November -2012 ISSN (Online): 1694-0814

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 4, April 2016

[9]. R. Sagayam, S.Srinivasan, S. Roshni ,A Survey of Text Mining: Retrieval, Extraction and Indexing Techniques, International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5, Issn 2250-3005(online), September| 2012
[10]. Wikipedia
[11]. Anna Stavrianou, PeriklisAndritsos, Nicolas Nicoloyannis, Overview and Semantic Issuses of Text Mining, Special Interest Group Management of Data (SIGMOD) Record, September-2007,Vol.36, No.3
[12]. Ronan Collobert, Jason Weston, A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning
[13]. Mahesh T R, Suresh M B, M Vinayababu, Text Mining: Advancements, Challenges and Future directions, International Journal of Reviews in Computing, ISSN: 2076-3328 E-ISSN: 2076-3336
[14].Anshika Singh , Dr. UdayanGhosh , Text Mining: A Burgeoning technology for knowledge extraction, International Journal of Scientific Research Engineering & Technology (IJSRET) Volume.1 Issue.12 pp 022-026 ,March 2013,ISN 278 - 082
[15]. RanveerKaur, ShrutiAggarwal, Techniques for Mining Text Documents, International Journal of Computer Applications (0975 – 8887), Volume 66, No.18, March 2013
[16]. K.L.Sumathy, M.Chidambaram, Texr Mining: Concepts, Applications, Tools and Issues- An Overview, International Journal of Computer Applications (0975- 8887), Vol-80, No.4, October-2013
[17].Mr. Rahul Patel, Mr. Gaurav Sharma, A survey on text mining techniques, International Journal Of Engineering And Computer Science, ISSN:2319-7242, Volume 3 ,Issue 5 ,May 2014
[18]. Ian H. Witten, Text Mining
[19].PatilMonali S, KankalSandip S, A Concise Survey on Text Data Mining, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 9, September 2014, ISSN (Online) : 2278-1021,ISSN (Print) : 2319-5940
[20]. K.Nalini, Dr.L.JabaSheela, Survey on Text Classification, International Journal of Innovative Research in Advanced Engineering (IJIRAE),ISSN: 2349-2163, Volume. 1,Issue 6, July2014
[21]. K.Thilagavathi, V.Shanmugapriya, A Survey on Text Mining Techniques, International Journal of Research in Computer Applications and Robotics (IJRCAR), ISSN: 2320-7345, Volume. 2, Issue 10, October 2014
[22]. Lokesh Kumar, ParulKalra Bhatia, Text Mining: Concepts, Process and Applications, Journal of Global Research in Computer Science (JGRCS), Volume 4, No. 3, March 2013
[23]. VandanaKorde, C NamrataMahender, Text Classification and Classifiers: A Survey,
International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012
[24].Michael Steinbach George KarypisVipin Kumar, A Comparison of Document Clustering Techniques, Department of Computer Science and Engineering, University of Minnesota.
[25]. PavelBerkhin, Survey of Clustering Data Mining Techniques, Accrue Software, Inc.
[26]. Atika Mustafa, Ali Akbar, Ahmer Sultan, Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization, International Journal of Multimedia and Ubiquitous Engineering
[27]. Umajancy.S,Dr. Antony S Elvadoss Thanamani, An analysis on text mining- text Retrieval and text extraction, International Journal of Advanced Research in Computer and Communication Engineering.

BIOGRAPHY

Dr.S.Vijayarani has completed MCA, M.Phil and Ph.D in Computer Science. She is working as Assistant Professor in the School of Computer Science and Engineering, Bharathiar University, Coimbatore. Her fields of research interest are data mining, privacy and security issues in data mining and data streams. She has published papers in the international journals and presented research papers in international and national conferences.

Ms. R. Janani has completed MCA and M.Phil in Computer Science. She is currently pursuing her Ph.D in Computer Science in the School of Computer Science and Engineering, Bharathiar University, Coimbatore. Her fields of interest are Data Mining, Text Mining and Natural Language Processing.