

# A Conceptual Based Approach in Text Mining: Techniques and Applications

G. L. Anand Babu, B. Srinivasu

**Abstract** - In increasing the development and application of digital data in various fields, the discovery of knowledge and the extraction of texts show great consideration for the most useful information and knowledge. The main concern is the application of appropriate schemes and activities to analyze text documents from a large volume of data. For decision making and future expectations, we use different methods and tools to undermine the text and determine the appropriate information. To improve speed and reduce the time and effort required to extract valuable information, correct and correct methods for extracting text must be applied. The article presents precisely on the evaluation of text mining techniques and their applications in different fields of life.

**Keywords:** Text Mining, Classification, Clustering, Summarization, Information Extraction, Information Retrieval

## I. INTRODUCTION

Text mining is a new area that tries to extract data well expressed by the text in a normal language. It tends to differentiate itself as the way to display content to recover data that is profitable for a specific reason. Text extraction normally has an impact on the text whose task is the matching of certain data or feelings and the motivating forces to instinctively request my data from that text are fascinating, regardless of whether the success is only partial.

Text mining is a practice for extracting substantive and stimulating models to find data from textual databases. Text mining is a multidisciplinary field based on information retrieval, data mining, machine learning and computational semantics. You can link different text mining techniques such as summary, classification, clustering, etc. to extract Text mining agrees with the natural language text that is kept in a semi-structured and unstructured format. Text-based techniques are constantly working in industry, academia, web applications, the Internet and other fields. Use of text mining in application areas, such as search tools, customer relationship management system, clean e-mails, product suggestion queries, fraud detection and social network analysis for the extraction of visualizations, extraction of features, emotions, predictive and trend analysis. Currently, most data in the business world, industry, government and various organizations are stored in a text frame in the database and this text database contains semi-structured data. In this context, the text extraction strategy begins with a grouping of files from a few resources. The text extraction mechanism would improve a particular file and process it by inspecting unusual formats and assemblies.

Revised Manuscript Received on May 07, 2019

G.L.Anand Babu, Department of Information Technology, Anurag Group of Institutions, Hyderabad, India.

Dr.B.Srinivasu, Department of Computer Science and Engineering, Stanley College of Engineering, Hyderabad, India.

Text search is a test to download more valuable information from the text. You can use many text search techniques that are subject to the organization's goal. Frame activities could be used. The resulting information can be placed in an information management system, with a large amount of knowledge for the consumer of that system. The extraction of valuable information from a number of various documents is an exhausting and irritating task.

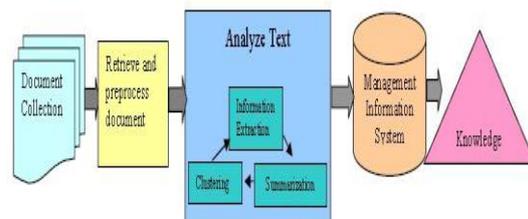


Fig.1: Text Mining Process.

## II. TECHNIQUES USED IN TEXT MINING

Text mining is an essential field that employs collective methods and approaches from many areas, such as

- ❖ Information Retrieval,
- ❖ Extraction of information,
- ❖ Categorization of Text,
- ❖ Summarization of Text And
- ❖ Clustering of Text.

### 2.1 Information Retrieval

Information retrieval is a relatively old-style research area. It has expanded the thought enhanced with the rise of the World Wide Web and the need for refined search engines. The most perceived IR (Information Retrieval) frameworks are the search tools, such as Google, which perceives those reports on the Web that are important for a given word arrangement. The pre-processing effort for web crawlers is the procedure for extracting information to be produced, organizing in a confusion of data. Google crawls the web to get statistics, understand them, and the provisions in a complete structure, so it typically recovers quickly when customers run search queries. It is the job of getting applicable data from a collection of various resources. The process to obtain, organize and examine the possible documents that can meet this data requires the recovery process. This framework is used by many universities, public libraries, governments and organizations to provide access to articles, books, journals and different documents.

# A Conceptual Based Approach in Text Mining: Techniques and Applications

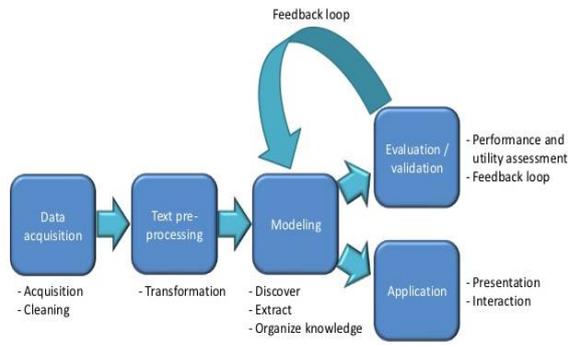


Fig.2: Information Retrieval

## 2.2 Extraction of Information

Information extraction (IE) is the task of immediately obtaining complete systematized information of unstructured or semi-structured natural language. Recognizes the extraction of units, such as names of individuals, association, area and connection between articles, foreground events and textual relationship. The precious information that is separated does not have an adequate relationship with the text, such as a person's name, association, position and gender. These are stored in the database as drawings and therefore are possible for further use. In the overwhelming majority of conditions, this activity alters the processing of text in the human language using natural language methods. The collected data is trained and automatically archived in a database. IE redraws a quantity of textual data in a more systematic database. The significant advantage of data extraction frameworks is query accuracy and openness of output. They can be professionally edited and visually displayed on the screen. It is useful for a variety of applications, especially for the continued proliferation of Internet and Web documents.

## 2.3 Summarization of Text

In recent years there has been an explosion in the dissemination of textual data from a variety of sources. This part of the text is a vital source of information and knowledge that you want to successfully summarize to be invaluable. The text summary is the task of shortening a text document into a shortened version keeping all the information and meaningful content of the original document. It is the process of creating a brief summary without problems, while maintaining key facts and full meaning. Summary systems are able to produce the exact two text requests and general schemes created by the machine that are based on the customer's prerequisites. The text summary is specified to express information for use in trivial mobile devices, such as PDAs, which require significant simplicity reduction.

## 2.4 Categorization of Text

Text categorization is an essential feature of text mining. It is a supervised process and uses predefined documents based on their content. The categorization provides to find exactly which domain category in use, a defined text file is retransmitted. For the implementation of text categorization, an extended tokenization is required. Tokenization refers to the extraction of functional terms in the document. Using categorization tools, systems can

assign unpublished documents to the most accurate category available, based on the classification or specific topic. [1]. It is a set of text documents, the method to find the topic or the precise topics for each document. Nowadays, the automatic categorization of texts is applied in a variety of contexts, from automatic or semi-automatic indexing of texts to the delivery of personalized spots, spam filtering and categorization of Web pages in hierarchical catalogs, automatic generation of metadata and text type detection, monitoring topic and many others [2].

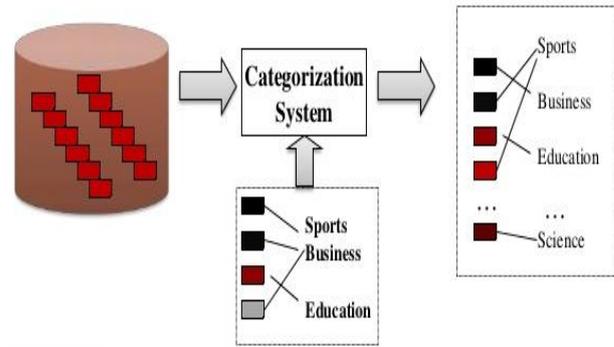


Fig.3 Categorization of Text

## 2.5 Clustering of Text

Grouping is a motivating and essential topic in text mining. Clustering is an unsupervised method on which objects are classified into sets called clusters. Its goal is to uncover basic information structures and insert them into the major subgroups for further study and exams. Clustering plays a key role in several application areas such as biology, image segmentation, data extraction, document retrieval, pattern classification, pattern recognition, security, business intelligence and research. on the Web. The cluster exam is used as a separate tool for text mining to achieve data distribution or as a pre-processing step for other text extraction algorithms that work in the identified clusters. Clustering is a data partition in collections of related objects.

Each set, called a group, is made up of elements related to each other and not related to the objects of other sets. Indicating data for smaller groups of numbers certainly omits certain acceptable details, but gets a simplification. Denotes different data objects for limited groups and then models the data according to their groups. Data modelling positions grouping into a historical vision embedded in mathematics, statistics, and arithmetic analysis.

<<<<

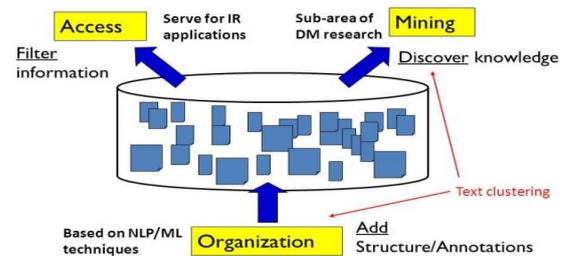


Fig.4 Clustering of Text



### III. COMPARISON OF TEXT MINING TECHNIQUES:

Text mining uses different techniques that play an important role. The techniques differ from each other.

Text Mining Technique	Characteristics	Algorithms	Model	Tools
Information Retrieval	Recover valuable information from unstructured text	Retrieval, Indexing, Filtering	-	Intelligent Miner, Text Analyst
Extraction of Information	Extract information from the structured database	Ripper, Apriori	-	Text Finder, Clear Forest Text
Summarization of Text	Reduce your length by keeping your main points and their general meaning as it is	Key phase Extraction, Text Rank, Page Rank, Grasshopper	Naïve Bayes Model	Tropic Tracking Tool, Sentence Ext Tool
Categorization of Text	Document-based categorization	K-NN, Support Vector Machine, Decision Tree Induction	Support Vector Machine, Probabilistic or generative model based	Intelligent Miner
Clustering of Text	Collection of group documents, grouping, classification and analysis of text documents	K-Mean & K-Medoids, DBSCAN	Statistical Model, Support Vector Machine	Carrot, Rapid Miner

### IV. APPLICATION OF TEXT MINING

Text-mining techniques greatly influence industry, from academia and healthcare to businesses and social networking platforms. Currently, some text mining applications are used all over the world :

#### 4.1 Educational and Research Field

In the academic field, some tools and methods of text mining are used to evaluate the specific educational models of the region, the state of alert of the academics in certain fields and the professional proportion. The formation of text mining in the field of research allows you to find and classify research documents and related material from different fields in one place. Determining models and styles in journals and procedures of large articles is an important task in the field of research [3]. The text mining tool is useful for determining trends in various topics that occur in procedures and to show how they change over time. Furthermore, it is used as a trace of arguments. Therefore, initiatives such as the Nature Proposal for a Open Text Mining (MTOO) interface and the National Documentation Definition (DTD) of the National Institutes of Health that provide semantic signals to the machines to respond to precise and limited queries within the text without remove the Obstacles editor for open access.

#### 4.2 Social Media Analysis

Several text mining software packages are fully considered to analyze the performance of social media platforms. The texts produced online by news, e-mails, blogs help to track and interpret the use of text mining software packages. The amount of publications, "likes" and followers is analyzed efficiently through the use of text mining on social networks and recognizing the reactions of people interacting with online content. Analysis also allows us to recognize what is in fashion and what is not for the target audience.

#### 4.3 Customer care service

Text-mining techniques, mainly NLP, are becoming increasingly important in customer service. Companies are capitalizing on text analysis software to enrich their overall customer experience by accessing

text data from a variety of sources, such as customer feedback, customer surveys and calls, etc. Customers quickly and efficiently.

#### 4.4 Resume Filtering

Text mining plays a key role in filtering the curriculum. Renowned companies receive thousands of CVs from job seekers every day. Resource mining information with extraordinary precision and recovery is not an easy task [4]. Although it establishes a limited domain, the curricula are written in a multitude of drawings (for example, organized tables or simple texts), in different languages (for example, Hindi and English) and in different file types (for example, plain text, PDF, Word, etc.). Furthermore, writing styles are very diverse. In the manual curriculum test, a recruiter searches for errors, qualifications, key words, professional curriculum, job titles, frequency of job changes and other personal information [5]. The automatic extraction of this information can be the first step to filter the curricula. Therefore, automating the practice of curriculum selection is a key task.

### V. CONCLUSION

Text mining deals with texts in natural language stored in semi-structured or unstructured formats. The document dealt with various text mining techniques, their applications in various fields and the comparison of different text mining techniques, which can be further improved.

### REFERENCES

1. Juan Jose Garcia Adeva and Rafael Calvo, "Mining Text with Pimiento", University of Sydney.
2. Rashmi Agrawal, Mridula Batra, "A Detailed Study on Text Mining Techniques", IJSCE, ISSN: 2231-2307, Vol. 2, Issue-6, January 2013.
3. Vallikannu Ramanathan, T. Meyyappan "Survey of Text Mining", International Conference on Technology and Business and Management, March 2013, pp. 508-514.
4. Daniel Waegel. —The Development of Text-Mining Tools and Algorithmsl. Ursinus College, 2006.
5. Text Mining Summit Conference Brochure, <http://www.textminingnews.com>, 2005