



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Text Mining

Priyanka Reddy¹, Deekshitha²

1. Presidency University, Email: deekshithal2002@gmail.com

2. Presidency University, Email: redhypriyanka281@gmail.com

DOI: <https://doi.org/10.55248/genppi.2022.3.6.22>

ABSTRACT

The unstructured texts which contain massive amount of information cannot simply be Used for further processing by computers. Therefore, specific processing methods and Algorithms are required in order to extract useful patterns. The process of extracting Interesting information and knowledge from unstructured text completed by using Text Mining. In this paper, we have discussed text mining, as a recent and interesting field with The detail of steps involved in the overall process. We have also discussed different Technologies that teach computers with natural language so that they may analyze, Understand, and even generate text. In addition, we briefly discuss a number of successful Applications of text mining which are used currently and in future.

INTRODUCTION



Text mining is a technique which extracts information from unstructured data and find Pattern which is novel and unknown earlier. It is also known as knowledge discovery from Text (KDT), deals with the machine supported analysis of text .Text documents are in semi Structured or unstructured format datasets such as emails, full-text documents, HTML files Etc. The problem of Knowledge Discovery from Text (KDT) is to extract explicit and implicit Concepts and semantic relations between concepts. Its aim is to get insights into large Quantities of text data.

Text mining process

A. Text Preprocessing:-

The text pre-processing step further divided into number of sub steps as Follows:

- 1) Tokenization: Text document has a collection of sentences. This step divide whole statement into words by removing spaces, commas etc.
- 2) Stop word Removal: This step involves removing of HTML,XML tags from Web pages. Then process of removal of Stop words like “a”, “of” etc. is performed. Finally word stemming is applied.
- 3)Stemming: These techniques are used to find out the root/stem of a word. Stemming converts words to their stems. E.g. Flying, Flew word to Fly. The algorithm proposed by Port, known as a Port’s stemming algorithm is widely used for the same .

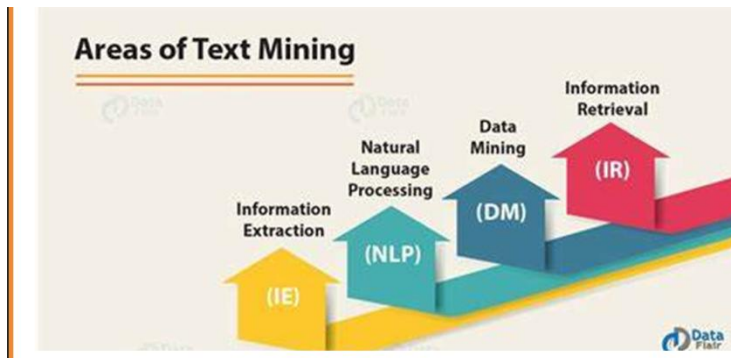
B. Text Transformation / Feature Generation :-

Text transformation means to convert text document into bag of words or Vector space document model notation, which can be used for further effective analysis task.

C. Feature Selection/Attribute Selection :-

This phase mainly performs removing features that are considered irrelevant for mining purpose. This procedure give advantage of smaller dataset size, less computations and minimum search space required.

D. Text mining methods There are different text mining methods as in Data mining had been proposed such as: Clustering, Classification, Information retrieval, Topic discovery



Basic Technologies

Information Extraction :-

Information extraction technique identifies key phrases and relationship within a text [5]. For that it uses pattern matching method. Pattern matching means matching predefined sequences of text with user text. This technique is very useful in analysing large text dataset. The extracted information by IE cannot be represented directly into a structured form. Hence post processing is required

Applications of textmining



I. Risk Management :-

One of the primary causes of failure in the business sector is the lack of proper or insufficient risk analysis. Adopting and integrating risk management software powered by text mining technologies such as SAS Text Miner can help businesses to stay updated with all the current trends in the business market and boost their abilities to mitigate potential risks. Since text mining tools and technologies can gather relevant information from across thousands of text data sources and create links between the extracted insights, it allows companies to access the right information at the right moment, thereby enhancing the entire risk management process.

2. Customer Care Service:-

Text mining techniques, particularly NLP, are finding increasing importance in the field of customer care. Companies are investing in text analytics software to enhance their overall customer experience by accessing the textual data from varied sources such as surveys, customer feedback, and customer calls, etc. Text analysis aims to reduce the response time of the company and help address the grievances of the customers speedily and efficiently.

3. Fraud Detection :-

Text analytics backed by text mining techniques provides a tremendous opportunity for domains that gather a majority of data in the text format. Insurance and finance companies are harnessing this opportunity. By combining the outcomes of text analyses with relevant structured data these companies are now able to process claims swiftly as well as to detect and prevent frauds.

CONCLUSION

Text mining, also known as Text Data Mining or Knowledge-Discovery in Text (KDT), refers generally to the process of extracting interesting and non-trivial information and knowledge from unstructured text. We also discuss the General Process of Text encoding and mining. Current text mining products and applications are designed for trained knowledge specialists.

References

-
- [1] Haralampos Karanikas and Babis Theodoulidis Manchester, "Knowledge Discovery in Text and Text Mining Software", Centre for Research in Information Management (2001), UK.
 - [2] Munyaradzi Chiwara, Mahmoud Al-Ayyoub, Mohammad Sajjad Hussain, Rajan Gupta, "CSE-634 Data mining: Text Mining", ppt presentation.
 - [3] M. Porter. An algorithm for suffix Stripping and stemming. Program, pages 130–137, 1980.
 - [4] Vishal Gupta and Guruprit Lehal, "A Survey of Text Mining Techniques and Applications", Journal Of Emerging Technologies In Web Intelligence, Vol. 1, No. 1, August 2009.