

Anekant Education Society's

**Tuljaram Chaturchand College
of
Arts , Science and Commerce**
(Autonomous)


Baramati

Department Of Computer Science

Ms. Kalyani Waman Londhe.
M.Sc. (Computer Science) NET



INTRODUCTION TO TEXT MINING

- 
- Sometimes it is also called as Text data mining or Text analytics
 - There are three different perspectives of text mining
 1. Text Mining as Knowledge Discovery in Databases Process
 2. Text Mining as Information Extraction
 3. Text Mining as Text Data Mining

Text Mining Definition


- Text mining is the process of exploring, analyzing and finding some pattern, trend or association from large amounts of text data which is available through various sources.

Sources for Text Mining

- Research papers
- News articles
- Blogs
- Books
- Digital libraries
- Email messages
- Web pages

Two basic information retrieval methods

- Document Selection Method :
 - In this method query is modeled as Boolean retrieval model in which document is represented by set of keywords.
 - User provides Boolean expression of keywords
 - Retrieval system takes such as Boolean query and returns documents that satisfy Boolean expression
 - For example keyword is “Car Showrooms in baramati”
 - User will get all relevant results regarding car showrooms in baramati.

- 
- Document Ranking Problem :
 - This method is specially used for organizing or ranking the document in specific order using query
 - More appropriate than document selection method
 - Different ranking methods are used based on probability, statistics, mathematics and logic.
 - Score for each document is decided based on the matching keywords.

Text Mining Approaches

- The keyword based approach
 - Here input is set of keywords or terms in the documents.
 - Discover relationships at shallow level
 - Less significant results may generate
- The tagging approach
 - Here input is set of tags
 - Completely rely on tagging
 - It is costly as compared to other 2
 - Not feasible for large datasets
- The information extraction approach
 - Here input is semantic information
 - For example events, facts or entities
 - Advanced approach
 - Discover relationship at deep level

Text Mining Applications

- **Security applications**

- Used for monitoring and analysis of online plain text sources such as Internet news, blogs, etc. for national security purposes.
- It is also involved in the study of text encryption and decryption

- **Marketing applications**

- Used in analytical customer relationship management
- Also applied in stock returns prediction



- **Biomedical applications**

- Text mining can be used in clinical studies and precision medicine
- GoPubMed is a knowledge-based search engine for biomedical texts.
- PubGene, a search engine that combines biomedical text mining with network visualization.

- **Sentiment analysis**

- Involve analysis of movie reviews for estimating how favorable a review is for a movie
- Used for affective computing in multiple corpora such as students evaluations, children stories and news stories.



- **Online media applications**

- Tribune Media Company is using text mining to clarify information and to provide readers with greater search experiences, which in turn increases site's "stickiness" and revenue.

- **Software applications**

- Used to create software for tracking and monitoring terrorist activities.
- Weka software is one of the most popular options in the scientific world
- For Python programmers, there is an excellent toolkit called NLTK for more general purposes.



Thank You