



ACTIVIDAD #3

Tipo actividad: Reading comprehension: "What is Text Mining: Techniques and Applications"

Socialization key vocabulary reading #3: "What is Text Mining: Techniques and Applications"

- 1. Text Mining:** Text Mining is the process of deriving high-quality information from unstructured text data, aiming to extract meaningful patterns and insights from textual sources.
- 2. Information Extraction:** Information Extraction is a text mining technique that focuses on identifying and extracting entities, attributes, and relationships from semi-structured or unstructured texts, storing the information in a database for future access.
- 3. Information Retrieval (IR):** Information Retrieval is a text mining technique involving the extraction of relevant patterns based on specific words or phrases, often using algorithms to track and monitor user behaviors to discover relevant data.



4. **Clustering:** Clustering is a crucial text mining technique that seeks to identify intrinsic structures in textual information, organizing them into relevant subgroups or clusters for further analysis, without prior information on the data.

5. **Sentiment Analysis:** Sentiment Analysis, also known as opinion mining, is an AI technique used to determine whether the sentiment expressed in data is positive, negative, or neutral. It involves analyzing text data to understand emotions and opinions.

10) Reading comprehension #3: "What is Text Mining: Techniques and Applications"

What is Text Mining: Techniques and Applications

Text Mining is one of the most critical ways of analyzing and processing unstructured data which forms nearly 80% of the world's data. Today a majority of organizations and institutions gather and store massive amounts of data in data warehouses, and cloud platforms and this data continues to grow exponentially by the minute as new data comes pouring in from multiple sources.

As a result, it becomes a challenge for companies and organizations to store, process, and analyze vast amounts of textual data with traditional tools. Upskilling yourself with data science programs will help you overcome the challenges. Let's talk more about text mining.



What is Text Mining?

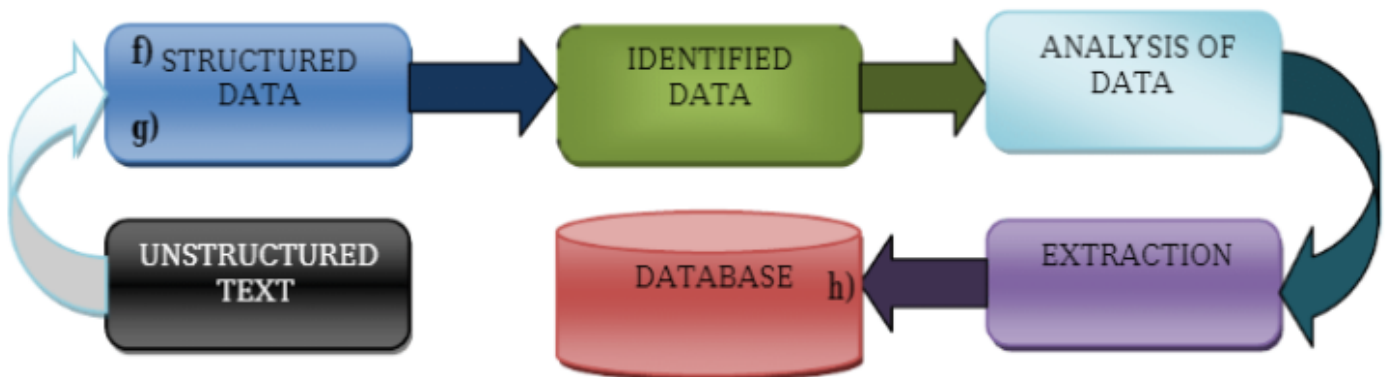
According to Wikipedia, “Text mining, also referred to as text data mining, roughly equivalent to text analytics, is the process of deriving high-quality information from text.” The definition strikes at the primary chord of text mining – to delve into unstructured data to extract meaningful patterns and insights required for exploring textual data sources.

Text mining incorporates and integrates the tools of information retrieval, data mining, machine learning, statistics, and computational linguistics, and hence, it is nothing short of a multidisciplinary field. Text mining deals with natural language texts either stored in semi-structured or unstructured formats. Now that we know what is Text Mining let us understand the steps involved in this.

The five fundamental steps involved in text mining are:

- Gathering unstructured data from multiple data sources like plain text, web pages, pdf files, emails, and blogs, to name a few.
- Detects and removes anomalies from data by conducting pre-processing and cleansing operations. Data cleansing allows you to extract and retain the valuable information hidden within the data and to help identify the roots of specific words.
- For this, you get a number of text mining tools and text mining applications.

- Convert all the relevant information extracted from unstructured data into structured formats.
- Analyze the patterns within the data via the Management Information System (MIS).
- Store all the valuable information into a secure database to drive trend analysis and enhance the decision-making process of the organization.



Text Mining Techniques

Text mining techniques can be understood at the processes that go into mining the text and discovering insights from it. These text mining techniques generally employ different text mining tools and applications for their execution. Now, let us now look at the various text mining techniques:

Let us now look at the most famous techniques used in text mining techniques:



1. Information Extraction

This is the most famous text mining technique. Information exchange refers to the process of extracting meaningful information from vast chunks of textual data. This text mining technique focuses on identifying the extraction of entities, attributes, and their relationships from semi-structured or unstructured texts. Whatever information is extracted is then stored in a database for future access and retrieval. The efficacy and relevancy of the outcomes are checked and evaluated using precision and recall processes. The technique which is useful for analyzing the textual data is Information Extraction.

2. Information Retrieval

Information Retrieval (IR) refers to the process of extracting relevant and associated patterns based on a specific set of words or phrases. In this text mining technique, IR systems make use of different algorithms to track and monitor user behaviors and discover relevant data accordingly. Google and Yahoo search engines are the two most renowned IR systems.



3. Categorization

This is one of those text mining techniques that is a form of “supervised” learning wherein normal language texts are assigned to a predefined set of topics depending upon their content. Thus, categorization or rather Natural Language Processing (NLP) is a process of gathering text documents and processing and analyzing them to uncover the right topics or indexes for each document. The co-referencing method is commonly used as a part of NLP to extract relevant synonyms and abbreviations from textual data. Today, NLP has become an automated process used in a host of contexts ranging from personalized commercials delivery to spam filtering and categorizing web pages under hierarchical definitions, and much more. The technique which is useful for analyzing the textual data is Categorization.

4. Clustering

Clustering is one of the most crucial text mining techniques. It seeks to identify intrinsic structures in textual information and organize them into relevant subgroups or ‘clusters’ for further analysis. A significant challenge in the clustering process is to form meaningful clusters from the unlabeled textual data without having any prior information on them. Cluster analysis is a standard text mining tool that assists in data distribution or acts as a pre-processing step for other text mining algorithms running on detected clusters. The most famous technique used in text mining is Clustering.

5. Summarisation

Text summarisation refers to the process of automatically generating a compressed version of a specific text that holds valuable information for the end-user. The aim of this text mining technique is to browse through multiple text sources to craft summaries of texts containing a considerable proportion of information in a concise format, keeping the overall meaning and intent of the original documents essentially the same. Text summarisation integrates and combines the various methods that employ text categorization like decision trees, neural networks, regression models, and swarm intelligence.

Technique	Characteristics	Tools
Retrieval	Retrievals valuable information from unstructured text	Intelligent Miner, Text Analyst
Extraction	Extract information from structured database	Text Finder, Clear Forest Text
Summarization	Reduce length by keeping its main points and overall meaning as it is	Tropic Tracking Tool, Sentence Ext Tool
Categorization	Document based categorization	Intelligent Miner
Cluster	Cluster collection of documents, Clustering, classification and analysis of text document	Carrot, Rapid Miner



Applications Of Text Mining

Text mining techniques and text mining tools are rapidly penetrating the industry, right from academia and healthcare to businesses and social media platforms. This is giving rise to a number of text mining applications. Here are a few text mining applications used across the globe today:

1. 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.



Top Data Science Skills to Learn

	Top Data Science Skills to Learn	
1	Data Analysis Course	Inferential Statistics Courses
2	Hypothesis Testing Programs	Logistic Regression Courses
3	Linear Regression Courses	Linear Algebra for Analysis

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.

4. Business Intelligence

Organizations and business firms have started to leverage text mining techniques as part of their business intelligence. Apart from providing profound insights into customer behavior and trends, text mining techniques also help companies to analyze the strengths and weaknesses of their rivals, thus, giving them a competitive advantage in the market. Text mining tools such as Cogito Intelligence Platform and IBM text analytics provide insights on the performance of marketing strategies, latest customer and market trends, and so on.



5. Social Media Analysis

There are many text mining tools designed exclusively for analyzing the performance of social media platforms. These help to track and interpret the texts generated online from the news, blogs, emails, etc. Furthermore, text mining tools can efficiently analyze the number of posts, likes, and followers of your brand on social media, thereby allowing you to understand the reaction of people who are interacting with your brand and online content. The analysis will enable you to understand 'what's hot and what's not' for your target audience.

Importance of Text Mining in Data Mining

In the article, we have covered the basics of what is Text Mining and what is the most famous techniques in Text Mining, now let's understand the importance of Text Mining in Data Mining.

Data and information have grown at an amazing rate due to the quick increase of computerized or digital information. Text databases, which include enormous collections of documents from diverse sources, are where a significant amount of the information that is now available is kept.



Read our popular Data Science Articles

Data Science Career Path: A Comprehensive Career Guide	Data Science Career Growth: The Future of Work is here	Why is Data Science Important? 8 Ways Data Science Brings Value to the Business
Relevance of Data Science for Managers	The Ultimate Data Science Cheat Sheet Every Data Scientists Should Have	Top 6 Reasons Why You Should Become a Data Scientist
A Day in the Life of Data Scientist: What do they do?	Myth Busted: Data Science doesn't need Coding	Business Intelligence vs Data Science: What are the differences?

The steps to Text Mining –

Assembling unstructured data from many sources that are available in different document formats, such as plain text, web pages, PDF documents, etc.

- To identify and remove discrepancies from the data, pre-processing and data cleansing procedures are carried out. In order to avoid stopping words stemming, the data clearing procedure ensures that the original text is captured.
- To examine and further clean the data collection, processing, and controlling activities performed.
- The data extracted from the information processed in the above mentioned processes are used for a strong and practical decision-making process and trend analysis.



Industries that use Text Mining efficiently –

Financial Services – The financial services industry is incredibly intricate. It involves a significant quantity of communication, paperwork, risk assessment, and compliance. Financial services companies use text analytics to examine client comments, assess claims, consider consumer interactions, and pinpoint compliance issues. Staff members may quickly and easily search internal legal papers for terms related to money or fraud using a text analytics system built on NLP. When compared to completing it manually, this can save a significant amount of time.

Healthcare and Pharma

Specialists in medical affairs assist in the transition of pharmaceutical goods from R&D to commercialization. Text mining is being used by experts in medical affairs to automatically interpret each of these and report changes. Depending on what these alterations indicate for the medication they are creating, the specialists can then adjust their direction. Instead of relying on human labor, text mining can track these changes more accurately and extensively while taking up less time.



Retail

The consumer is always correct in the retail industry. With the surge in online sales during the pandemic, e-Commerce sellers, in particular, have to make sure that the consumer experience is as favorable as possible. Even more so than at physical establishments that people visit, a bad experience makes a client unwilling to return. Text mining is being used by many e-tailers to collect, organize, and analyze consumer input that identifies places of friction while using an e-commerce website or interacting with customer care.