Seminar
On
Text Mining

Submitted To:

MS.SANA BHARTI

Submitted By:

DHRUV GARG

Content

Introduction

- In text mining, the goal is to discover unknown information, something that no one yet knows and so could not have yet written down.

- Text mining is different from what are familiar with in web search.

- In search, the user is typically looking for something that is already known and has been written by someone else.

What is Text mining?

"The objective of Text Mining is to exploit information contained in textual documents in various ways, including …discovery of patterns and trends in data, associations among entities, predictive rules, etc." (Grobelnik et al., 2001)
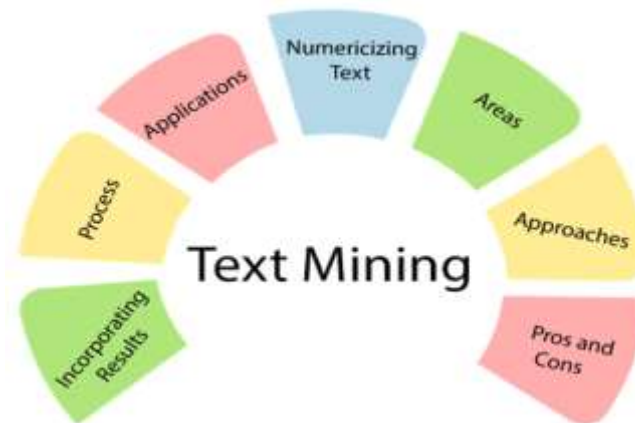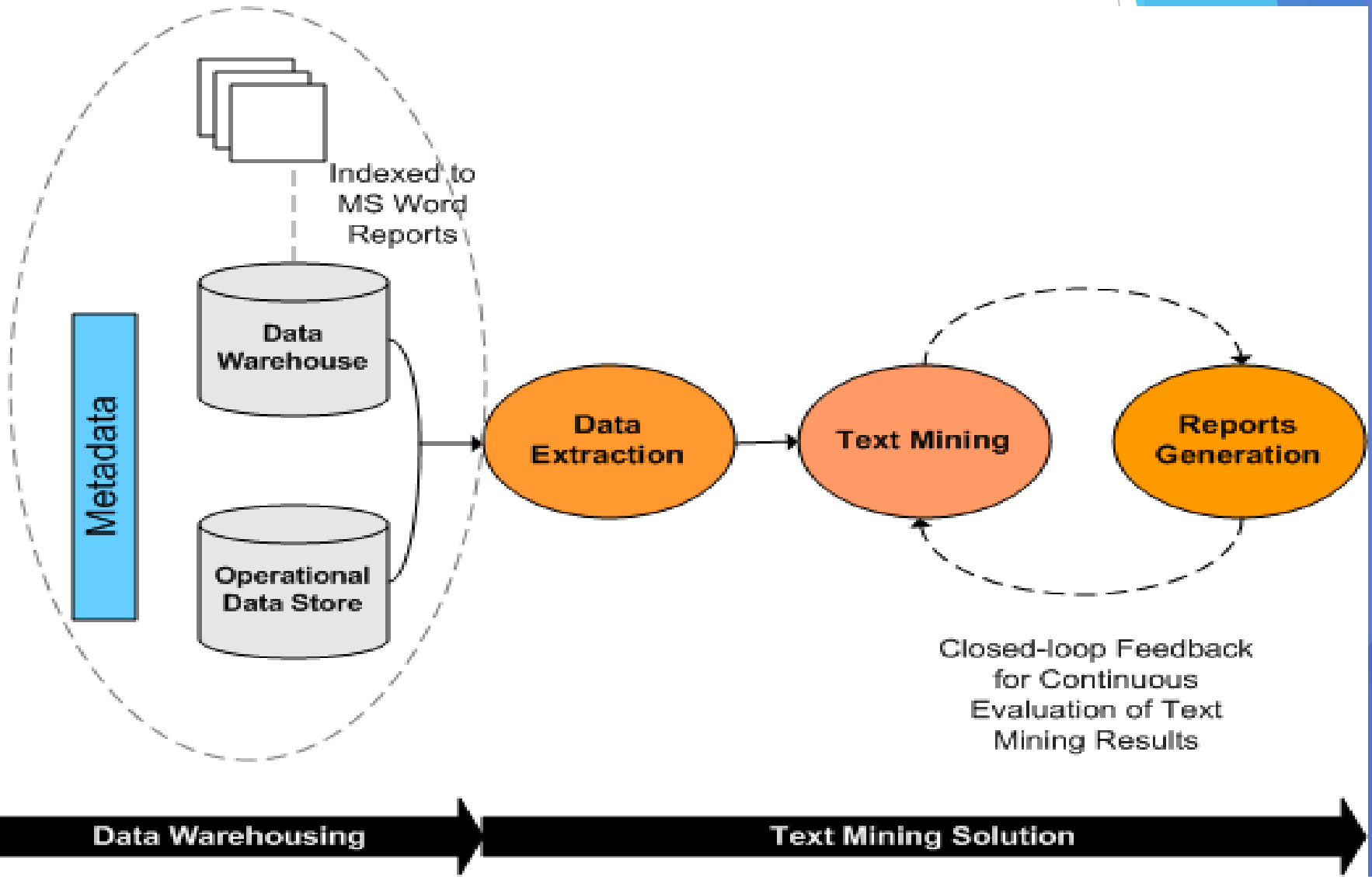
Text Mining Methods

- Data Mining Style: View text as high dimensional data
  - Frequent pattern finding
  - Association analysis
  - Outlier detection

- Information Retrieval Style: Fine granularity topical analysis
  - Topic extraction
  - Exploit term weighting and text similarity measures
  - Question answering
- Natural Language Processing Style: Information Extraction
  - Entity extraction
  - Relation extraction
  - Sentiment analysis

- Machine Learning Style: Unsupervised or semi-supervised learning
  - Generative models
  - Dimension reduction
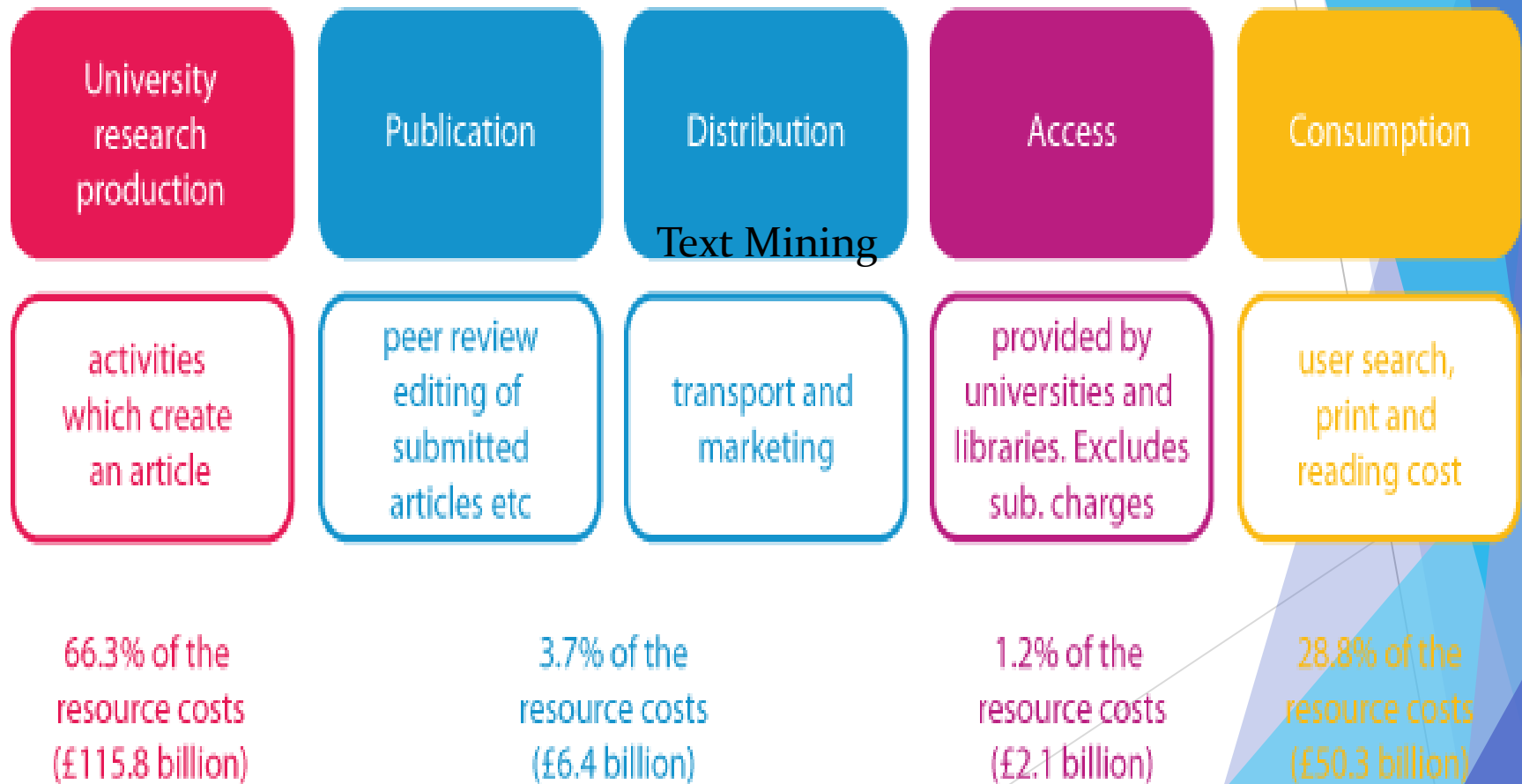 - Classification & prediction

Applications

- ▶ Spam filtering

- ▶ Monitoring public opinions (for example in blogs or review sites)

- ▶ Customer service, email support

- ▶ Automatic labeling of documents in business libraries

- ▶ Fraud detection by investigating notification of claims

Text Mining Process

Text Mining Resource

University research production	Publication	Distribution	Access	Consumption
		Text Mining		
activities which create an article	peer review editing of submitted articles etc	transport and marketing	provided by universities and libraries. Excludes sub. charges	user search, print and reading cost
66.3% of the resource costs (£115.8 billion)	3.7% of the resource costs (£6.4 billion)		1.2% of the resource costs (£2.1 billion)	28.8% of the resource costs (£50.3 billion)

Text Mining Features

- **Data Access**
- **Feature Extraction**
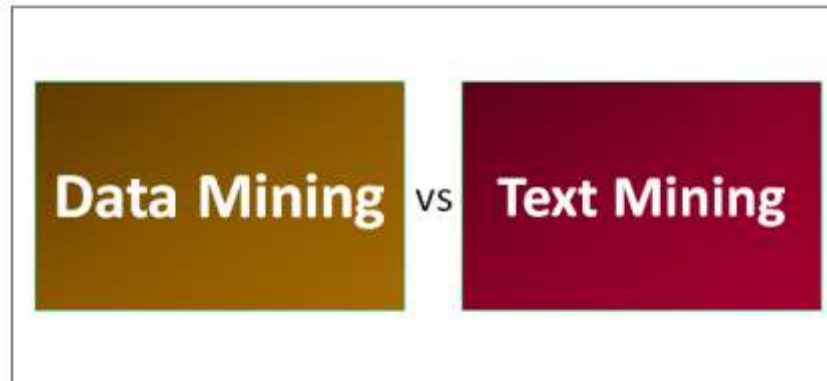- **Text Processing Capabilities**
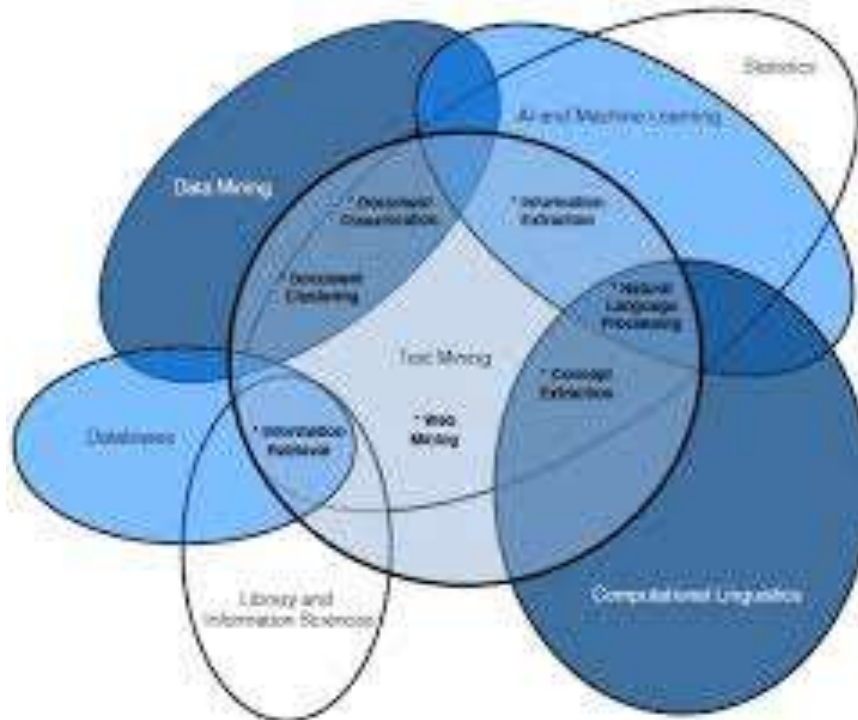
Text mining and Data mining

<u>Data Mining</u>

- ▶ Identify data sets
- ▶ Select features
- ▶ Prepare data
- ▶ Analyze distribution

<u>Text Mining</u>

- ▶ Identify documents
- ▶ Extract features
- ▶ Select features by algorithm
- ▶ Prepare data
- ▶ Analyze distribution

Data Mining vs Text Mining

Value and Benefits of Text Mining

✓ Unlock hidden information
✓ Analyze the behavior of people
✓ Analyze customers and competitors to take better decisions
✓ Spot emerging trends in market
✓ Targeted identification of information
✓ Find topics, opinions and ideas
✓ Predict the feature

1%
4%
12%
15%
68%

- UK public
- UK charity
- Non-UK sources
- UK private
- Other
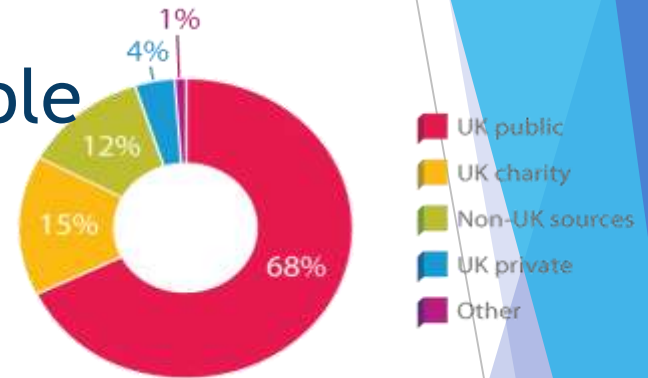
Approaches to Text Mining

- Using well-tested methods and understanding the results of text mining.

- "Black-box" approaches to text mining and extraction of concepts.

- Text mining as document search.

Conclusion

▶ Text mining is a burgeoning technology that is still, because of its newness and intrinsic difficulty, in a fluid state—akin, perhaps, to the state of machine learning in the mid-1980s.

▶ Generally accepted characterizations of what it covers do not yet exist.

Reference

- https://monkeylearn.com/text-mining/

- https://www.rightsdirect.com/3-reasons-consider-text-mining-4/

- https://www.analyticssteps.com/blogs/what-text-mining-process-methods-and-applications

- https://guides.library.uq.edu.au/research-techniques/text-mining-analysis/types-of-text-analysis

- https://www.upgrad.com/blog/what-is-text-mining-techniques-and-applications/

thank you!