

Text Mining, Sentiment Analysis, and Social Analytics

Prepared by Dr. Mohamed Sleem

Subject code : FNT 631
Data Analysis

Lecture 10



Text Mining, Sentiment Analysis, and Social Analytics

Lecture 10

Subject code : FNT 631
Data Analysis

Prepared by
Dr. Mohamed Sleem

www.midocean.km

What is a decision model?

Answer:

- A decision model is an abstraction or representation of a real system, idea, or object used to understand, analyze, or facilitate making a decision. It includes data, uncontrollable variables, and decision variables, and can be represented as pictures, spreadsheets, or mathematical relationships.

Learning Objectives (1 of 2)

- 7.1** Describe text mining and understand the need for text mining
- 7.2** Differentiate among text analytics, text mining and data mining
- 7.3** Understand the different application areas for text mining
- 7.4** Know the process of carrying out a text mining project
- 7.5** Appreciate the different methods to introduce structure to text-based data
- 7.6** Describe sentiment analysis

Learning Objectives (2 of 2)

- 7.7** Develop familiarity with popular applications of sentiment analysis
- 7.8** Learn the common methods for sentiment analysis
- 7.9** Become familiar with speech analytics as it relates to sentiment analysis
- 7.10** Learn three facets of Web analytics—content, structure, and usage mining
- 7.11** Know social analytics including social media and social network analyses

Text Analytics and Text Mining (1 of 2)

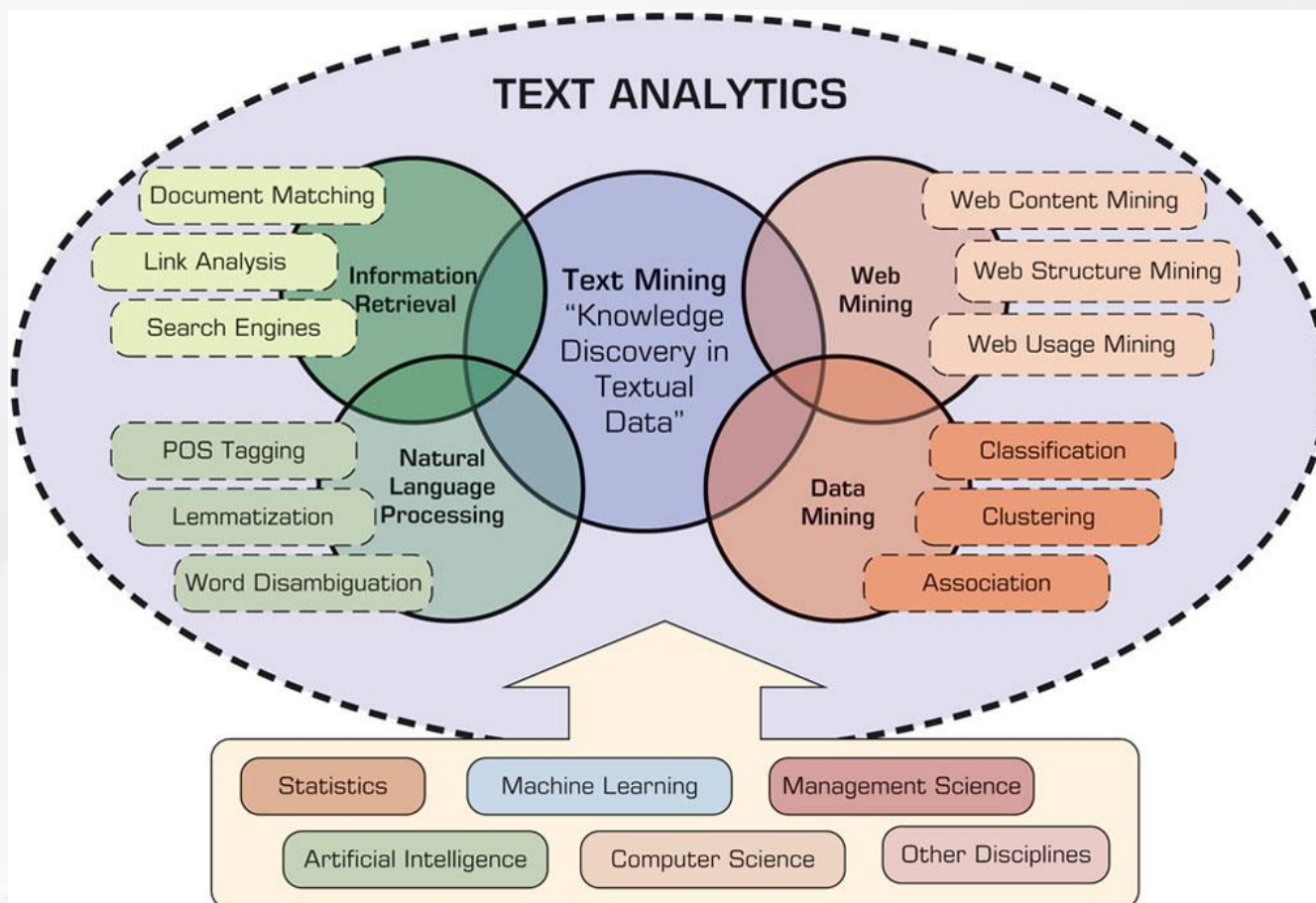
- Text Analytics versus Text Mining
- Text Analytics =
 - Information Retrieval +
 - Information Extraction +
 - Data Mining +
 - Web Mining

or simply

- ***Text Analytics = Information Retrieval + Text Mining***

Text Analytics and Text Mining (2 of 2)

Figure 7.2 Text Analytics, Related Application Areas, and Enabling Disciplines.



Text Mining Concepts (1 of 2)

- 85-90 percent of all corporate data is in some kind of unstructured form (e.g., text)
- Unstructured corporate data is doubling in size every 18 months
- Tapping into these information sources is not an option, but a need to stay competitive
- Answer: text mining
 - A semi-automated process of extracting knowledge from unstructured data sources
 - a.k.a. text data mining or knowledge discovery in textual databases

Data Mining versus Text Mining

- Both seek for novel and useful patterns
- Both are semi-automated processes
- Difference is the nature of the data:
 - Structured versus unstructured data
 - **Structured data:** in databases
 - **Unstructured data:** Word documents, PDF files, text excerpts, XML files, and so on
- **To perform text mining** – first, impose structure to the data, then mine the structured data.

Text Mining Concepts (2 of 2)

- Benefits of text mining are obvious especially in text-rich data environments
 - e.g., law (court orders), academic research (research articles), finance (quarterly reports), medicine (discharge summaries), biology (molecular interactions), technology (patent files), marketing (customer comments), etc.
- Electronic communication records (e.g., Email)
 - Spam filtering
 - Email prioritization and categorization
 - Automatic response generation

Text Mining Application Area

- Information extraction
- Topic tracking
- Summarization
- Categorization
- Clustering
- Concept linking
- Question answering

Text Mining Terminology (1 of 2)

- Unstructured or semistructured data
- Corpus (and corpora)
- Terms
- Concepts
- Stemming
- Stop words (and include words)
- Synonyms (and polysemes)
- Tokenizing

Text Mining Terminology (2 of 2)

- Term dictionary
- Word frequency
- Part-of-speech tagging
- Morphology
- Term-by-document matrix
 - Occurrence matrix
- Singular value decomposition
 - Latent semantic indexing

Natural Language Processing (NLP) (1 of 4)

- Structuring a collection of text
 - Old approach: bag-of-words
 - New approach: natural language processing
- NLP is ...
 - a very important concept in text mining
 - a subfield of artificial intelligence and computational linguistics
 - the studies of "understanding" the natural human language
- Syntax- versus semantics-based text mining

Natural Language Processing (NLP) (2 of 4)

- What is “Understanding” ?
 - Human understands, what about computers?
 - Natural language is vague, context driven
 - True understanding requires extensive knowledge of a topic
 - Can/will computers ever understand natural language the same/accurate way we do?

Natural Language Processing (NLP) (3 of 4)

- Challenges in NLP
 - Part-of-speech tagging
 - Text segmentation
 - Word sense disambiguation
 - Syntax ambiguity
 - Imperfect or irregular input
 - Speech acts
- Dream of AI community
 - to have algorithms that are capable of automatically reading and obtaining knowledge from text

NLP Task Categories

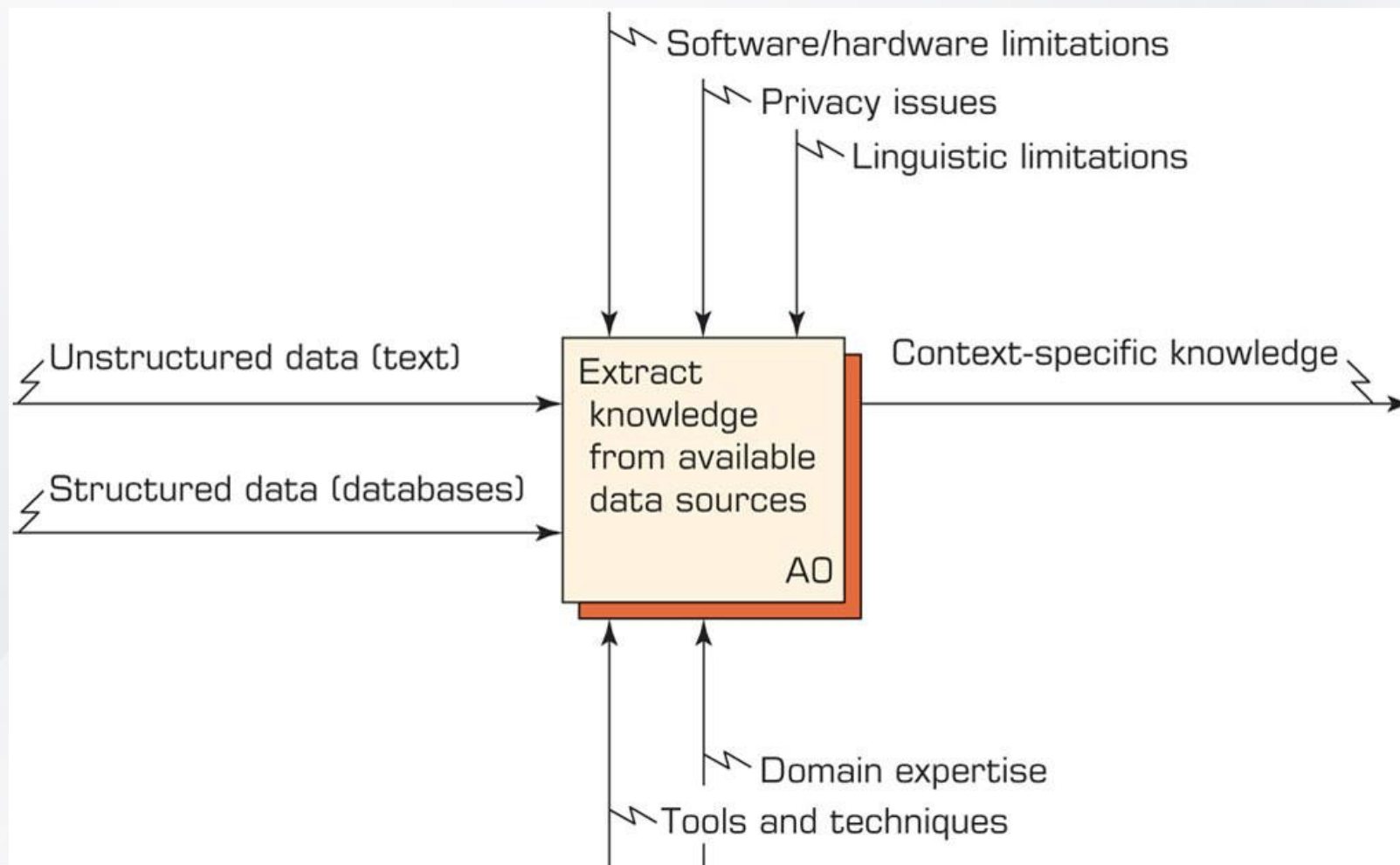
- Question answering
- Automatic summarization
- Natural language generation & understanding
- Machine translation
- Foreign language reading & writing
- Speech recognition
- Text proofing, optical character recognition
- Optical character recognition

Text Mining Applications (1 of 2)

- Marketing applications
 - Enables better CRM
- Security applications
 - ECHELON, OASIS
 - Deception detection (...example coming up)
- Medicine and biology
 - Literature-based gene identification (...)
- Academic applications
 - Research stream analysis

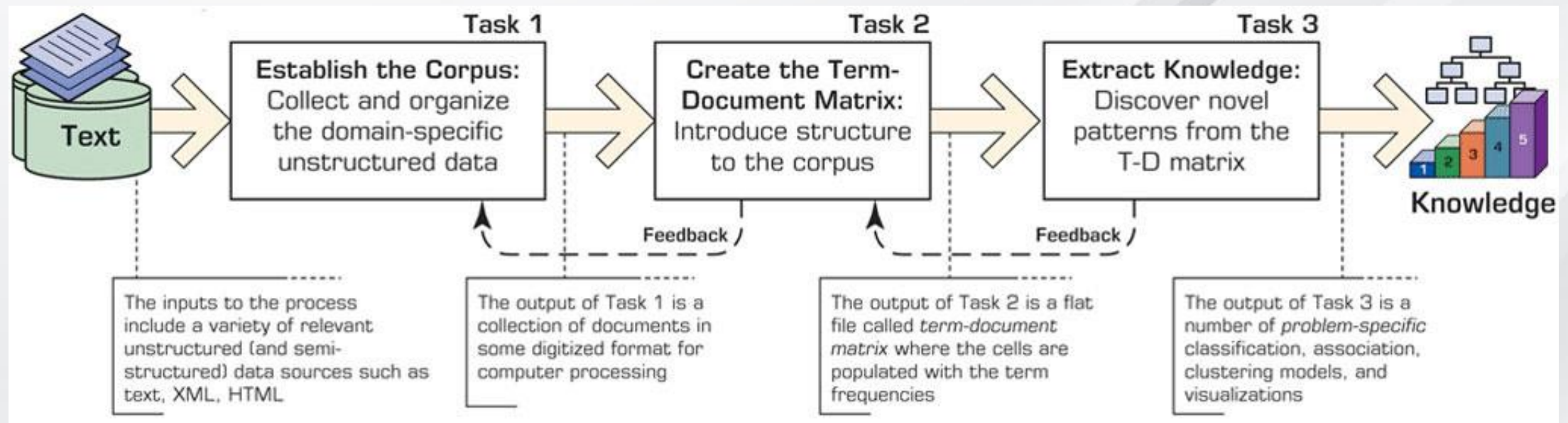
Text Mining Process (1 of 7)

- A Context Diagram for Text Mining Process



Text Mining Process (2 of 7)

Figure 7.6 The Three-Step/Task Text Mining Process.



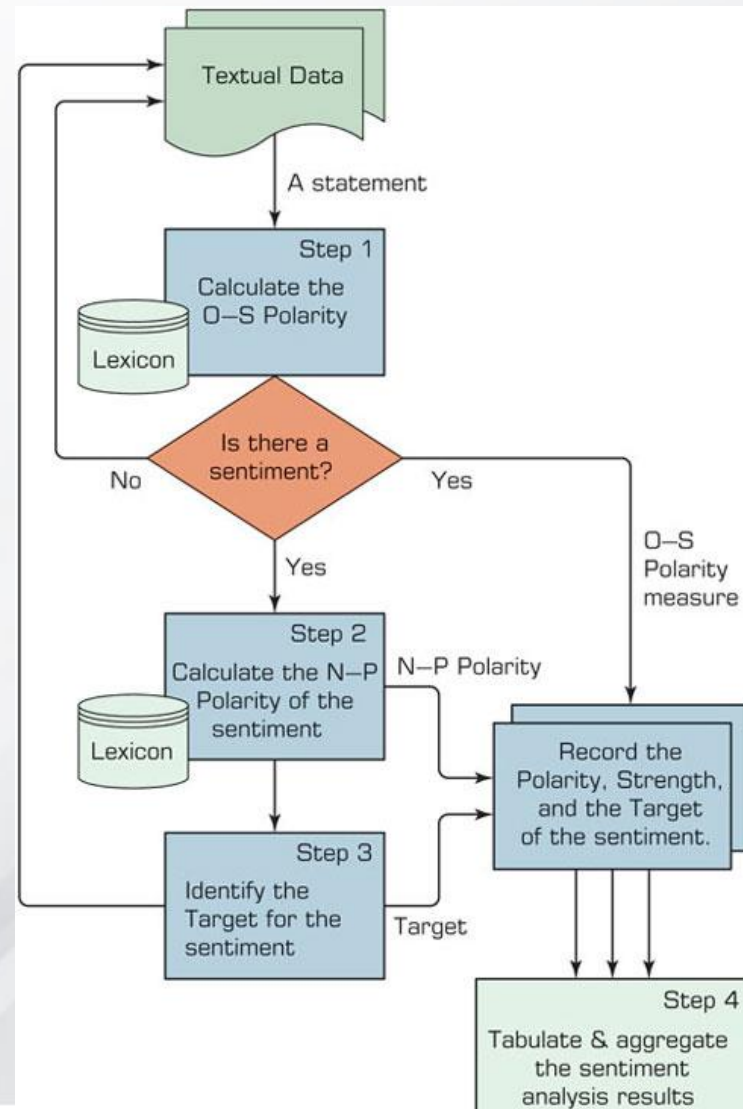
Sentiment Analysis

- Sentiment → belief, view, opinion, and conviction
- Sentiment analysis is trying to answer the question “What do people feel about a certain topic?”
- By analyzing data related to opinions of many using a variety of automated tools
- Used in variety of domains, but its application in CRM are especially noteworthy (which related to customers/consumers’ opinions)

Sentiment Analysis Applications

- Voice of the customer (VOC)
- Voice of the Market (VOM)
- Voice of the Employee (VOE)
- Brand Management
- Financial Markets
- Politics
- Government Intelligence
- ... others

Sentiment Analysis Process (1 of 3)

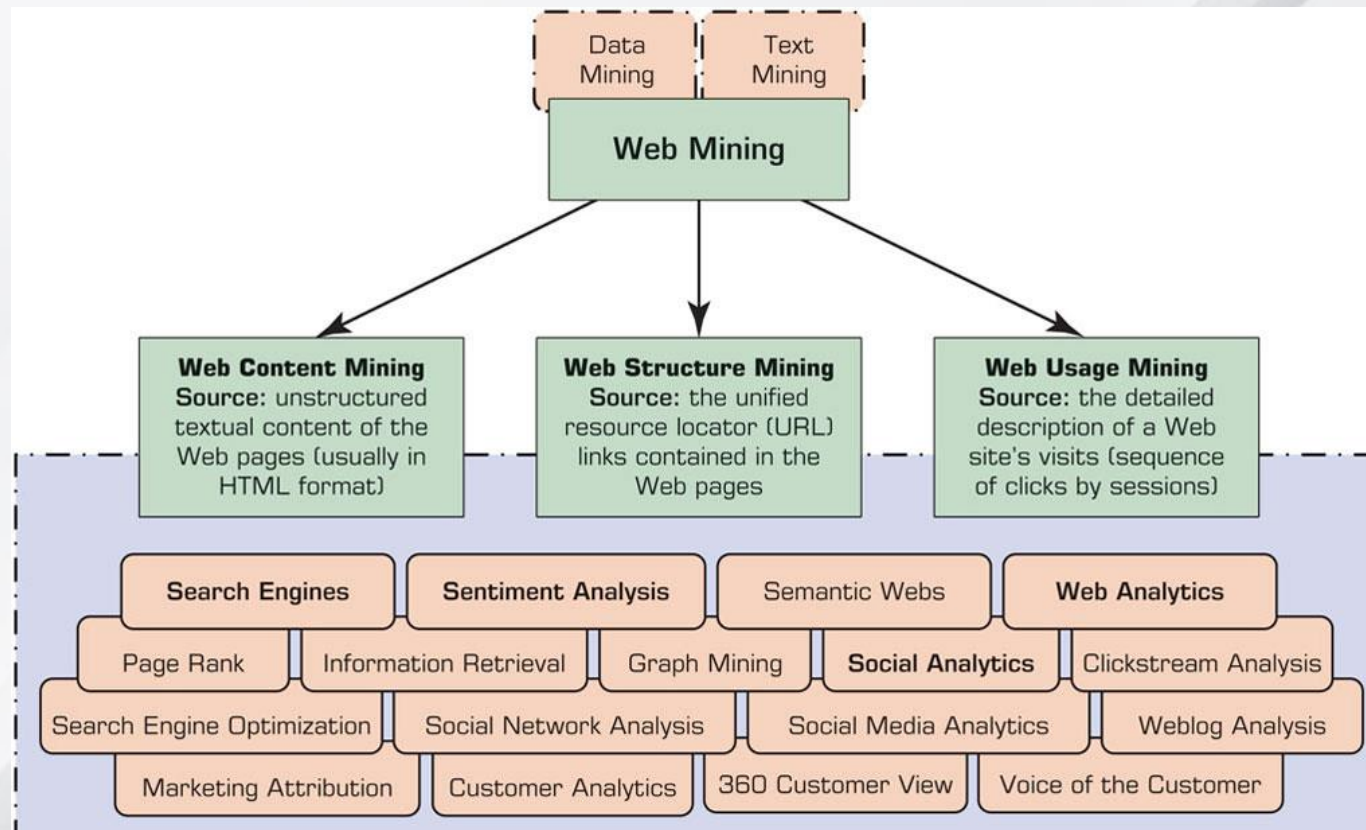


Web Mining Overview

- Web is the largest repository of data
- Data is in HTML, XML, text format
- Challenges (of processing Web data)
 - The Web is too big for effective data mining
 - The Web is too complex
 - The Web is too dynamic
 - The Web is not specific to a domain
 - The Web has everything
- Opportunities and challenges are great

Web Mining

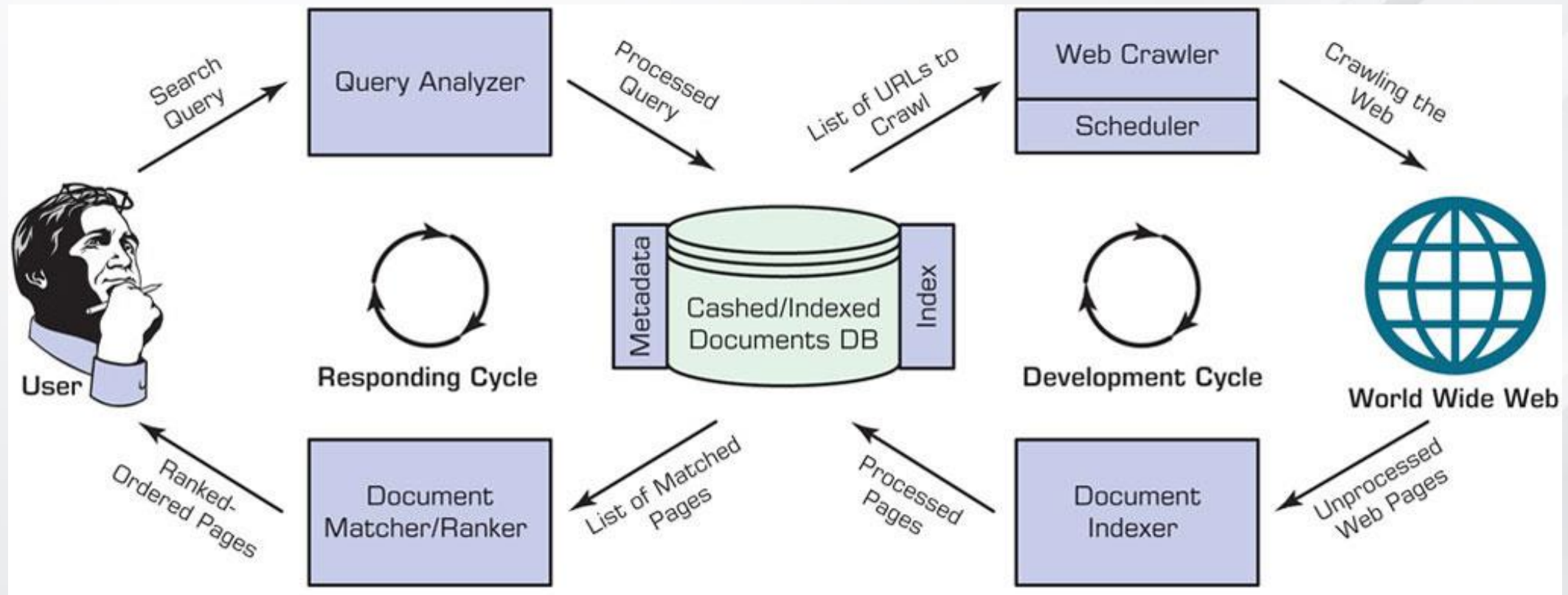
Web mining (or Web data mining) is the process of discovering intrinsic relationships from Web data (textual, linkage, or usage)



Search Engines

- Google, Bing, Yahoo, ...
- For what reason do you use search engines?
- **Search engine** is a software program that searches for documents (Internet sites or files) based on the keywords (individual words, multi-word terms, or a complete sentence) that users have provided that have to do with the subject of their inquiry
- They are the workhorses of the Internet

Structure of a Typical Internet Search Engine



Search Engine Optimization

- It is the intentional activity of affecting the visibility of an e-commerce site or a Web site in a search engine's natural (unpaid or organic) search results
- Part of an Internet marketing strategy
- Based on knowing how a Search Engine works
 - Content, HTML, keywords, external links, ...
- Indexing based on ...
 - Webmaster submission of URL
 - Proactively and continuously crawling the Web

Top 15 Most Popular Search Engines

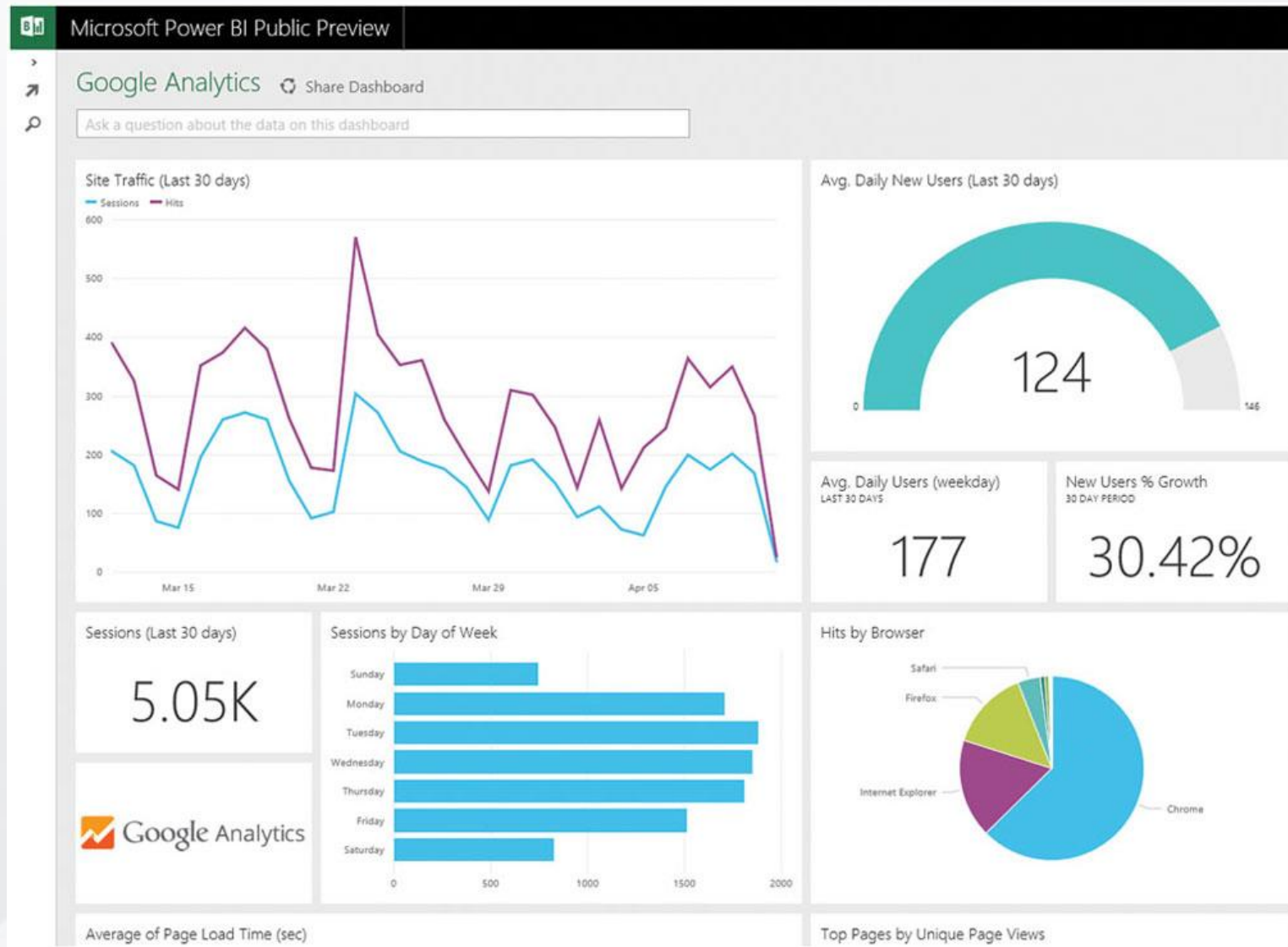
(by eBizMBA, August 2016)

Rank	Name	Estimated Unique Monthly Visitors
1	Google	1,600,000,000
2	Bing	400,000,000
3	Yahoo! Search	300,000,000
4	Ask	245,000,000
5	AOL Search	125,000,000
6	Wow	100,000,000
7	WebCrawler	65,000,000
8	MyWebSearch	60,000,000
9	Infospace	24,000,000
10	Info	13,500,000
11	DuckDuckGo	11,000,000
12	Contentko	10,500,000
13	Dogpile	7,500,000
14	Alhea	4,000,000
15	ixQuick	1,000,000

Web Analytics Metrics (1 of 3)

- Web site usability
 - How were the visitors using my Web site?
- Traffic sources
 - Where did they come from?
- Visitor profiles
 - What do my visitors look like?
- Conversion statistics
 - What does it all mean for the business?

A Sample Web Analytics Dashboard



Social Analytics Social Network Analysis (1 of 2)

- **Social Network** - social structure composed of individuals link to each other
- Analysis of social dynamics
- Interdisciplinary field
 - Social psychology
 - Sociology
 - Statistics
 - Graph theory

Social Analytics Social Network Analysis (2 of 2)

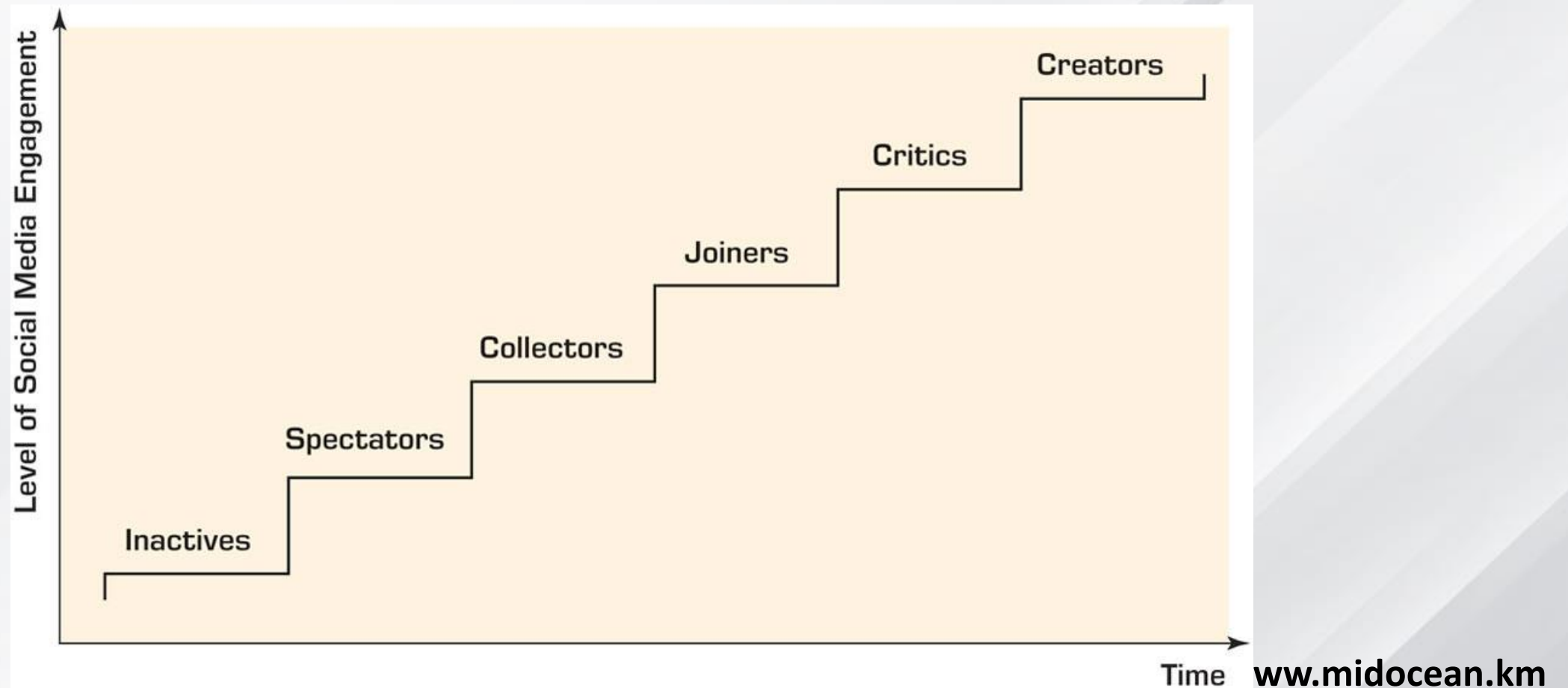
- **Social Networks** help study relationships between individuals, groups, organizations, societies
 - Self organizing
 - Emergent
 - Complex
- Typical social network types
 - Communication networks, community networks, criminal networks, innovation networks, ...

Social Media Definitions and Concepts

- Enabling technologies of social interactions among people
- Relies on enabling technologies of Web 2.0
- Takes on many different forms
 - Internet forums, Web logs, social blogs, microblogging, wikis, social networks, podcasts, pictures, video, and product reviews
- Different types of social media
 - Based on media research and social process

How Do People Use Social Media?

- Different engagement levels



Social Media Analytics

- It is the systematic and scientific ways to consume the vast amount of content created by Web-based social media outlets, tools, and techniques for the betterment of an organization's competitiveness
- Tools to measure social media impact:
 - Descriptive analytics
 - Social network analysis
 - Advanced analytics

Best Practices in Social Media Analytics

- Think of measurement as a guidance system, not a rating system
- Track the elusive sentiment
- Continuously improve the accuracy of text analysis
- Look at the ripple effect
- Look beyond the brand
- Identify your most powerful influencers
- Look closely at the accuracy of your analytic tool
- Incorporate social media intelligence into planning

Review Questions:

- What is the primary difference between text mining and data mining?
- How does sentiment analysis benefit Customer Relationship Management (CRM)?
- What are the main challenges associated with Natural Language Processing (NLP) in text mining?



Thank You