 **ST.ANNE’S**

**COLLEGE OF ENGINEERING AND TECHNOLOGY**

ANGUCHETTYPALAYAM, PANRUTI – 607106.

**QUESTION BANK**

**JULY 2019 - NOV 2019 / ODD SEMESTER**

**BRANCH:** CSE **YR/SEM:** IV/VII **BATCH**: 2016 - 2020

**SUB CODE/NAME:** CS6007 - INFORMATION RETRIEVAL

**UNIT I**

**INTRODUCTION**

**PART – A**

1. What is Information Retrieval? (Nov/Dec 2016)
2. Specify the purpose/role of an IR system. (Nov/Dec 2016)
3. Outline the impact of the web on Information Retrieval. (April/May 2017)
4. What are the two kinds of Information Retrieval systems?
5. What are the issues/major challenges in Information Retrieval systems.
6. List some famous Open source Search engines?
7. What are the components of any search engine?
8. What are the performance measures for search engine?(Nov/Dec 2017, Nov/Dec 2018)
9. Give an example of Information Retrieval process?
10. What are the major sub systems in Information Retrieval process?
11. Draw the broad outline of an Information Retrieval System.
12. Draw the Query Model for Information Retrieval.
13. Define Zipf’s law.(May-19)
14. What is open source software?
15. What is a search engine?
16. Explain difference between data retrieval and information retrieval.
17. List the advantage of open source.
18. Draw the architecture diagram for Web Search using IR.
19. Compare Information Retrieval and Web Search. (April/May 2017)
20. What are the two major dimensions in the model representation?
21. What is the first step towards the knowledge discovery of IR?
22. What is peer-to-peer search?(Nov/Dec 2017, Nov/Dec 2018,May-19)

**PART – B**

**Introduction -History of IR- Components of IR**

1. (a) What is Information Retrieval? Explain the history of Information Retrieval.(7) (April/May 2017)

(b) Explain the purpose of Information Retrieval System. (6)

1. Explain in detail about the components of Information Retrieval System with neat diagram. (13) (Nov/Dec 2016, Nov/Dec 2017)

**Issues - Open source Search engine Frameworks**

1. Explain in detail about need of Open source software.(13)
2. Explain detail about free open source software.(13)
3. Explain detail about Widely used open source software license.(13)
4. Explain any four open source search engine frameworks that use the IR principles.(13)

**The impact of the web on IR - The role of artificial intelligence (AI) in IR:**

1. i) Identify the various issues in IR system.(7) (Nov/Dec 2018,May-19)

ii) Examine the various impact of WEB on IR (6)

1. (a) Differentiate between Information Retrieval and Web search.(6) (Nov/Dec 2017)

(b) Explain the issues in the process of Information Retrieval? (7)

1. Demonstrate the role of Artificial Intelligence in Information Retrieval Systems.(Dec-2018) (13)
2. Explain in detail about the components of Information Retrieval and search engine. (13)

(Nov/Dec 2016, April/May 2018)

**Components of a Search engine- Characterizing the web**

1. Explain in detail about the components of search engine.(13) (April/May 2019)
2. Explain how to characterize the web in detail. (13)(Dec-16)

**PART – C**

1. Create an open source search engine like Google with suitable functionalities.

**UNIT II**

**INFORMATION RETRIEVAL**

**PART – A**

1. Define an Information Retrieval model.(Apr/May 2019)
2. List the retrieval models. (Nov/Dec 2016)
3. Define Document Preprocessing. (Nov/Dec 2016)
4. What is Boolean model?
5. List the pros and cons of Boolean model.
6. Explain Vector Space model.
7. List the process of vector space model.
8. Define Term Weighting and list its two factors.
9. Define tf-idf.
10. Define similarity and various similarity coefficients.
11. What is data preprocessing?
12. What is the need for Preprocessing?
13. List the two key statistics that are used to assess the effectiveness of an IR system.
14. Define Precision and Recall.
15. Define Query Optimization.(May-19)
16. Define Language model.
17. List the types of language models.
18. What is probabilistic retrieval model?
19. Explain Relevance Feedback.
20. What is zone index?(Nov/Dec 2017)
21. Differentiate between relevance feedback and pseudo relevance feedback. ?(Nov/Dec 2018)
22. What is inverted index? (Apr/May 2017) or what is inversion in indexing process? (Nov/Dec 2017)
23. What is Stemming? Give Example.(Apr/May 2017,Nov/Dec 2018)

**PART – B**

**Boolean and vector-space retrieval models**

1. (a) Discuss the Boolean retrieval model in detail with diagram. (7)

(b) Explain vector-space retrieval model. (6). (Apr/May 2017)

1. Explain various information retrieval models in detail.(13) (May-18)

**Term weighting**

1. Develop on example to implement term weighting.(13) (May-19)
2. Briefly explain weighting and cosine similarity.(13) (Nov/Dec 2016)

**Preprocessing - Inverted indices**

1. i) Discuss the structure of inverted indices. (7)

ii) Discuss the searching process in inverted file (6)

**Language Model based IR**

1. Analyze the language model based IR and its probabilistic representation. (13) (Nov/Dec 2018)

**Probabilistic IR –Latent Semantic Indexing**

1. i) ExplainLatent semantic Indexing.(7) .(Apr/May 2017)

ii) Explain about language model. (6)

**Relevance feedback and query expansion**

1. Write about relevance feedback and query expansion. (Nov/Dec 2016,Apr/May 2017,May-19)
2. i) Explain tf-idf Scheme in detail.(7)

ii) Explain in detail about conflation algorithm. (6)

**PART-C**

1. Compose the information Retrieval services of the internet with suitable design.(13)

**UNIT III**

**WEB SEARCH ENGINE – INTRODUCTION AND CRAWLING**

**PART – A**

1. Explain the basic operation of a Web?
2. What is a web graph?(Dec-18)
3. What is a Spam?
4. Why spam is a problem?
5. Explain the difference between CPM and CPC.
6. What is Algorithmic search?
7. Define Lossless vs. lossy compression
8. Explain the three categories of common search queries.
9. Describe benefit of SEO.
10. What is Web Indexing?(May-18)
11. Why compression need?
12. What is the purpose of Web Crawler? (Nov/Dec 2016)
13. Define Web Crawling.(Apr/May 2017,May-19)
14. What are the challenges in Web Crawling?(May-18)
15. What are the features that must be added to any crawler?
16. What are the two major jobs of any Web Crawler? How do you calculate the age of any Crawler?
17. What is focused crawling?
18. What are requirements of XML information retrieval systems? (Nov/Dec 2016)
19. Define Search Engine Optimization.
20. Explain paid submission of search service.(May-19)
21. What are politeness policies used in web crawling.(Nov/Dec 2017)

**PART – B**

**Web search overview**

1. Briefly explain web search architectures?(13) (Dec-16,May-19)

Search engine optimization/spam

1. Discuss the Search Engine Optimization/SPAM in detail.(13)
2. Summarize on the working of WEB CRAWLER with its diagram.(13) (Apr/May 2017,May-18)
3. i)What is P4P? Elaborate on Paid Placement.(7)

ii) What is the purpose of Index Compression?(6)

**Crawling**

1. Describe Meta and focused crawling.(13) (Nov/Dec 2016)
2. Explain in detail about Meta searchers.(13)

**Index Compression**

1. i) What is Index Compression? Explain the various ways to achieve it.(7) (Apr/May2017)

ii) Explain in detail about near duplicate detection. (6)(Dec-18)

**XML retrieval**

1. Explain in detail about XML retrieval? (13)(nov/dec 2016,Nov/Dec 2018,May-19)
2. Explain in detail about: (13)
3. Types of Queries
4. Pattern matching
5. Structural Queries
6. (i) What is the difference between SEO and Pay-Per-Click.(6)

(ii) How to combine Searching and browsing. (7)

**UNIT IV**

**WEB SEARCH – LINK ANALYSIS AND SPECIALIZED SEARCH**

**PART – A**

1. What is link analysis?(May-19)
2. What is a meta search engine?
3. What are the issues in link analysis?
4. What are link analysis solutions?
5. What is Page Rank?
6. Define authorities. (Nov/Dec 2016)
7. What are hubs?
8. What is the approach taken for hubs and authorities?
9. What is HITS?
10. What is ranking?
11. What is relevance?
12. What is the Invisible Web?(Nov/Dec 18)
13. What is relevance feedback?
14. What is collaborative filtering?(Nov/Dec 18)
15. Define Snippets.
16. What are the CLIR techniques can be classified into different categories based on different translation resources?
17. Define user based collaborative Filtering. (Nov/Dec 2016)
18. What do you mean by item based collaborative filtering? (Apr/May 2017)
19. What are the Hadoop Distributed File System?
20. Define MapReduce.(May-19)
21. List the characteristics of MapReduce.(Apr/May 2017, Nov/Dec 2017)
22. What are the limitations of Hadoop/MapReduce?
23. What is Snippet Generation?(Nov/Dec 2017)

**PART – B**

**Link Analysis –hubs and authorities**

1. i)Define Link Analysis and explain in detail.(7)

ii)Describein detail about HUBS and Authorities.(6)

1. What is link analysis? How do you perform link Analysis?(May-19)(13)

**Page Rank and HITS algorithms -Searching and Ranking**

1. Give the concept of PAGE Ranking in detail. (Apr/May 2017)(13)
2. Discuss in detail about HITS Algorithm with necessary examples.(Dec-18)(13)
3. (i) Compare HITS and Page rank in detail. (7) (Nov/Dec 2016)

(ii) Explain in detail about the HITS. (6)

**Searching-Hadoop & Map Reduce-Evaluation**

1. Describe the Searching using hyperlinks and Ranking process in detail with necessary examples. (13)
2. i)Formulate the working of HADOOP . (Apr/May 2017, May-19)(7)

ii) Compose the Map Reduce in detail. (6)

1. Explain in detail about TREC collection.(13)

**Collaborative filtering and content-based recommendation of documents and products**

1. Explain in detail about Collaborative filtering.(13)(Nov/Dec- 2018)
2. i)Analyzecontent based recommendations of documents and products.(7)

ii) How to handle the invisible web? (6)

**Snippet generation- Cross- Lingual Retrieval**

1. i)Describe in detail about of SNIPPET Generation along with example.(6)

ii) Explain in detail about Cross- Lingual Retrieval. (7) (Nov/Dec 2016)

**UNIT V**

**DOCUMENT TEXT MINING**

**PART – A**

1. Define Information Filtering
2. List the two major approaches for Information Filtering.
3. What are the characteristics of information filtering? (Nov/Dec 2016)
4. Define Text Mining.(May-19)
5. Explain text classification.
6. Explain Naïve Bayes text classification.
7. Explain Bernoulli model.
8. List the properties of Naïve Bayes theorem.
9. Define Feature Selection.
10. Define Text clustering.(Dec-18)
11. Explain K-Means in detail.(May-19)
12. Define Model based clustering.
13. Define Expectation-Maximization algorithm.
14. List the two types of hierarchical clustering.
15. List the four different agglomerative algorithms.
16. Explain top-down clustering or divisive clustering.
17. What are the desirable properties of a clustering algorithm? (Nov/Dec 2016)
18. What is a decision tree?
19. Explain difference between information filtering and information Retrieval.(May 2017,Dec-18)
20. What are the desirable properties of a clustering algorithm? (Apr/May 2017)
21. State Bayes rule.(Nov/Dec 2017)
22. Differentiate supervised learning and unsupervised learning. (Nov/Dec 2017)
23. What is dendrogram?(Nov/Dec 2017)

**PART – B**

**Information filtering**

1. Explain in detail about web data mining. (13)
2. i) Explain in detail about Information filtering.(May-18)(7)

ii) Explain difference between information filtering and information Retrieval. (6)

**Text Mining**

1. i) Expressthe process of Text Mining. (5)

ii) Examine the process of Mining with detailed example. (8)

1. i) Discuss in detail about Text Classification.(7)

ii) SummarizeText Clustering in detail**.** (6)

**Categorization algorithms**

1. Explain in detail about the working of Naïve Bayesian classifier with an example.(Nov/Dec2016,Apr/May 2017,Dec-18,May-19)
2. Explain in detail about nearest neighbours algorithm.(May-18)(13)

**Clustering algorithms**

1. i) Analyze the procedure involved in Expectation Maximization along with the steps involved in it. (7) (Dec-16,Dec-18,May-19)

ii) Explain in detail K-means clustering. (Dec-18)(13)

1. Explain in detail about hierarchical clustering. (13)
2. Rank the impacts of Categorization and clustering of text in the mining with the suitable examples.
3. Explain in detail about decision trees.