

## CS8080 INFORMATION RETRIEVAL TECHNIQUES

### IMPORTANT QUESTIONS AND QUESTION BANK

#### UNIT-I INTRODUCTION

##### 2-Marks

1. Define information retrieval.
2. Identify the need of information retrieval.
3. List and explain the components of IR block diagram.
4. List the fundamental concepts in IR.
5. Express the need of tiered indexes.
6. Interpret the role of Artificial Intelligence (AI) in IR.
7. Differentiate data retrieval and information retrieval.
8. Give the components of search engine and the performance measures
9. What is an extractor?
10. Show the issues that affects IR.
11. Give the purpose of query interface.
12. Summarize the quires of IR.
13. Design the IR architecture diagram.
14. State the impact of WEB on IR.
15. What is search engine?

##### 13-Marks

1. 1.summarize the history of information retrieval.  
2.Explain the purpose of information retrieval system
2. Describe the various components of information retrieval system with neat diagram.
3. 1.Define information retrieval system and its features.  
2.formulate the working of search Engine.
4. 1. Identify the various issue in IR system.  
2.Examine the various impact of WEB on IR.
5. Demonstrate the framework of open-source search engine with necessary diagrams.
6. Compare in detail information retrieval and web search with example
7. Develop the role of Artificial Intelligence in information retrieval system.
8. Describe the various components of a search engine.
9. Describe the different stages of IR system.
10. 1.Demonstrate the working of IR architecture with a diagram.  
2. infer how designing parsing and scoring functions works in detail.
11. Describe in detail the IR system, fundamental concepts, need and purpose of the system.

12. Explain how to characterize the web in detail.
13. Explain the different types of computer software used in computer architecture.
14. Demonstrate database and information retrieval with example.
15. Generalize the process of search engine in detail.

## **UNIT-II MODELING AND RETRIEVAL EVALUATION**

### **2-Marks**

1. Identify probabilistic information retrieval.
2. Analyze the Boolean model.
3. Construct the vector space model representation.
4. List the classes of retrieval model.
5. Define retrieval model.
6. Explain language modelling with example.
7. Illustrate similarity measure.
8. Analyze the problems in lexical semantics.
9. Demonstrate language model and Naïve Bayes.
10. Formulate the Bayesian rule.
11. What is meant by sparse vector?
12. Design an inverted file with an example.
13. Evaluate the goals of LSI.
14. What is smoothing and stemming?
15. What is meant Zone Index?

### **13-Marks**

1. Describe the document pre-processing steps in detail.
2. Illustrate the vector space retrieval model with example.
3. Describe about basic concepts of cosine similarity.
4. Develop an example to implement term weighting (min docs=5).
5. Discuss the Boolean retrieval in detail with diagram.
6. Discuss in detail about term frequency and inverse document frequency.
7. Explain latent semantic indexing and latent semantic space with an illustration.
8. 1. Examine how to form a binary term-document incidence matrix.  
2. Give an example for an above.
9. Describe document pre-processing and its stages in detail.
10. Discuss the structure of inverted indices and the basic Boolean Retrieval model.
11. Discuss the searching process in inverted file.
12. Estimate sparse vectors and its efficiency with diagram.

13. Compare language model-based information retrieval and its probabilistic representation.
14. Explain in detail about binary independence model for probability ranking principles.
15. Illustrate the following:
  1. Probabilistic relevance feedback.
  2. pseudo relevance feedback.
  3. indirect relevance feedback.

### **UNIT-III TEXT CLASSIFICATION AND CLUSTERING**

#### **2-Marks**

1. Integrate the problem of k-means method.
2. Define the characterization of text classification.
3. Summarize Evaluation metrics with example.
4. What are the types of data in clustering analysis?
5. Point out the advantages and disadvantages of decision tree algorithm.
6. Show the applications SVM classifiers
7. Illustrate the advantages of Naïve Bayes.
8. Assess how to measure distance of cluster?
9. Distinguish supervised learning and Unsupervised learning/
10. Summarize the good clustering approaches.
11. Define supervised and Unsupervised algorithms.
12. What is hash-based Dictionary in indexing?
13. Describe two types of indexes in detail.
14. Explain Sequential search in detail.
15. Describe Brute Force in Sequential search

#### **13-Marks**

1. 1. Define Topic detection and tracking, clustering in TDT.  
2. Examine in detail about cluster analysis in text Clustering.
2. Define clustering in metric space with application to information retrieval.
3. Evaluate the agglomerative clustering and HAC in detail.
4. 1. Summarize on clustering Algorithms  
2. Evaluate on the various classification methods of text.
5. 1. Analyze the working of nearest neighbor algorithms along with one representation.  
2. Analyze the K-means Clustering method and the problem in it.
6. Analyze about Decision tree algorithm with illustration.

7. Examine inverted index and forward index.
8. Discuss in detail about text classification.
9. Apply Naïve Bayes algorithm for an example.
10. Demonstrate its working in detail.
11. Analyze single dimension index in detail.
12. Define Knuth Morris Pratt algorithm in detail.
13. 1. construct B+ tree index in detail.  
2. summarize the significance of SVM classifiers in detail.
14. Examine single dimensional and multi-dimensional index.
15. Explain Sequential search in detail.

### **UNIT-IV WEB RETRIEVAL AND WEB CRAWLING**

#### **2-Marks**

1. Express the basics of web search with a neat diagram.
2. Explain pay for placement.
3. What is meant by search engine optimization?
4. List the need of web search engine.
5. Demonstrate the architecture of search engine.
6. Compare parallel crawler and meta crawler.
7. List the SPAM techniques.
8. Evaluate use of inversion in indexing process.
9. State the issues in search engines.
10. Design the politeness policies used in web crawler
11. Classify the ways to identify duplication.
12. How to apply duplicate deduction to web pages?
13. Assess the need for keyword stuffing.
14. Identify the application of web crawlers.
15. What is focused crawler?

#### **13-Marks**

1. Demonstrate the search engine optimization/SPAM in detail?
2. Describe in detail about vector space model for XML retrieval
3. List the types of search engine and explain them.
4. Distinguish visual vs programmatic crawler.
5. Design and develop a web search architecture and the components of search engine and its issues.
6. Recommend the need for near duplication detection by the way of finger point algorithm.
7. Summarize the process of index compression in detail.
8. Examine the need for web search engine.
9. Explain the overview of web search.

10. What is P4P? elaborate on paid placement.
11. Demonstrate about search engine optimization.
12. List the challenges in date traversing by search engine and how will you overcome
13. Describe the following with example: bag of words and shingling
14. Illustrate the following:
  1. focused crawling
  2. Deep web
  3. Distributed crawling
  4. sitemap.
15. Analyse on URL normalization.

### **UNIT-V RECOMMENDER SYSTEM**

#### **2-Marks**

1. Examine the board classification of recommendation systems?
2. Justify content-based recommendation systems.
3. Classify collaborative filtering system
4. Define knowledge-based Recommendation.
5. Give the definition hybrid recommendation.
6. Define meta level and cascade Recommendation system
7. Examine knowledge-based configuration in detail.
8. Define web Recommendation in detail.
9. Describe hybrid Recommendation system.
10. What are the types of hybrid Recommendation system?
11. Describe Recommendation techniques in detail.
12. Describe weighted recommenders
13. Construct collaborative and content-based recommendation systems.
14. Estimate some examples for content-based recommendation systems.
15. Examine disadvantage content-based recommendation systems

#### **13-Marks**

1. Define recommendation based on user rating using appropriate example
2. Explain the technique of matrix factorization.
3. Estimate the content-based recommendation systems.
4. Differentiate collaborative filtering and content-based system.
5. Explain about high level architecture.
6. Explain in significance of collaborative filtering in detail.
7. Illustrate the advantages and disadvantages of content-based and collaborative filtering recommendation systems.
8. Describe the knowledge-based recommendation systems in detail
9. Difference between hybrid and collaborative Recommendation.

10. Define the steps involved in collaborative filtering.
11. Describe web-based recommendation system.
12. When can be collaborative filtering be used?
13. Define in detail about matrix factorization models.
14. Discuss neighboring model in detail.
15. Discuss the approaches of recommender system.

www.binils.com