CS6007-INFORMATION RETRIEVAL

Question Bank

Unit-I

Introduction

Part A

1) Define information retrieval.

2) Explain difference between data retrieval and information retrieval.

3) List and explain components of IR block diagram.

4) What is objective term and nonobjective term?

5) Explain the type of natural language technology used in information retrieval.

6) What is search engine?

7) What is conflation?

8) What is an invisible web?

9) Define Zipf's law.

10) What is open source software?

11) What is proprietary software?

12) What is closed software?

13) List the advantage of open source.

14) List the disadvantage of open source.

15) What are the reasons for selecting open software?

16) What do you mean by Apache License?

17) Explain features of GPL version2.

Part B

1) i) Explain the components of IR in detail. (8)

   ii) Explain the centralized architecture in detail. (8)

2) Explain the steps HITS algorithm with hubs and authorities. (16)

3) Explain the following (16)

   i) Measuring the web.

   ii) Characterizing the web

4) i) with neat diagram , Explain the components of information retrieval (8)

   ii) Distinguish between IR versus Web search (8)

5) Write on the open search engine frameworks with neat diagram. (16)

6) Discuss about the various ways in characterizing the web and its impact with IR. (16)

Unit-II

INFORMATION RETRIEVAL

Part A

1. What do you mean information retrieval models?

2. What is cosine similarity?

3. What is language model based IR?

4. Define unigram language.

5. What are the characteristics of relevance feedback?

6. What are the assumptions of vector space model?

7. What are the disadvantages of Boolean model?

8. Define term frequency.

9. Explain Luhn's ideas

10. Define stemming.

11. What is Recall?

12. What is precision?

13. Explain Latent semantic Indexing.

Part B

1) Explain on latent semantic indexing. (16)

2) Differentiate the various query expansion method with relevance feedback. (16)

3) Explain probabilistic IR with neat derivations. (16)

4) i) what are the steps in preprocessing (8)

   ii) Explain about Inverted indices (8)

Unit-III

WEB SEARCH ENGINE – INTRODUCTION AND CRAWLING

Part A

1. Define web server.

2. What is web Browsers?

3. Explain paid submission of search service.

4. Explain paid inclusion programs of search services.

5. Explain in pay-for-placement of search services.

6. Define Search Engine Optimization.

7. Describe benefit of SEO.

8. Explain the difference between SEO and Pay-per-click

9. What is web crawler?

10. Define focused crawler.

11. What is hard and soft focused crawling?

12. What is the Near-duplicate detection?

13. What are requirements of XML information retrieval systems?

Part B

1) i) Explain about the web structure and paid placement (8)

   ii) Explain about crawling and Meta crawlers (8)

2) i) Write short notes on Index compression (8)

   ii) Explain about crawling and Meta crawlers (8)

3) Discuss the detailed architecture of search engine optimization/spam   (16)

4) Explain how Web search architecture is implemented for near duplicate detection. (16)

UNIT IV

WEB SEARCH – LINK ANALYSIS AND SPECIALIZED SEARCH

1. What is link analysis?

2. What is in query independent ranking?

3. What is query dependent ranking?

4. Define authorities?

5. Define hubs.

6. What is Hadoop?

7. What is the Hadoop Distributed File System?

8. Define MapReduce.

9. List the characteristics of MapReduce?

10. What are the limitations of Hadoop/Map Reduce?

11. What is Cross-Lingual Retrieval?

12. Define Snippets.

13. List advantages of invisible web content.

14. What is collaborative filtering?

15. What do you mean by item-based collaborative filtering?

16. What are problem of user based CF?

17. Define user based collaborative Filtering.

Part B

1) Explain on relevance scoring and ranking of web with block diagram.   (16)

2) Use an example and explain the method of Page ranking algorithm   (16)

3) Explain about Collaborative filtering.   (16)

4) What is Cross Lingual Retrieval? Explain its role in IR.   (16)

5) Explain model of Hadoop and map reduce.   (16)

6) Explain briefly about   (16)

        i)   Personalized search

        ii)   Page Rank

7) i) Explain focused crawling in detail.   (10)

    ii) Explain index compression in detail.   (6)

UNIT V

DOCUMENT TEXT MINING

1. What do you mean by information filtering?
2. What are the characteristics of information filtering?
3. Explain difference between information filtering and information Retrieval.
4. What is text mining?
5. What is classification?
6. Explain clustering.
7. What are the desirable properties of a clustering algorithm?
8. What is decision tree?
9. List the advantages of decision tree.
10. List the disadvantages of decision tree
11. What is supervised learning?
12. What is unsupervised learning?
13. What is dendrogram?

Part B

1) Explain XML retrieval method in detail.   (16)

2) i) With neat diagram explain the meta-searchers in detail.   (10)

   ii) Explain the various types of paid search services in detail.   (6)

3) Discuss the various text classification techniques in detail.   (16)

4) Discuss the naive Bayes and nearest neighbor algorithm in detail.   (16)

5) Discuss the various clustering techniques in detail.   (16)