

**Fifth semester B.E., CSE / AI&DS, AIML Examinations 2024-25**

**Information Retrieval - S5CCSPE01**

**QUESTION BANK**

Q.No.		BL	CO	PO	PSO	Marks
<b>UNIT-1</b>						
1.	Discuss any 4 types of search in the process of information retrieval.	2	1	2	2	4
2.	Explain the major challenges in the design of information retrieval systems.	4	1	2	2	6
3.	Highlight the critical requirements for a search engine.	4	1	2	2	6
4.	With a neat diagram explain the architecture of a search engine.	2	1	2	1	8
5.	What is inverted index. How is it built? Draw the inverted index that would be built for the following document collection. Doc 1 - new home sales top forecasts Doc 2 - home sales rise in july Doc 3 - increase in home sales in july Doc 4 - july new home sales rise	3	1	2	2	08
6.	With a suitable example explain how a Boolean query is processed using inverted index.	3	1	2	2	06
7.	What are the challenges in tokenization? Explain	2	1	2	2	06
8.	What is token normalization? Explain with suitable examples.	2	1	2	2	06
9.	How are skip pointers used with postings lists? What is the advantage of using such pointers? Write a suitable algorithm to implement skip pointers in the inverted index.	4	1	2	2	08
10.	Differentiate between stemming and lemmatization with suitable examples.	3	1	2	2	08
11.	Write the algorithm for intersection of two postings. How is it useful in Boolean retrieval ?	3	1	2	2	06
12.	Assume certain availability of words w1,w2,.....w6 in documents d1,d2-----d5. Draw the term-document incidence matrix. Given the query w3 AND w5 AND NOT w6, determine the documents satisfying the query.	3	1	2	2	06
13.	List any two issues related to the following i) Choosing a document unit ii) Tokenization	2	1	2	2	08
14.	How does a term-document incidence matrix help in handling Boolean queries? Explain with an example. What are its limitations.	4	1	2	2	08
15.	How are positional indexes created? How do they assist in processing phrase queries?	4	1	2	2	06

16.	For the current requirement of free text queries in IR systems, what additional features can be provided on top of Boolean retrieval model.	4	1	2	2	06
17.	What are stop words? How are they handled in various IR systems?	2	1	2	2	06
18.	How do biword indexes help in processing phrase queries? Explain with suitable examples	3	1	2	2	06
19.	How are positional indexes created? How do they assist in processing phrase queries?	3	1	2	2	06

Q.No.		BL	CO	PO	PSO	Marks
UNIT-2						
1.	Demonstrate how a binary search tree is suitable for storing and retrieving vocabulary terms.	4	2	1	2	08
2.	What is the importance of edit distance in spelling correction? Write the algorithm for computing edit distance.	3	2	1	2	06
3.	With a suitable example explain how distributed indexing is achieved using Map Reduce technique.	3	2	1	2	06
4.	How are B-trees used in handling trailing and leading wild card queries?	3	2	1	2	08
5.	How is phonetic correction achieved in Information Retrieval Systems?	3	2	1	2	06
6.	Why is SPIMI (Single pass in memory indexing) considered as more scalable than blocked sort based indexing?	4	2	1	2	06
7.	Discuss the suitability of Binary search tree and B tree for dictionaries.	4	2	1	2	08
8.	What is edit distance? Determine edit distance between FRIED and FRESH using Levenshtein edit distance computation.	3	2	1	2	08
9.	Discuss the suitability of Distributed indexing for index construction? How is it achieved using map reduce technique?	4	2	1	2	08
10.	What are permuterm indexes? How are they used in handling general wild card queries?	4	2	1	2	06
11.	How are k-gram indexes constructed? How are they used in handling wild card queries?	4	2	1	2	06
12.	How is Context sensitive spelling correction achieved in IR systems?	3	2	1	2	06
13.	Why is there a need for Phonetic correction in IR systems? Write an algorithm to achieve it.	4	2	1	2	08
14.	What computer hardware characteristics are considered in designing an IR system? Explain	2	2	1	2	06
15.	How does BLOCKED SORT-BASED INDEXING handle construction of index for very large collections of documents? Write the algorithm for the same.	2	2	1	2	08
16.	Compare the following approaches of Dictionary compression: i) Dictionary as a string ii) Blocked storage iii) Blocked storage and front coding	2	2	1	2	08

17.	How does Variable byte codes approach provide Postings file compression? Explain with an example	2	2	1	2	06
18.	How do $\gamma$ codes provide Postings file compression? Explain with an example	2	2	1	2	06

Q.No.		BL	CO	PO	PSO	Marks
<b>UNIT-3</b>						
1.	How is the weighted zone score computed from two postings lists? Write the algorithm for the same?	3	3	3	2	08
2.	What is a champion list? How is it created for a term $t$ ?	3	3	3	2	06
3.	How can vector space model be used in Boolean Retrieval schemes?	3	3	3	2	06
4.	What is inverse document frequency? What is its importance in IR system? How is it computed?	4	3	3	2	08
5.	What is cluster pruning? Demonstrate with suitable examples.	3	3	3	2	06
6.	Explain why vector space model cannot be used to handle phrase queries.	3	3	3	2	06
7.	What are parametric and zone indexes? Explain with suitable examples	2	3	3	2	08
8.	How is Pivoted document length normalization done. Explain with relevant diagrams.	2	3	3	2	08
9.	How does <b>Tf-idf weighting</b> help to produce a composite weight for each term in each document? Write the relevant expressions.	4	3	3	2	06
10.	How is precisely retrieving the $K$ highest-scoring documents for a query possible? Write the algorithm for the same.	2	3	3	2	08
11.	Write a suitable approach such that $K$ documents are produced that are <i>likely</i> to be among the $K$ highest scoring documents for a query.	3	3	3	2	06
12.	What is impact ordering? Explain the same considering term frequency(tf) as a parameter.	2	3	3	2	06
13.	What are tired indexes? How are they constructed?	2	3	3	2	06
14.	How are parsing and scoring functions designed for free text queries? Explain	3	3	3	2	08
15.	With a neat diagram, explain the components necessary for a basic search system.	2	3	3	2	08

Q.No.		BL	CO	PO	PSO	Marks
<b>UNIT-4</b>						
1.	How is the evaluation of ranked retrieval results done? Explain.	3	4	3	2	8
2.	Discuss the challenges in XML retrieval.	2	4	3	2	6
3.	Compare language modelling with other approaches in Information Retrieval.	4	4	3	2	6
4.	What are the requirements to evaluate an IR system? Discuss the evaluation parameters for unranked retrieval sets.	4	4	3	2	8
5.	Compare text centric with data centric XML retrieval.	3	4	3	2	6
6.	Describe the query likelihood model.	3	4	3	2	6

7.	Discuss the requirements to evaluate an IR System	2	4	3	2	6
8.	What is relevance assessment? How is it achieved using pooling and Kappa statistic	3	4	3	2	8
9.	What is a snippet w.r.t IR system? How is a static snippet different from a dynamic snippet?	2	4	3	2	6
10.	Describe a suitable approach to refine a deployed IR system.	3	4	3	2	6
11.	Apart from formal evaluation measures, how does <b>System quality and user utility impact the rating of</b> IR systems?	4	4	3	2	8
12.	Describe the XML DOM Model.	2	4	3	2	8
13.	Describe the Vector space model for XML retrieval.	2	4	3	2	8
14.	Explain how XML retrieval is evaluated.	2	4	3	2	6
15.	Discuss the challenges in using language models.	2	4	3	2	6
16.	Describe the extended language model approaches for IR.	2	4	3	2	6

Q.No.		BL	CO	PO	PSO	Marks
<b>UNIT-5</b>						
1.	How do search engines index the web pages? Discuss the capture recapture method to estimate the size of index .	3	5	3	2	8
2.	What is a web crawler? Discuss the features of a web crawler.	3	5	3	2	6
3.	Describe the procedure to compute subset of web pages.	2	5	3	2	6
4.	Why duplication is an issue in determining index size? Discuss a method to detect duplicate web pages.	3	5	3	2	8
5.	What is the role of connectivity servers in web search engines?	4	5	3	2	6
6.	With a suitable example demonstrate topic specific page rank computation.	3	5	3	2	6
7.	Discuss any 3 major web characteristics that has made web search crucial.	2	5	3	2	6
8.	Compare web graph and Bowtie structure of web.	2	5	3	2	6
9.	What is a spam? Discuss the technique of cloaking and the role of SEOs in combatting spams.	4	5	3	2	8
10.	Compare CPC model and Goto model used in Web Advertising.	2	5	3	2	8
11.	How are pure search engines and sponsored search engines combined in the current search engines.	4	5	3	2	6
12.	How is Google able to attract more user traffic? Explain	4	5	3	2	6
13.	How are web search queries categorized? List and explain each category.	2	5	3	2	8
14.	With a neat diagram explain the various components of a web search engine.	2	5	3	2	8
15.	Outline the attempts made in sampling web pages.	2	5	3	2	6
16.	Explain the random queries sampling approach. Why is it considered as one of the best sampling approaches.	4	5	3	2	8
17.	What are near duplicates? Discuss a solution to detect near duplicates.	2	5	3	2	8
18.	Describe the web crawling operation? With a neat diagram explain the basic crawler architecture.	2	5	3	2	8

19.	Discuss the two implementations related to distribution of the indexes across a large computer cluster.	2	5	3	2	6
20.	What is the importance of link analysis in web search engines.	4	5	3	2	6
21.	What is the importance of the following in link analysis : i) Page Rank ii) Markov Chain	4	5	3	2	8
22.	Justify that Pagerank is a query independent measure of the static quality of each page.	4	5	3	2	6
23.	Differentiate between hub score and authority score w.r.t. web pages.	2	5	3	2	6