

(2 ½ Hours)

[Total Marks: 75]

- N.B. 1) All questions are compulsory.
2) Figures to the right indicate marks.
3) Illustrations, in-depth answers and diagrams will be appreciated.
4) Mixing of sub-questions is not allowed.

Q. 1 Attempt All

(a) Select the correct alternative from the options given:

(10M)

- (i) One of the main issues of IR is _____
(a) Scale (b) Size
(c) Quality (d) Relevance
- (ii) Dictionary data structure can be _____.
(a) Hash table and Tree (b) Linked list and Tree
(c) Arrays (d) Stack
- (iii) _____ was named after the Larry page, one of the founders of Google.
(a) Page Rank (b) Ranking
(c) Permuterm (d) Boolean Retrieval
- (iv) _____ function is responsible for consolidating the results produced by each of the map() function
(a) Mapper() (b) Reducer()
(c) Reduce() (d) Map()
- (v) PPC stands for _____
(a) Pay per context (b) Pay Per customer
(c) Pay per click (d) Pay per content
- (vi) SERP stands for _____
(a) Search Engine Records Page (b) Search Engine Result Page
(c) Search Engine Ranked Page (d) Search Engine Repeated Page
- (vii) Web search engines stores the information about many web pages by a _____.
(a) Web Indexer (b) Web Organizer
(c) Web Router (d) Web Crawler

(viii) The _____ is categorized as an unknown segment of the Deep Web which has been purposely kept hidden & is inaccessible using standard web browsers.

- (a) Haunted web (b) Dark web
(c) World Wide Web (d) Surface web

(ix) _____ filtering recommends the products which are similar to the ones the user has liked in the past.

- (a) Context Based filtering (b) Collaborative filtering
(c) Content based filtering (d) None of these

(x) Recommendation system uses _____.

- (a) User's History of search (b) User's Profile
(c) Both users history and users profile (d) Neither users history nor users profile

(b) **Fill in the blanks by selecting from the pool of options:** (5M)

(position, surface web, Isolated term distance, 1, Corpus, degree, dark web, Levesthein distance, Lexicon, 0)

(i) The normal google search is a part of _____

(ii) Given the two strings s1 and s2 the edit distance between them is known as _____

(iii) A similarity measure is a function that computes the _____ of similarity between two vectors.

(iv) A dictionary of terms is sometime also referred as _____.

(v) Initially Page Rank of all web pages is _____.

Q. 2 Attempt the following (Any THREE) (15M)

(a) Define Information Retrieval. Discuss and differentiate between the terms precision and recall with respect to information retrieval.

(b) Discuss the issues and challenges in information retrieval.

(c) Draw the inverted index for following document.

Doc1-new home sales top forecast

Doc2-home sales rise in july

Doc3-increase in home sales in july

Doc4-july new home sales rise

(d) Explain how k-gram index can be used for spelling correction.

(e) With the help of a diagram, explain the components of an information retrieval system.

(f) Write a short note on phonetic correction.

Q. 3 Attempt the following (Any THREE) (15M)

- (a) Write a short note on Cross – Lingual Information Retrieval.
- (b) Explain the concept of collaborative filtering, with the help of an example.
- (c) Discuss the HITS Algorithm in detail.
- (d) What is invisible web? Explain in detail.
- (e) What is text summarization? Discuss the stages of summarization.
- (f) Write a short note on the Page Rank Algorithm.

Q. 4 Attempt the following (Any THREE) (15)

- (a) Explain the web search architecture.
- (b) State and explain the challenges in XML retrieval.
- (c) Discuss the characteristics of web search engine?
- (d) Write a short note on Search Engine Optimization.
- (e) Explain the vector space model for XML retrieval.
- (f) List and explain various SPAM Techniques.

Q. 5 Attempt the following (Any FIVE) (15)

- (a) Compare Information retrieval and Data retrieval.
- (b) Define following terms
 - i) Posting list
 - ii) Corpus
 - iii) Lexicon
- (c) Differentiate between Content based and Collaborative Filtering
- (d) Explain the following in brief:
 - i) hubs
 - ii) authorities
- (e) Discuss the challenges in Question and Answering.
- (f) Define the following
 - i) cloaking
 - ii) stopping
 - iii) stemming
- (g) Discuss the role of XML in Information Retrieval.
- (h) Compare Text Centric vs Data Centric XML.
