

- 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 (Each of 5Marks) (15M)

(a) Multiple Choice Questions (5M)

i) The group of documents over which user perform retrieval as the _____

- a. terms b. collection c. information d. data

ii) **SEO stands for** _____

- b. Search engine order b. Search engine organizer
 c. Search engine optimization d. Search engine option

iii) **CPC stands for** _____

- d. Cost per click b. Cost per customer
 e. Cost per coverage d. Cost per content

iv) _____ **filtering recommends products which are similar to the ones that a user has liked in the past.**

- f. Collaborative based b. Context based
 g. Collection based d. Content based

v) _____ **is the fraction of the relevant documents in the collection returned by the system.**

- h. Recall b. Reconnect c. Reciprocal d. Retrieved

(b) Fill in the blanks (5M)

(Stemmer, Ranking , PageRank, Google, Permuterm)

i) _____ was named after Larry Page, one of the founders of Google.

ii) The _____ component of the search engine transformed query from the user interaction component and generates a ranked list of documents using scores based on a retrieval model.

iii) The first web search engine known to apply link analysis on a large scale was _____.

iv) The first special index for general wildcard queries is the _____.

v) The task of the _____ is to group words that are derived from a common stem.

(c) Short Answers- Define the following terms: (5M)

- i) Boolean retrieval model
 ii) Invisible Web
 iii) Cloaking
 iv) XML schema
 v) Crawler

Q. 2 Attempt the following (Any THREE)(Each of 5Marks) (15M)

- (a) What is the need of Information retrieval? Discuss.
- (b) Brief the history of Information retrieval.
- (c) What are the components of Information retrieval? Explain.
- (d) List the forms of spelling correction in Information retrieval. Explain.
- (e) Give in detail the technique for tolerant retrieval.
- (f) 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

Q. 3 Attempt the following (Any THREE) (Each of 5Marks) (15M)

- (a) Discuss in detail Hubs and Authorities.
- (b) Define Page rank. How to compute page rank for a webpage? Give example.
- (c) Write a note on Personalized search.
- (d) Explain the concept of Collaborative filtering.
- (e) What are snippet generation and two kinds of summary? Explain.
- (f) Give the meaning of cross lingual retrieval. Analyse its process.

Q. 4 Attempt the following (Any THREE) (Each of 5Marks) (15M)

- (a) Brief the structure of Web.
- (b) Discuss categories of user needs in web queries for query analysis.
- (c) Write a note on Spam.
- (d) What are the basic building blocks of Search Engine Architecture? Explain.
- (e) Differentiate Text Centric vs. Data Centric XML retrieval.
- (f) Discuss the challenges in XML retrieval.

Q. 5 Attempt the following (Any THREE) (Each of 5Marks) (15M)

- (a) Compute the Levenshtein edit distance between "paris" and "alice".
- (b) Write a note on paid placement.
- (c) Discuss evaluation of XML retrieval.
- (d) What is MapReduce? Explain its paradigm.
- (e) Briefly explain the concept of Question answering in IR.
