TYBSC Computer Science **Assignment**

1. Describe the aim and purpose of Question Answering in detail.
2. Evaluate the concept of Personalized Search and its methodology
3. Analyse the process of cross lingual retrieval
4. Discuss the concept invisible web.
5. Elaborate on summarization with its types.
6. What is snippet generation? Consider the document with 500 sentences. Interpret that the word “activity” with frequency 35 is significant or not.
7. What is the significance of external sorting algorithms in web based information retrieval system? Justify block-sort based indexing as external sorting algorithm.
8. Elaborate block storage and front-coding as dictionary compression methods.
9. Evaluate collaborative filtering with its advantages and limitations.
10. The PageRank algorithm uses a model of a “random surfer” to calculate the importance or validity of a page. Justify how the random surfer can be modelled and need of it using suitable example.
11. Draw and discuss bow-tie structure of the web.
12. Elaborate on Map-reduce programming paradigm with suitable example.
13. Elaborate on CPC, CPM and click spam with respect to advertising on web.
14. Evaluate the working of URL frontier for web crawler.
15. Describe Hubs and Authorities.
16. What is ranked Boolean retrieval? Consider the query “python” in a collection in which each document has three zones: abstract, title and body. Three weights g1=0.4 , g2=0.3 and g3=0.3, respectively corresponding to the abstract, title and body zones. Find document score for the query.
17. Given a document containing terms with the given frequencies A(5) ,B(10), C(8). Assume document collections 12,000 and document frequencies of these terms are A(60), B(1500), C(300). Compute TF-IDF.
18. Discuss pseudo relevance feedback and indirect relevance feedback.
19. Discuss precision at k, R-precision and break-even point with respect to ranked retrieval.
20. What is SPAM from the context of web search? Explain different ways used by spammers.