Introduction:

This week's focus was on various searching and indexing techniques in data structures, including algorithms for both unsorted and sorted arrays, hashing techniques, linear and tree-based indexing. I explored linear probing, open and closed hashing, bucket hashing, and studied search techniques like interpolation and jump search. The practical emphasis on ISAM, B-trees, and 2-3 trees provided deeper insight into how databases manage large datasets efficiently.

Difficulties Faced:

Understanding the nuances between different tree-based indexing methods, especially distinguishing B-trees, B+ trees, and 2-3 trees, was initially challenging. It required multiple reviews of visual aids and external tutorials. Implementing collision resolution methods in hashing, especially differentiating between open and closed hashing and managing bucket overflows, also proved complex. Debugging self-quiz questions involving hashing algorithms required extra effort to track index changes after collisions.

Activities Performed:

I engaged in various activities including completing the self-quiz and review quiz to prepare for the final exam. The quizzes helped reinforce theoretical knowledge and highlighted areas needing revision. I also participated in the discussion forum by posting a detailed explanation of the Zipf distribution. In the post, I explored its significance in indexing and its "rank-frequency" relationship, giving real-world applications such as SEO and data compression (Powers, 1998; Newman, 2005). The act of writing this post enhanced my understanding of how naturally occurring data sets often follow this distribution pattern.

Conclusion:

This week solidified my grasp on how search and indexing techniques work in real-world systems, especially in databases and file systems. I realized the practicality of these concepts, especially Zipf's law, which offers insight into how data is distributed and accessed in large systems. The structured approach to indexing and the logic behind various search strategies has broadened my analytical thinking as a data analyst.

References:

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46(5)*, 323–351. https://doi.org/10.1080/00107510500052444

Powers, D. M. W. (1998). Applications and explanations of Zipf's law. *In Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning*. https://dl.acm.org/doi/10.5555/1603899.1603924

Word count: 291