

Chapter No. 4 Index Construction (Article 4.1 to 4.4 are covered)

Chapter No. 5 Index Compression (Article 5.1 to 5.3 are covered)

<Food for Thoughts>

1. What are the problems with traditional inverted index approach introduced in chapter 1? Illustrate them.
2. How Block-Sort Based Indexing(BSBI) gain advantages over traditional inverted indexing approach?
3. What are some of the drawbacks on BSBI?
4. How would you create the dictionary in blocked sort-based indexing on the fly to avoid an extra pass through the data?
5. Explain how Single-Pass In Memory Indexing(SPIMI) gain advantages over BSBI?
6. Compare SPIMI and BSBI in term of time complexities.
7. Compare the differences between Block-Sort Based Indexing(BSBI) and Single-Pass In Memory Indexing(SPIMI)
8. State Heaps Law? Explain its importance in IR.
9. State Zipf's Law? Explain its importance in IR.
10. How Heap's Law and Zipf's Law are related?
11. Why compression is important for dictionary and posting list together?
12. Given a dataset for the IR? How would you estimate the dictionary size and posting size? If it is known that it is a collection of English short stories, how would you justify the estimates?