

Chapter No. 3 Dictionaries and Tolerant Retrieval

<Food for Thoughts>

1. What are the two data structures choices while implementing dictionaries? which one you preferred and why?
2. What do we mean by "Wild Card Queries"? what are their different types? give examples.
3. What is a Permuterm Index? How it can support the general Wild Card Queries?
4. What is a k-gram index? How it can support the general Wild Card Queries?
5. When Soundex based search is required? give example?
6. Write down the entries in the permuterm index dictionary that are generated by the terms { "mama", "chacha", "Tullo" }
7. If you wanted to search for s*ng in a permuterm wildcard index, what key(s) would one do the lookup on?
8. Give an example of a sentence that falsely matches the wildcard query mon*h if the search were to simply use a conjunction of bigrams.
9. What is so special about context sensitive spelling correction?
10. Why do you think soundex approach is not good for IR?