Name : Dhiraj Bodake
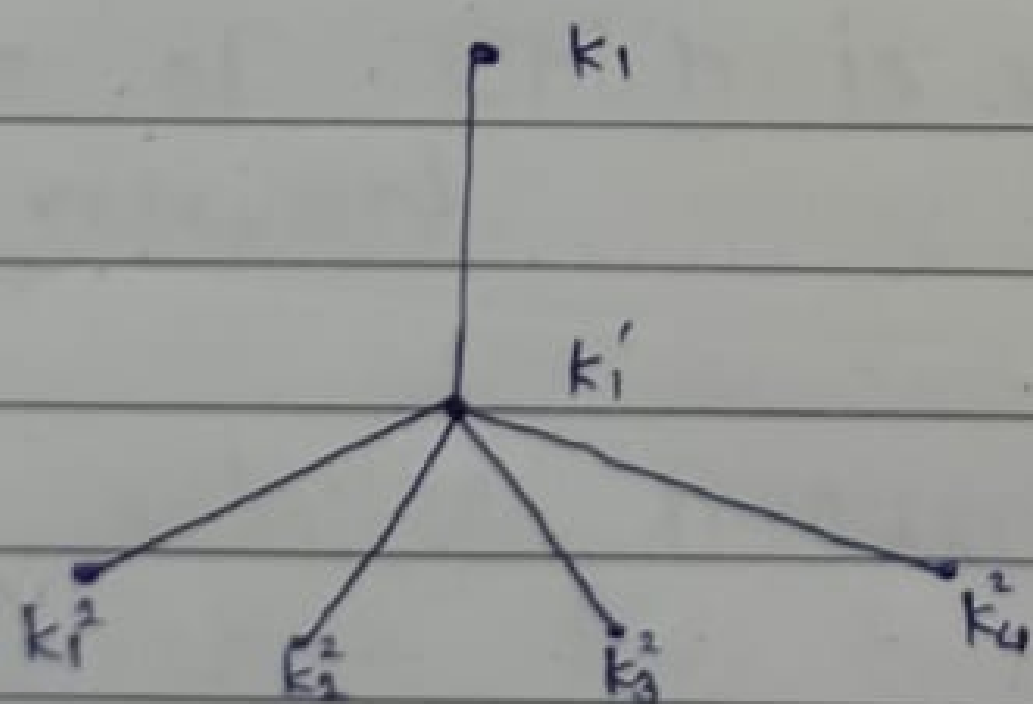
Roll No : 18141216

TUTORIAL NO : 04

Q.1 Explain Boolean search & serial search an search strategies.

- Boolean Search : It retreieven those documents which are true for query. Queries are expressed in term of index & combined by usual logical connections AND, OR, NOT.

- System which operate by boolean search allow even user to narrow / broden search by giving user access to structured dictionary which stores keywords related to any given of keyword.

ex.

- way to implement this search is through inverted file. we store list of keywords in vacabulary of in each list put addresses of documents containing that particular particular words.

There are two ways:-
① matching function is given threshold, retrieving docs above threshold & discard below one
$$\{ D_i \mid M(Q, D_i) \geq T \}$$
② docs are ranked in increasing order of matching function value.

Q.2] What is pattern matching? Explain it with the help of different types of patterns.

→ Pattern matching is checking & locating specific sequences of data of some pattern among raw data/sequence of tokens.
The most used types of patterns are :-
① Words: A sequence of characters which must be a word in the text. This is the most basic pattern.

② Prefixes: A string which must form the beginning of text word. For instance, given prefix 'comput' all documents containing words like 'computer', 'computation' etc are retrieved.

③ Suffixes: A string which must form the termination of text word.
ex: suffix 'ters' all docs containing words like 'testers', 'computers' are retrieved.

④ Substrings : A string which can appear within a text word.

ex. for substring 'tal' all docs containing words such as 'talk', 'metallic', etc. are retrieved.

⑤ Ranges : A pair of string which matches any word lying between them in lexographical order.

ex: range between words 'held' & 'hold', will retrieve strings such as 'hoax' & 'hissing'.

Q.3. Write short note on cluster file.

→ - When two or more records are stored in the same file, it is known as clusters.

~ A file which is processed by clustering is called clustered file & the resulting structure as file structure.

- Cluster methods are more profitably discussed at the level of abstraction at which relations are discussed in connection with data bases that is in thoroughly data independent way.

- Selecting any appropriate cluster method & implementing it are two seperate problems

- Unfortunately not all users of clustering techniques see it this way & so current scene is rather confused.

- Those who see clustering merely to increase retrieval efficiency will tend to see a classification structure as file structure.
- While those who see clustering as means of discovering some inherent structure in data will look upon same structure as description of data.