

## **Solution – 1**

**The System follow the given below steps to search the relevant entries for the query entered by the user.**

- a) System takes the user query as the input.**
- b) Now system will parse the query and fetches the important keywords on which searching can be done.**
- c) After fetching relevant words, the prioritization of words has to be done on the basis of their accessibility and popularity.**
- d) After prioritization, the words will be grouped into two or three to get some combination, which can be searched, on the network.**
- e) Once done with the grouping, system searches the words and combination of words on the network using Comment Collector.**
- f) Comment Collector (CC) searches heavily used entries and follows their thread of comments to find new entries.**
- g) Multiple Comment Collector searched the network on a rate of 500 entries per second.**
- h) Comment Collector (CC) index each entry they find according to the words that are used in the entries**
- i) The result are rank according to relevance, the CC will note where the words are found (with more importance given to initial and ending of the entries) and the number of times the word is used in the entry.**

**1) The System differs from search engine because in the search engine web pages are being crawled with the help of web crawler and, here I used Comment Collector to crawl through the entries.**

**2) Ranking is very important for the system to decide the order of entries that to be placed in front of the user.**

**3) The way in which the keywords are combine can easily compromise the security of the system because the combing words like union, etc can easily be bypass by the hackers.**

**4) Yes, Spamming of wrong entries results into wrong piece of information to the user. This can be easily be detect using Near-Duplicate Detection where we use some threshold value for some similarity measure between pairs of document.**

**5) System ensure a good user experience by giving the searching result in the well sorted manner from highly visited to low visited.**

