UNIT – IV

TOPIC: User Search Techniques

USER SEARCH TECHNIQUES

Search Statements and Binding

- Search statements o are the statements of an information need generated by users to specify the concepts they are trying to locate in items.

- uses traditional boolean logic and/or natural language. o in generation of the search statement, the user may have the ability to weight to different concepts .

Binding :

- Binding is to the vocabulary and past experiences of the user.
- A more abstract form is redefined into a more specific form is binding.

- The search statement is the user's attempt to specify the conditions needed to logically subset the total item space to cluster of items that contains the information needed by the user.

- **Similarity Measures and Ranking**
- Searching is concerned with calculating the similarity between a user's search statement and the items in the database.
- The similarity may be applied to the total item or constrained to logical passages in the item.
- For example, every paragraph may be defined as a passage or every 100 words.
- The highest similarity for any of the passages is used as the similarity measure for the item.

- **Relevance Feedback** :
- Major problems in finding relevant items o It lies in the difference in vocabulary between the authors and the user.
- Thesauri and semantic networks provide utility in expanding a user's search statement to include potential related search terms.
- But this still does not correlate to the vocabulary used by the authors that contributes to a particular database.

- **Selective Dissemination of Information Search:**
- Search system o a search system is called a "pull" system.
- In a search system, the user proactively makes a decision that he needs information.
- Directs the query to the information system to search.
- In the search system, an existing database exists. Corpora statistics exist on term frequency within and between terms and these can be used for weighting factors in the indexing process and the similarity comparison.

- **Weighted Searches of Boolean Systems**
- The two major approaches for generating queries are Boolean and natural language.
- Natural language o Queries are easily represented within statistical models and are usable by the similarity measures.
- **Boolean queries**
- Issues that arise when Boolean queries are associated with weighted index systems are
- how the logic (AND, OR, NOT) operators function with weighted values. how weights are associated with the query terms.

- **Searching the INTERNET and Hypertext**

- The primary techniques for search of items are associated with servers on the Internet that create indexes of items on the Internet and allow search of them.
- Some of the most commonly used nodes are
- YAHOO
- AltaVista
- Lycos
- In those systems, there are active processes that visit a large number of Internet sites and retrieve textual data which they index.

- Lycos and AltaVista automatically go out to
- other Internet sites and return the text at the sites for automatic indexing.
- Lycos returns home pages from each site for automatic indexing.
- Altavista indexes all of the text at a site.
- The retrieved text is used to create an index to the source items storing the Universal Resource Locator (URL).
- All the systems use some form of ranking algorithm to assist in display the retrieved items.

- **Introduction to Information Visualization**

- System designers need to treat the display of data as visual computing instead of treating the monitor as a replica of paper.
- Functions that are available with electronic display and visualization of data that were not previously provided are:
- modify representations of data and information or the display condition(e.g., changing color scales)
- use the same representation while showing changes in data (e.g., moving between clusters of items showing new linkages)

Cognition and Perception:

- **Visualization**
- The transformation of information into a visual form that enables the user to observe and understand the information.
- The mind follows a set of rules to combine the input stimuli to a mental representation that differs from the sum of the individual inputs:
- Proximity - nearby figures are grouped together
- Similarity - similar figures are grouped together
- Continuity - figures are interpreted as smooth continuous patterns rather than discontinuous concatenations of shapes .

Information Visualization Technologies:

Information visualization in Information Retrieval Systems considers how best to display

1. results of searches
2. structured data from DBMSs
3. results of link analysis correlating data.

The goals for displaying the result from searches fall into two major classes:

1.document clustering

2.Search statement analysis

The goal of document clustering

1. to present the user with a visual representation of the document space constrained by the search criteria.
2. Within this constrained space there exist clusters of documents defined by the document content.