

Name: Dhiraj Bodake  
 Roll No: 18141216

## TUTORIAL NO: 09

Q.1 Explain Query processing in detail with respect to distributed IR.  
 → Query processing in a distributed IR system proceed as follows:

- (i) Select collections to search
- (ii) Broadcast query to selected collections.
- (iii) Evaluate query at distributed collections in parallel.
- (iv) Combine result from distributed collections into final result.

- The step may be eliminated if query is always broadcast to every document collection in system.

- Each all the participating search servers then evaluates the query on the selected collections using its own local search algorithm. Finally result are merged.

- Improvement on this process is to merge hit lists based on relevant score.

- IF documents are randomly distributed such global term statistics are consistent across all of distributed collections.

Explain in brief : Collection partitioning in distributed IR

When the distributed system is centrally administered more options are available. The first option is simple replication of the collection across all of the search servers. This is appropriate when collection is small enough to fit on a single search server, but high availability and query processing throughput is handled one of two ways.

- In the first method, each search server separately indexes its replica of the documents.
- In the second method, each server is assigned a mutually exclusive subset of documents to index and index subsets are replicated across search servers.

The second option is random distribution of documents. This is appropriate when large document collection must be distributed for performance reasons but the docs will always be viewed and searched as if they are part of a single logical collection. The final option is explicit semantic partitioning of the documents. Here documents are either already organized into semantically meaningful collections such as by technical discipline or an automatic clustering procedure is used.

a.3 Explain MULTUS data model as the data modelling technique in multimodal IR.

→ MULTUS is a multimodal document server with advanced document retrieval capabilities, developed in context of an ESPRIT project in the area of office systems.

- MULTUS is based on client-server architecture. Three different types of document servers are supported. current servers, dynamic servers, and archive servers, which differ in storage.
- The MULTUS data model allows representation of high level concepts present in the documents contained in the database, the grouping of documents into classes & documents having similar content and structure.
- Each document is described by a logical structure, a layout structure and a conceptual structure.
- The logical structure determines arrangement of logical document components.
- The layout structure deals with layout of the document.
- The conceptual structure allows a semantic oriented description of the document content as opposed to description of the document content provided by logical and layout structure.