Chapter 7

Conclusion and Future Work

7.1. SUMMARY

The work focuses on designing P2P catalog system to extract cluster and node information in order to initiate the search process and retrieve documents efficiently. The documents distributed over P2P network is conceptually clustered and characterized by cluster keys to represent the dimension of the cluster. Every peer will carry $n$ number of dimensions. Where, these dimensions cannot be mapped directly on to peer nodes due to curse of dimensionality. Therefore, the clusters are mapped on to logical nodes in DHT based chord ring. On initiating a query from an arbitrary peer, the peer node consults its own finger table to discover similar clusters available on neighboring peers. The cluster indices are shipped to get the reduct from neighbors. The peers containing relevant content are the candidate peers. The indices corresponding to the query are shipped to candidate peers and semi join is performed to retrieve the content in response to initiated query.

7.2. CONTRIBUTIONS

The contributions from our work include:

a. P2P Coordination: Peer Nodes are designed to coordinate in order to extract information required like cluster location and peer node location semantically.

b. Document movement: The document movement is minimized and only indices of clusters are shipped.

c. Semantic Clustering of documents: Documents are clustered according to semantics. Therefore, even a document of unknown class can also be clustered under the proper class.

d. Cluster aware and node aware search initiation: The search for content is not initiated at random point. Instead the search starts with the node where the relevant content is available based on semantic distribution of clusters.
e. Content directed search: The peer nodes containing relevant content are gathered first. (Information gathering). Then the search query is communicated only to those peers on the virtual ring. This reduces communication overhead and increases the response time.

f. Minimum resource consumption: The number of nodes visited indicates the resource consumption in the network.

g. Adaptive Indexing: The indices available are different peers are updated according to changing configuration of P2P overlay network incorporating extension of cooperative mirroring scheme.

7.3. FUTURE WORK

Our current work confines to only clustering documents based on generic hierarchies. It can be further extended to specific domain hierarchies.

The P2P catalog system is confined to semantically cluster, generate indices, organize indices for documents distributed in structured P2P system and does not consider network related issues. The reliability of communication, network delay and quality of service are beyond the scope of current research work.

Though the indices are aimed to be distributed evenly, with the increasing cluster keys threshold leads to uneven load on peer nodes. It may be considered in our future work.

Replication of indices are required if the configuration is highly dynamic, which is not considered significantly. It may be extended in our future work.