Information at World Wide Web (WWW) is retrieved through queries that are submitted by the users to the search engine. For a given query, various techniques are followed to access the WWW to process it and retrieve the most related documents. All techniques retrieve more or less same documents, but vary in time that it consumes to generate such results. Therefore, the best method is needed to reduce the time required for information retrieval. This requirement is focused by enormous researchers to discover and invent many techniques. Though the invented techniques retrieved the documents quickly, they did not concentrate on the relevancy of retrieved documents. Hence, an essential technique that depletes less time and retrieves more accurate document is necessary. To address this issue, a framework named Semantic and Feature Aggregated Information Retrieval (SFAIR) is proposed in this dissertation. The proposed framework comprises the following components, namely (1) clustering, (2) indexing, (3) information retrieval and (4) ranking. Each component is described in four different phases.

At the first phase, the documents are indexed through a technique named multilevel hashing that helps to find the location of information easily. This indexing technique properly maintains the web documents in two levels: a database named BulidKeyDB is built as the first and keywords are indexed in the second level. In second phase, the documents are clustered, which is another way for increasing the speed of the searching operation. Clustering is
taken through context-based helps to cluster the documents depending on the meaning of the information. The clustering is carried out in three levels, namely sentence level, document level and at corpus level. The first two phases help to improve the retrieval speed, whereas in the third and fourth phases two different retrieval techniques are developed along with ranking methodology. The third phase retrieval approach is supported by ontology-semantic and synonym determination. It highly depends on the semantic structure of the information. Ranking approach for displaying to the user also follows on the semantic nature. Document weight is computed for ranking the document. Similarly, feature-based information retrieval method is proposed in the fourth phase, and it ranks the documents or web pages depending on the score that is computed for all documents. These individual components are grouped to frame the fifth phase. In phase five, a framework named SFAIR is framed and all the four phases are combined. SFAIR pre-processes the documents and provides those pre-processed documents as inputs to the clustering. The documents are clustered as in phase one and each cluster is indexed using the multilevel hashing. These components act as supporting method for information retrieval technique. In phase five, feature-based information approach as followed in phase four is used, and for ranking the documents, the weight that is implemented in phase three is used. The SFAIR framework is analyzed through various experiments, and the result of the experiment shows that it outperforms the existing techniques.