XML has become the most common means for publication, storage and exchange of data over the Internet. As a result, a huge amount of information is stored and represented in XML, and research on keyword search in XML documents is on the increase as it allows users to find information they are interested in without having to know the underlying database schema or complex query language. A lot of research has been conducted in XML keyword search. The main problem of XML keyword search is the identification of user search intention of an XML keyword query precisely in the presence of keyword ambiguities. In XML keyword search, the accurate identification of user search intention and grading of the result in the presence of keyword ambiguities have been challenging problems. It is quite clear that the most essential part in keyword search is effectiveness in terms of result relevance.

In this thesis, we propose a Dual Indexing and Mutual Summation based keyword search Algorithm for XML databases, XDMA, to address the problems in XML keyword search approaches, specifically, XReal. The first step in our proposed approach is to build dual indices, namely, tag information table and data node information table for structural node and data node in XML database respectively. In the second step, we propose a keyword searching technique to select all possible T-typed nodes for a given query using the two-level matching between the two indices. Using this searching technique, each keyword in a given query can be identified and distinguished as tag or data while searching for an input query. Furthermore, another new keyword ambiguity is identified in this thesis. Subsequently, by incorporating the concept of dependence of two indices and the concept of mutual summation, we define the Mutual Score (MScore) between selected tags and query keywords to find the desired T-typed node. Successively, we introduce the similarity function for random variable and utilized it to propose a formula to estimate similarity among the leaf nodes of XML data and the query so as to determine the precise data via the
selected $T$-typed node. In addition, we propose the grading for the query results having comparable relevance scores. Finally, comprehensive experiments have been conducted to demonstrate the effectiveness of our proposed approach XDMA using the different XML datasets and the performance of our proposed approach with other XML keyword search approaches has been compared.

Our proposed dual indexing approach, XDMA, builds dual indices that deal with each structural node and data node separately in XML database so as to simplify the query processing. Moreover, the data node information table is dependent on the tag information table in relation with the tag name. Hence, mutual score which incorporates the concept of mutual summation among tag and data keywords is applied to XML keyword search. Mutual Score between selected tags and query keywords is designed to enable the selection of desired $T$-typed node for the input query. Subsequently, similarity score computes similarity among leaf nodes of XML data and the query to determine the precise data through the selected $T$-typed node. It involves the mutual score and similarity function of keyword matching data value contained in a leaf tag. Based on similarity score, grading determines the most eminent similarity score for every prefix path with respect to its keyword. In this thesis, XDMA addresses the limitations and problems in XReal.