CHAPTER 7

CONCLUSION AND FUTURE WORK

Most of the business in the internet use XML format to represent and exchange data efficiently. The semi-structured nature of XML allows the user to update the document in a flexible manner. The query processing systems use the labeling method to get answer for the queries. The operations such as insertion, deletion and updation of nodes affect the lexicographic order of the nodes. When such order-sensitive updates take place in XML documents, the existing labeling schemes need to perform the re-computation of the labels of the affected nodes.

Research has been made on this issue, for which three approaches have been proposed. In first approach, NLSX, the labels are made with the combinations of upper case letters (A-Z), lower case letters (a-z) and digits (0-9). The generated labels are in lexicographic order, which is not changed even when changes occur in the document. This is helpful for getting the relationships like parent-child and ancestor-descendant relationships. Also, it is useful in query processing system for getting quick response.

To improve the performance further, the thesis proposed the second approach, NLSXU, in which the labels are generated by using the combinations of digits (0-9), upper case letters (A-Z), lower case letters (a-z), and some of the characters in Unicode character set. The lexicographic order of the labels is useful in maintaining the relationships and improving the query processing system.
For further enhancement, the third approach, BSIX has been proposed. The binary sequences are used to generate the labels. The synthetic datasets and real world dataset have been used for experimental purposes. The lexicographic order of label of the siblings is maintained always and also, it preserves the relationship among the nodes. Moreover, the deleted labels can also be reused. The experiment results show that the BSIX scheme reduced the space and time requirements by an average of 73% and 48% respectively when compared with present systems. It also reduced the time taken for query processing by an average of 34%.

In future, a new strengthened, powerful schema may be developed by which the re-computation problem may be solved without sacrificing the nature of XML. That is, without losing the flexibility of XML, a new way of defining schema may be designed. From the schema information, the clusters of same structural data may be formed. Additionally, ranking the nodes within the clusters may further reduce the query processing time. The clusters of same type of structural data may reduce the overhead of handling large size index. Then, anyone of the powerful compressing algorithms may be applied for further reducing the size of the label.