COXML: A COOPERATIVE XML QUERY ANSWERING SYSTEM

Shaorong Liu and Wesley W. Chu
UCLA Computer Science Department

Introduction:

XML schema is often too complex for users to specify the exact query conditions. Thus, we develop an XML approximate query answering system that cooperates with users (*CoXML)* [1].

Methods:

The key features of the CoXML system include: 1) a new query language that allows users to specify approximate conditions and to control the relaxation process; 2) relaxation indexes to provide systematic guidance to query relaxation; and 3) a new similarity metric that use both content and structure aspects in measuring the relevancy of the returned approximate answers.

When a user posts a query, the *Relaxation Engine* first sends the query to the *XML Database Engine* to search for exact matching answers. If there are no answers or insufficient results, the *Relaxation Engine* consults the *Relaxation Indexes* for the best relaxed query based on the relaxation specifications in the query. The relaxed query is then resubmitted to the *XML Database Engine* to search for approximate answers. The *Ranking Module* ranks the returned approximate answers based on their structure and content relaxations. This process is repeated until either there are enough approximate answers returned or the query is no long relaxable.

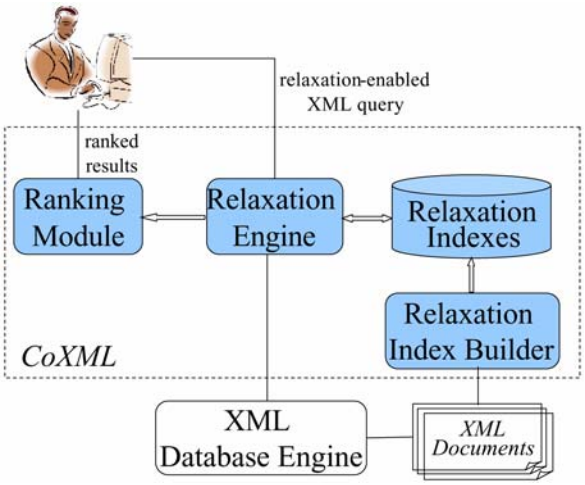

Figure 1: The CoXML system architecture

Results:

We used the INEX [2] 05 benchmark test collection for evaluating the effectiveness of approximate query answering generated by CoXML. The results reveal that the relaxation control feature provided by our query language is very effective in searching for user-specific approximate answers. Further, our similarity metric including both content and structure aspects yields better relevancy than that of using content only as used by other systems in INEX 05.

Conclusions:

We developed a CoXML system for approximate query answering by relaxing query conditions based on users' relaxation specifications and guided by relaxation indexes. CoXML ranks the returned approximate answers based on their corresponding levels of content and structure relaxation. Empirical evaluations validate the effectiveness of the proposed methodologies.

Reference

[1] S. Liu and W. W. Chu. CoXML: A Cooperative XML Query Answering System. Submitted to VLDB, 2006
[2] INEX. http://inex.is.informatik.uni-duisburg.de/.