

## Effective predicate identification algorithm for XML retrieval

### ABSTRACT

Query structuring systems are keyword search systems recently used for effective retrieval of XML documents. Existing systems fail to put keyword query ambiguity problems into consideration during query preprocessing. Thus, the systems return irrelevant user search intentions. A search intention consists of entity nodes and predicate nodes of XML data. In this paper, an entity based query segmentation (EBQS) method which interprets a user query as a list of keywords and/or named entities to resolve ambiguity. Then, segment terms proximity scorer (STPS) that assigns relevance scores to XML fragments that contains query keywords is proposed. Fragments containing the keywords as interpreted by EBQS are assigned higher scores. Finally, an effective predicate identification algorithm (EPIA) which uses EBQS and STPS to return relevant predicates is introduced. The effectiveness of the algorithm is demonstrated through experimental performance study on some real world XML documents.

**Keyword:** Query structuring; XML retrieval; Predicate node; Keyword query