

## **Relationships in structured text retrieval**

Mounia Lalmas, Department of Computer Science, Queen Mary, University of London

### **SYNONYM**

None

### **DEFINITION**

In structured text retrieval, the relationship between text components may be used in ranking components relative to a given query.

### **MAIN TEXT**

In a structured text document, there exists a relationship between the document components. In the context of XML retrieval, the relationships between elements are provided by the logical structure of the XML mark-up. An element, unless it is the root element (the document itself), has a parent element, which itself may have a parent element. Similarly, non-leaf elements have children elements, and so on. Considering relationships between elements appears to be beneficial for XML retrieval. For instance, in a collection of scientific articles, it is reasonable to assume that the “abstract” of an article is a better indicator of what the article is about than a “future work” section in the same article. The challenge in XML retrieval is what types of relationship should be considered, and how this information can be used to score elements according to how relevant they are for a given query. In the contextualization approach [1], considering the “root element – element” relationship to rank an element in addition to the element own content has shown to improve retrieval effectiveness [2,3].

### **CROSS REFERENCES**

XML Retrieval  
Logical structure  
Contextualization  
Structure weight

### **RECOMMENDED READING**

[1] Arvola, P., Junkkari, M., & Kekäläinen, J. (2005). Generalized contextualization method for XML information retrieval. In *Proceedings of the ACM Fourteenth Conference on Information and Knowledge Management (CIKM)*, 20-27.

[2] Mass, Y., & Mandelbrod, M. (2005). Component ranking and automatic query refinement for XML retrieval. In *Advances in XML Information Retrieval and Evaluation*, Lecture Notes in Computer Science 3493, 73-84.

[3] Sigurbjörnsson, B., Kamps J., & de Rijke, M. (2003). An element-based approach to XML retrieval. In *INEX 2003 Workshop Proceedings*, 19-26. Available at: <http://inex.is.informatik.uni-duisburg.de:2003/proceedings.pdf>.