INEX: Initiative for the Evaluation of XML Retrieval summarize

Carlos Andrés Reyes Evangelista - 157068

The paper is written around the inherent relevance of the XML markup language, the alternatives in techniques for retrieving information from an XML document and the importance of obtaining a trustworthy testbed of XML documents to being able to test XML retrieval systems. It is in this last point where INEX performs its appearance by surging as an alternative to generate a reliable document collection to be used for testing purposes.

The ***document-centric view*** is used when the document is structured in a paradigmatic form, i.e., the constructions within the document follow logic rules that concern the own nature of the document itself. The paper provides a few examples such as hierarchical chains: *chapter -> section -> paragraph*. Whereas ***data-centric view*** is more likely to be used for documents that will be exchanged among several nodes and that is required for them the existence of a standard way to structure and de-structure one specific interchangeable document. ***Content-and-structure*** queries try to limit the results of a query to a set of specific structures either for obtaining more concise results or for restricting them to follow a required structure. On the other hand, ***content-only*** queries do not require the retrieval system to look for the structure. The importance of these kind of queries lies in the fact that users often either do not know the exact structure of the XML document or do not desire to limit their results to a specific field.

`DTD` stands for Document Type Definition. It is used to define a standardized set of valid building blocks to be used in a specific kind of XML documents. Within the paper it is demonstrated that the DTD -called topic DTD for their business- plays an important role as the one in charge of defining how the documents and the queries within the INEX should be structured to provide a better achieve of relevant retrievals.