

Structural Features in XML Retrieval

**Speaker: Georgina Ramirez,
CWI, Nederlands**

Abstract:

The structure of documents provides a new source of information that retrieval systems may exploit to improve their search effectiveness. This seminar summarizes the main ideas and results of our research, which investigates the use of several structural features -- markup and (derived) metadata -- for effective XML retrieval. Our retrieval framework is based on the principle of polyrepresentation (Ingwersen, 1994) and makes use of the available evidence collected from documents, queries, and contextual features to rank components of XML documents. We will present our approaches on three main topics: (1) new retrieval strategies that use structural information, (2) the use of relevance feedback techniques to refine the structural information given a user need, and (3) the study of the relationships between user search tasks and contextual factors and the structural characteristics of the relevant information. We evaluate these approaches using the INEX benchmark and show that structural information can be further exploited to improve retrieval effectiveness.