Introduction

XML is an ideal model for structure and text. XML structural queries are augmented with keyword search to query both structured data and textual content. We study the ranking of XML tree pattern queries with text predicates. While query efficiency has received a lot of attention and many systems and algorithms have been developed, ranking functions (e.g., TF-IDF) are borrowed from IR and adapted to structural queries. The foundation of these adaptations is to treat all textual elements as a “document” collection, so that elements that match one or more keywords can be assigned to relevance scores. The element scores are then aggregated to the ranking score of the result fragment.

In this paper, we argue that such an adaptation is questionable and violate IR heuristics. Though various IR ranking functions evolved separately in the history, they are all closely related to a small set of IR heuristics. In particular, Fang et al. [6] formally define a set of basic constraints that any reasonable ranking functions should satisfy. We argue that ranking functions of tree pattern queries should be derived from these constraints, so that they are consistent with IR heuristics.

I count on the abstract to have told what we do.