Extended XQuery for Digital Libraries*

Alex Dekhtyar Ionut E. Iacob Kevin Kiernan Dorothy C. Porter

University of Kentucky
Lexington, KY 40506

{dekhtyar,eiaco0}@cs.uky.edu, kiernan@uky.edu,dporter@uky.edu

ABSTRACT

Documents have, in general, a multihierarchical structure (such as physical organization in the form of pages and lines, content organization in the form of paragraphs and sentences, etc.). Searching multihierarchical XML encoding presents a number of unique challenges for both computer scientists and document experts. We present an extension of the XQuery language suitable for searching multihierarchical XML.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Query formulation

General Terms

Languages, Management, Design

Keywords

XQuery, document-centric XML

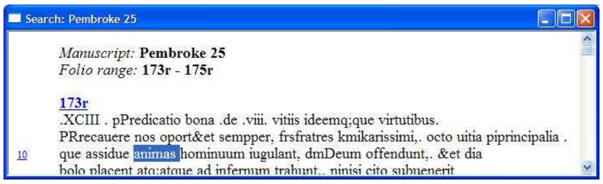
1. INTRODUCTION

In our previous work [1] we described the Edition Production Technology which was recently enhanced towards an Edition Production and Presentation Technology (EPPT[1]) – a complex platform for preparing and presenting electronic editions of documents. This platform includes a data model for multihierarchical document-centric XML, an extensive suite of editing tools, and a query language, Extended XPath [3, 2], for searching multihierarchical documents. Extended XPath, an extension of the XPath query language, offers us the benefit of clear semantics for path expressions over multihierarchical XML. At the same time, to achieve full query language functionality, path expression languages need to be "wrapped" by a declarative query language, such as XQuery. XQuery is an attractive XML query language for document-centric XML processing as it provides built-in support for presentation of query results. EPPT has HTML presentation capabilities and we have enhanced it with an extended version of XQuery FLWOR expressions over multihierarchical XML documents. In our demonstration we show the benefits of using XQuery as the underlying query language at the electronic edition preparation, presentation, and delivery time.

*The work of the first author was supported, in part, by the NSF grant ITR-0219924 and the NSF grant ITR-0325063. The work of the second and third authors is supported in part by a Collaborative Research Award from the NEH (RZ-20887-02) and the Andrew W. Mellon Foundation.

[1]Available at: http://beowulf.engl.uky.edu/∼eft/

JCDL'06, June 11–15, 2006, Chapel Hill, North Carolina, USA.
ACM 1-59593-354-9/06/0006.




```
let $a := doc("data.xml")//overlapping::*
for $content in doc("data.xml")//text()
  return
    if ($content/parent::* intersect $a)
    then
      <font style="background-color: blue"> {$content}
      </font>
    else $content
```

Figure 1: Extended XQuery and html formatted results.

2. EXTENDED XQUERY

We extended the XQuery language to support queries over multihierarchical documents. Two main extensions were incorporated: extended axes (from Extended XPath) and sequences of items from multiple XML hierarchies. For instance, in extended XQuery we can search for overlapping markup and highlight the results in a HTML presentation, as shown in the query in Figure 1. In this example we are using an Extended XPath `overlapping` axis to select parts of overlapping markup, which are then displayed on blue background.

3. DEMO OVERVIEW

We present the Edition Production and Presentation Technology (EPPT), enhanced with the implementation of FLWOR expressions of XQuery over multihierarchical XML documents. We demonstrate the extended XQuery ability to search multihierarchical documents and applications of its presentation features. The implementation of extended XQuery extends the ability of EPPT to show HTML-formatted search results (Figure 1).

4. REFERENCES

[1] A. Dekhtyar, I. E. Iacob, J. Jaromczyk, K. Kiernan, N. Moore, and D. C. Porter. Building image-based electronic editions using the Edition Production Technology. In *JCDL '05*, pages 406–406, New York, NY, USA, 2005. ACM Press.

[2] I. E. Iacob and A. Dekhtyar. Processing xml documents with overlapping hierarchies. In *JCDL '05*, pages 409–409, New York, NY, USA, 2005. ACM Press.

[3] I. E. Iacob and A. Dekhtyar. Towards a Query Language for Multihierarchical XML: Revisiting XPath. In *In Proc. of WebDB*, pages 49–54, 2005.