

Grouping in MetaXQuery

Hao Jin and Curtis Dyreson
School of E.E. and Computer Science
Washington State University
USA

WISE 2004 - Brisbane

Outline

- Motivation
- Data grouping
- Metadata grouping
- Summary

Grouping in MetaXQuery - Jin and Dyreson

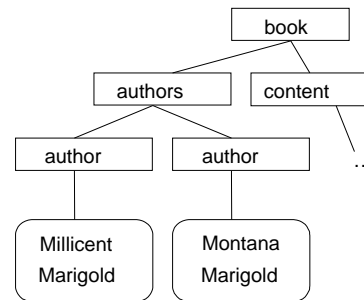
Book Data in XML

```
<book>
  <authors>
    <author>Millicent Marigold</author>
    <author>Montana Marigold</author>
  </authors>
  <content>
    <chapter>Usability testing is...</chapter>
    ...
  </content>
</book>
```

Grouping in MetaXQuery - Jin and Dyreson

Parsed into DOM

Query: List authors and for each author, the books they've written.
Solution: Group books by author.



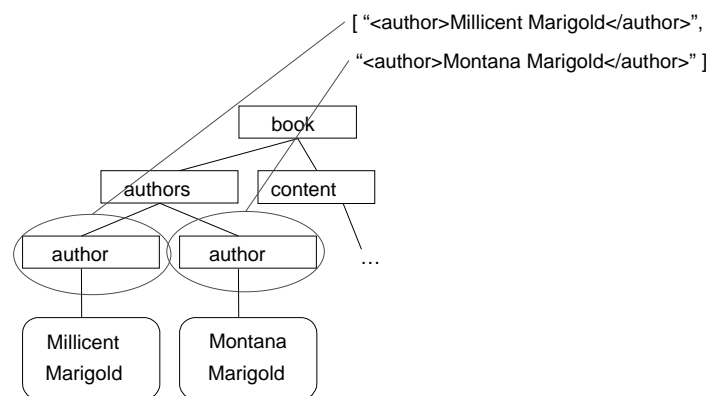
Grouping in MetaXQuery - Jin and Dyreson

Grouping in XQuery

```
for each $author in distinct-values(//author)
return
  <author> $author
  {for each $book in //book
   where $book/authors/author = $author
   return $book
  }
</author>
```

Grouping in MetaXQuery - Jin and Dyreson

Distinct Values



Grouping in MetaXQuery - Jin and Dyreson

Desirable Grouping in XQuery

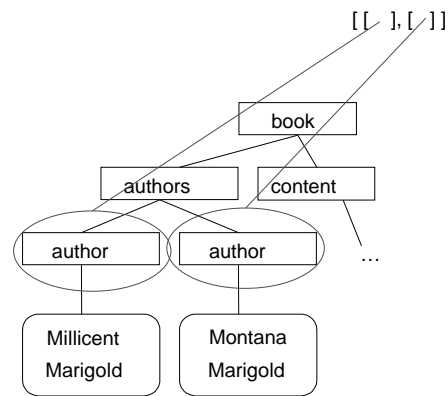
```

for each $author in data-group(//author)
return
  <author> $author {
    for each $book in $author/../../book
    return <book> $book </book>
  }
</author>

```

Grouping in MetaXQuery - Jin and Dyreson

Data-Group Builds a List of List of Nodes



Grouping in MetaXQuery - Jin and Dyreson

Metadata

- Metadata is data about data
- Descriptive
 - author, language, subject (Dublin Core), privacy
- Proscriptive
 - security, time

Grouping in MetaXQuery - Jin and Dyreson

Embedding Metadata in XML

```

<book>
  <metadata>
    <subjects xml:lang="en">
      <subject>Usability testing</subject>
    </subjects>
    ...
  </metadata>
  <content>
    <chapter>Usability testing is...</chapter>
    ...
  </content>
</book>

```

Grouping in MetaXQuery - Jin and Dyreson

Using RDF

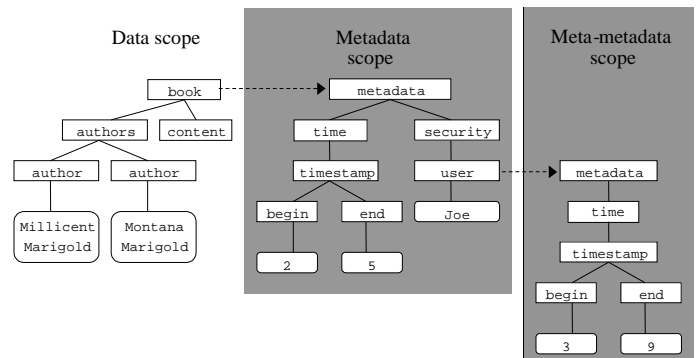
```

<RDF xmlns="rdf-syntax-ns#">
  <Description
    about="document(books.xml)//book[1]"
    xml:lang="en">
    <subject>Usability testing</subject>
  </Description>
  ...
</RDF>

```

Grouping in MetaXQuery - Jin and Dyreson

MetaDOM



Grouping in MetaXQuery - Jin and Dyreson

MetaDOM Features

- Support arbitrary levels of metadata
- Reuses DOM at each level
- Separates metadata and data into different scopes
 - Data-only
 - Group books by author
 - Metadata-only
 - List the subjects
 - Data and metadata conditions
 - Group books by time
- Upwards compatible with DOM/XPath/XQuery/XSLT etc.

Grouping in MetaXQuery - Jin and Dyreson

Grouping with Metadata

- Evaluate data group conditions
 - Semantics differs for each kind of metadata
 - Language – equality
 - Time – intersection
- Recursively evaluate each level of metadata

Grouping in MetaXQuery - Jin and Dyreson

Desirable Grouping in MetaXQuery

```
for each $author in categorize(//author)
return
  <author> $author {
    for each $book in get-parent($author)/..
    return <book> $book </book>
  }
</author>
```

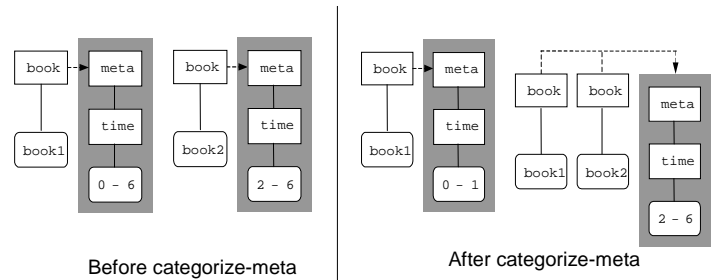
Grouping in MetaXQuery - Jin and Dyreson

Reorganizing the Data in MetaXQuery

Query: Group books by time

Solution: for each \$book in categorize-meta (//book, "time")

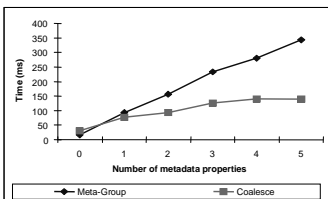
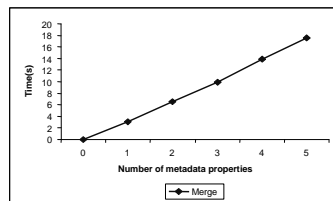
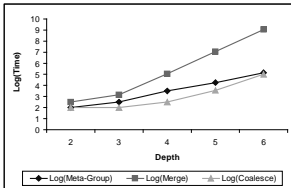
return \$book



Grouping in MetaXQuery - Jin and Dyreson

Prototype Implementation and Experiments

- Overhead is linear in amount of metadata



Grouping in MetaXQuery - Jin and Dyreson

Conclusions

- Extend XQuery to support metadata
 - Include grouping semantics for metadata
- Categorize-meta to reorganize data by metadata
- Performance results suggest overhead is linear in amount of metadata

Grouping in MetaXQuery - Jin and Dyreson

Thanks!