fast Faceted Search in XML

Anne Schuth & Maarten Marx
University of Amsterdam
Faceted Search in XML

Anne Schuth & Maarten Marx
University of Amsterdam
structural Faceted Search in XML

Anne Schuth & Maarten Marx
University of Amsterdam
Outline

- **Motivation**
- **Fast** Faceted Search in XML
  - Implementation in eXist
  - Benchmarked
- **Structural** Faceted Search in XML
  - Work in progress
Motivation
Why Faceted Search?
Motivation

Why Faceted Search?

• Lots of data
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
  • Why am I looking into (faceted) search?
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
  • Why am I looking into (faceted) search?

• In XML
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
  • Why am I looking into (faceted) search?

• In XML
  • Why XML?
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
  • Why am I looking into (faceted) search?

• In XML
  • Why XML?

• With metadata
Motivation

Why Faceted Search?

• Lots of data
  • Millions of political documents
  • Why am I looking into (faceted) search?
• In XML
  • Why XML?
• With metadata
  • How can we exploit metadata?
We are not alone (1) well..
We are not alone (1) well..
We are not alone (2) well.
We are not alone (2) well..
We are not alone (3) well..
We are not alone (3) well..
We are not alone (4) well..
We are not alone (4) well..
We are not alone (4) well..
We are not alone (4) well..
Outline

- **Motivation**
  - **Fast** Faceted Search in XML
    - Implementation in eXist
    - Benchmarked
  - **Structural** Faceted Search in XML
    - Work in progress
What is the Problem?
What is the **Problem**?

- A free full-text Query
What is the **Problem**?

- A *free full-text* Query
- Counts for *each Query*, for *each FacetValue*
What is the **Problem**?

- A **free full-text** Query
- Counts for **each Query**, for **each FacetValue**
- There can be **many** FacetValues
What is the Problem?

- A free full-text Query
- Counts for each Query, for each FacetValue
- There can be many FacetValues
- Online calculation is not feasible
Fast Faceted Search

towards a solution
Fast Faceted Search

towards a solution

- Bit vector
Fast Faceted Search towards a solution

• Bit vector
  • pre-calculated for each facet value
Fast Faceted Search

towards a solution

• Bit vector
  • pre-calculated for each facetvalue
  • online for the query
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query
**Fast Faceted Search**

**towards a solution**

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query

<table>
<thead>
<tr>
<th></th>
<th>&lt;location&gt;Prague&lt;/location&gt;</th>
<th>&lt;year&gt;2011&lt;/year&gt;</th>
<th>&lt;desc&gt;An XML Conf.&lt;/desc&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Doc 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
<td>&lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
</tr>
<tr>
<td><strong>Doc 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2012&lt;/year&gt;</td>
<td>&lt;desc&gt;coming up&lt;/desc&gt;</td>
</tr>
<tr>
<td><strong>Doc 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fast Faceted Search

towards a solution

- Bit vector
  - pre-calculated for each facetvalue
  - online for the query

<table>
<thead>
<tr>
<th></th>
<th>location</th>
<th>Ams</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;An XML Conf.&lt;/desc&gt;</td>
<td>1</td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2012&lt;/year&gt; &lt;desc&gt;coming up&lt;/desc&gt;</td>
<td>1</td>
</tr>
</tbody>
</table>
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facetvalue
  - online for the query

<table>
<thead>
<tr>
<th>Doc</th>
<th>Location</th>
<th>Year</th>
<th>Description</th>
<th>Ams</th>
<th>Pra</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;An XML Conf.&lt;/desc&gt;</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2012&lt;/year&gt; &lt;desc&gt;coming up&lt;/desc&gt;</td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Fast Faceted Search  
towards a solution

- Bit vector
  - pre-calculated for each facet value
  - online for the query

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;location&gt;Prague&lt;/location&gt;</th>
<th>&lt;year&gt;2011&lt;/year&gt;</th>
<th>&lt;desc&gt;An XML Conf.&lt;/desc&gt;</th>
<th>Ams</th>
<th>Pra</th>
<th>'11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
<td>&lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2012&lt;/year&gt;</td>
<td>&lt;desc&gt;coming up&lt;/desc&gt;</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Fast Faceted Search

towards a solution

- Bit vector
  - pre-calculated for each facet value
  - online for the query

<table>
<thead>
<tr>
<th>Doc</th>
<th>location</th>
<th>year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2012&lt;/year&gt;</td>
</tr>
</tbody>
</table>
Fast Faceted Search towards a solution

- Bit vector
  - pre-calculated for each facet value
  - online for the query

<table>
<thead>
<tr>
<th>Doc</th>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;An XML Conf.&lt;/desc&gt;</td>
<td>1 0 1 0 1</td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt; &lt;year&gt;2011&lt;/year&gt; &lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td>0 1 1 0 1</td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt; &lt;year&gt;2012&lt;/year&gt; &lt;desc&gt;coming up&lt;/desc&gt;</td>
<td>1 0 0 1 0</td>
<td></td>
</tr>
</tbody>
</table>
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query

- **Boolean AND operation**

<table>
<thead>
<tr>
<th></th>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>&lt;location&gt;Prague&lt;/location&gt;&lt;year&gt;2011&lt;/year&gt;&lt;desc&gt;An XML Conf.&lt;/desc&gt;</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;&lt;year&gt;2011&lt;/year&gt;&lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt;&lt;year&gt;2012&lt;/year&gt;&lt;desc&gt;coming up&lt;/desc&gt;</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Fast Faceted Search

• Bit vector
  • pre-calculated for each facet value
  • online for the query

• Boolean AND operation

Location=Amsterdam AND query="XML"

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>location=Prague</th>
<th>year=2011</th>
<th>query=&quot;XML&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>location=Amsterdam</td>
<td>year=2011</td>
<td>query=&quot;XML&quot;</td>
</tr>
<tr>
<td>Doc 3</td>
<td>location=Prague</td>
<td>year=2012</td>
<td>query=&quot;XML&quot;</td>
</tr>
</tbody>
</table>
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query

- **Boolean AND operation**

<table>
<thead>
<tr>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ams</td>
<td>Pra</td>
<td>'11</td>
</tr>
<tr>
<td>Doc 1</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;</td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>&lt;year&gt;2012&lt;/year&gt;</td>
</tr>
</tbody>
</table>

\[\text{location=Amsterdam AND query="XML"} \rightarrow 1 \ AND \ 1 = 1\]

0 \ 1 \ 0
Fast Faceted Search towards a solution

- Bit vector
  - pre-calculated for each facet value
  - online for the query

- Boolean AND operation

<table>
<thead>
<tr>
<th></th>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>Prague</td>
<td>2011</td>
<td>XML</td>
</tr>
<tr>
<td>Doc 2</td>
<td>Amsterdam</td>
<td>2011</td>
<td>XML</td>
</tr>
<tr>
<td>Doc 3</td>
<td>Prague</td>
<td>2012</td>
<td>coming up</td>
</tr>
</tbody>
</table>

location=Amsterdam AND query=“XML” → 1 AND 1 = 1 → Doc 2
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query

- **Boolean AND operation**

<table>
<thead>
<tr>
<th></th>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>Prague</td>
<td>2011</td>
<td>XML</td>
</tr>
<tr>
<td>Doc 2</td>
<td>Amsterdam</td>
<td>2011</td>
<td>XML</td>
</tr>
<tr>
<td>Doc 3</td>
<td>Prague</td>
<td>2012</td>
<td></td>
</tr>
</tbody>
</table>

- **Cardinality calculations**

<table>
<thead>
<tr>
<th></th>
<th>location=Amsterdam AND query=“XML”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 AND 1 = 1 AND 0 = 0</td>
</tr>
</tbody>
</table>

Doc 2
Fast Faceted Search towards a solution

- **Bit vector**
  - pre-calculated for each facet value
  - online for the query

- **Boolean AND operation**

- **Cardinality calculations**

<table>
<thead>
<tr>
<th>Doc</th>
<th>location</th>
<th>year</th>
<th>query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>2011</td>
<td>“XML”</td>
</tr>
<tr>
<td></td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;desc&gt;An XML Conf.&lt;/desc&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>&lt;location&gt;Amsterdam&lt;/location&gt;</td>
<td>2011</td>
<td>“XML”</td>
</tr>
<tr>
<td></td>
<td>&lt;year&gt;2011&lt;/year&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;desc&gt;The XML Conf.&lt;/desc&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>&lt;location&gt;Prague&lt;/location&gt;</td>
<td>2012</td>
<td>“XML”</td>
</tr>
<tr>
<td></td>
<td>&lt;year&gt;2012&lt;/year&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;desc&gt;coming up&lt;/desc&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Example:**

- **location=Amsterdam AND query=“XML”**
  - Doc 2: 1 AND 1 = 1

**Cardinality calculations:**

- For **query=“XML”** the number of docs with **year=2011**
Fast Faceted Search

towards a solution

- Bit vector
  - pre-calculated for each facet value
  - online for the query

- Boolean AND operation

- Cardinality calculations
  
  for query=“XML” the number of docs with year=2011
  
  cardinality(1 AND 1) = 2
Faceted Search in XQuery
Faceted Search in XQuery

- let $hits := collection("/db/xmark")//item[facet:query(., "gold")]

Faceted Search in XQuery

- `let $hits := collection("/db/xmark")/item[facet:query(., "gold")]
- `let $counts := facet:counts($hits)"
Faceted Search in XQuery

- let $hits := collection("/db/xmark/)//item[facet:query(.) , "gold")

- let $counts := facet:counts($hits)

```xml
<facets>
  <facet name="location">
    <value name="United States">1103</value>
    <value name="Barbados">5</value>
    <value name="Gabon">4</value>
    <value name="Gambia">4</value>
    <value name="Palau">4</value>
    <value name="Cape Verde">3</value>
    <value name="Grenada">3</value>
    <value name="Myanmar">3</value>
    <value name="Namibia">3</value>
    <value name="New Zealand">3</value>
  </facet>
  <facet name="quantity">
    <value name="1">1355</value>
    <value name="2">116</value>
    <value name="3">5</value>
    <value name="4">1</value>
  </facet>
  <facet name="featured">
    <value name="yes">159</value>
  </facet>
</facets>
```
Faceted Search in XQuery

- let $hits := collection("/db/xmark/")/item[facet:query(., "gold")]
- let $counts := facet:counts($hits)
Faceted Search in XQuery

- let $hits := collection("/db/xmark/"")//item[facet:query(., "gold")]
- let $counts := facet:counts($hits)
- let $filtered := $hits[facet:filter(., ("location", "barbados"))]
Faceted Search in XQuery

- let $hits := collection("/db/xmark/")//item[facet:query(., "gold")]
- let $counts := facet:counts($hits)
- let $filtered := $hits[facet:filter(., ("location", "barbados"))]
- let $added := facet:add($filtered, "user", "gold-barbados")
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        \<text>\</text>
      </text>
    </bobo>
    <create path="/location" type="xs:string"/>
    <create path="/quantity" type="xs:string"/>
    <create path="/@featured" type="xs:string"/>
  </index>
</collection>
collection.xconf  (eXist-db specific)

<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        <facet name="location" select="location"/>
      </text>
    </bobo>
    <create path="//location" type="xs:string"/>
    <create path="//quantity" type="xs:string"/>
    <create path="//@featured" type="xs:string"/>
  </index>
</collection>
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        <facet name="location" select="location"/>
        <facet name="quantity" select="quantity"/>
      </text>
    </bobo>
    <create path="/location" type="xs:string"/>
    <create path="/quantity" type="xs:string"/>
    <create path="/featured" type="xs:string"/>
  </index>
</collection>
collection.xconf  (eXist-db specific)

```xml
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        <facet name="location" select="location"/>
        <facet name="quantity" select="quantity"/>
        <facet name="featured" select="@featured"/>
      </text>
    </bobo>
    <create path="//location" type="xs:string"/>
    <create path="//quantity" type="xs:string"/>
    <create path="//@featured" type="xs:string"/>
  </index>
</collection>
```
collection.xconf (eXist-db specific)

```xml
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        <facet name="location" select="location"/>
        <facet name="quantity" select="quantity"/>
        <facet name="featured" select="@featured"/>
        <facet name="category" select="incategory/@category" type="multi"/>
      </text>
    </bobo>
    <create path="/location" type="xs:string"/>
    <create path="/quantity" type="xs:string"/>
    <create path="/featured" type="xs:string"/>
  </index>
</collection>
```
collection.xconf (eXist-db specific)

```xml
<collection xmlns="http://exist-db.org/collection-config/1.0">
  <index>
    <fulltext default="none" attributes="no"/>
    <bobo>
      <text qname="item">
        <facet name="location" select="location"/>
        <facet name="quantity" select="quantity"/>
        <facet name="featured" select="@featured"/>
        <facet name="category" select="incategory/@category" type="multi"/>
        <facet name="date" select="/metadata/date"/>
      </text>
    </bobo>
    <create path="/location" type="xs:string"/>
    <create path="/quantity" type="xs:string"/>
    <create path="/featured" type="xs:string"/>
  </index>
</collection>
```
<collection xmlns="http://exist-db.org/collection-config/1.0">
    <index>
        <fulltext default="none" attributes="no"/>
        <bobo>
            <text qname="item">
                <facet name="location" select="location"/>
                <facet name="quantity" select="quantity"/>
                <facet name="featured" select="@featured"/>
                <facet name="category" select="incategory/@category" type="multi"/>
                <facet name="date" select="/metadata/date"/>
                <facet name="structure" type="structural"/> <!--later-->
            </text>
        </bobo>
        <create path="/location" type="xs:string"/>
        <create path="/quantity" type="xs:string"/>
        <create path="/@featured" type="xs:string"/>
    </index>
</collection>
Indexing Issues
Indexing Issues

• Just like Lucene Full-Text Indexer
Indexing Issues

• Just like Lucene Full-Text Indexer

• Supporting root selects: “//metadata”
Indexing Issues

• Just like Lucene Full-Text Indexer
• Supporting root selects: “//metadata”
• Facet values are dictated by data
Indexing Issues

• Just like Lucene Full-Text Indexer
• Supporting root selects: “//metadata”
• Facet values are dictated by data
• Normalization (none as of yet)
Indexing Issues

• Just like Lucene Full-Text Indexer
• Supporting root selects: “//metadata”
• Facetvalues are dictated by data
• Normalization (none as of yet)
• Structural facetvalues (more later)
using **Bobo-Browse** (1)
using **Bobo-Browse** (1)
using Bobo-Browse (2)
using **Bobo-Browse** (2)

- A Faceted Search Implementation, an extension of Apache Lucene
using Bobo-Browse (2)

• A Faceted Search Implementation, an extension of Apache Lucene
• Developed by and powering LinkedIn People Search
using **Bobo-Browse** (2)

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search
  - No need for *cache warm-up* for the system to perform.
using **Bobo-Browse** (2)

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search

  - No need for *cache warm-up* for the system to perform.
  - **Multi value** sort, sort documents on fields that have multiple values per doc
using **Bobo-Browse** (2)

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search
  - No need for *cache warm-up* for the system to perform.
  - **Multi value** sort, sort documents on fields that have multiple values per doc
  - **Fast** field value *retrieval* - over 30x faster than `IndexReader.document(int docid)`
using **Bobo-Browse (2)**

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search
  - No need for **cache warm-up** for the system to perform.
  - **Multi value** sort, sort documents on fields that have multiple values per doc
  - **Fast** field value **retrieval** - over 30x faster than `IndexReader.document(int docid)`
  - Facet **count distribution analysis**
using **Bobo-Browse (2)**

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search
  - No need for *cache warm-up* for the system to perform.
  - **Multi value** sort, sort documents on fields that have multiple values per doc
  - **Fast** field value *retrieval* - over 30x faster than `IndexReader.document(int docid)`
  - Facet *count distribution analysis*
  - Stable and *small memory* footprint
using Bobo-Browse (2)

• A Faceted Search Implementation, an extension of Apache Lucene
• Developed by and powering LinkedIn People Search
  • No need for cache warm-up for the system to perform.
  • Multi value sort, sort documents on fields that have multiple values per doc
  • Fast field value retrieval - over 30x faster than IndexReader.document(int docid)
  • Facet count distribution analysis
  • Stable and small memory footprint
  • Support for runtime faceting
using **Bobo-Browse (2)**

- A Faceted Search Implementation, an extension of Apache Lucene
- Developed by and powering LinkedIn People Search

- No need for **cache warm-up** for the system to perform.
- **Multi value** sort, sort documents on fields that have multiple values per doc
- **Fast** field value **retrieval** - over 30x faster than `IndexReader.document(int docid)`
- Facet **count distribution analysis**
- Stable and **small memory** footprint
- Support for **runtime faceting**
- **Result merge** library for distributed facet search
using Bobo-Browse (2)

• A Faceted Search Implementation, an extension of Apache Lucene

• Developed by and powering LinkedIn People Search

  • No need for **cache warm-up** for the system to perform.
  • **Multi value** sort, sort documents on fields that have multiple values per doc
  • **Fast** field value **retrieval** - over 30x faster than `IndexReader.document(int docid)`
  • Facet **count distribution analysis**
  • Stable and **small memory** footprint
  • Support for **runtime faceting**
  • **Result merge** library for distributed facet search
  • ...

Bobo Method

Indexing

facet:counts($hits)

facet:filter(., ("location", "barbados"))
Benchmark (1)

- XMark


<table>
<thead>
<tr>
<th>XMark scaling</th>
<th>Size (MB)</th>
<th># Files</th>
<th># items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>12</td>
<td>69</td>
<td>2175</td>
</tr>
<tr>
<td>0.2</td>
<td>24</td>
<td>136</td>
<td>4350</td>
</tr>
<tr>
<td>0.4</td>
<td>46</td>
<td>272</td>
<td>8700</td>
</tr>
<tr>
<td>0.8</td>
<td>92</td>
<td>542</td>
<td>17400</td>
</tr>
<tr>
<td>1.6</td>
<td>184</td>
<td>1084</td>
<td>34800</td>
</tr>
<tr>
<td>3.2</td>
<td>368</td>
<td>2166</td>
<td>69600</td>
</tr>
<tr>
<td>6.4</td>
<td>735</td>
<td>4330</td>
<td>139200</td>
</tr>
<tr>
<td>12.8</td>
<td>1500</td>
<td>8660</td>
<td>278400</td>
</tr>
</tbody>
</table>
Benchmark (2)

- Three Facets
  - Location up to 232 values
  - Quantity up to 7 values
  - Featured 2 values
Comparing Bobo to ...

facet:counts($hits)
BENCHMARK (4)

Comparing **Bobo** to ...

\[ \text{facet:counts} (\$\text{hits}) \]

... to **Naive Method**
Comparing Bobo to ...

facet:counts(\$hits)

... to Naive Method

... to Rvdb Method
Benchmark (4)

... to **Naive Method**
... to Naive Method

```
declare function local:counts($hits){
  let $facets := ( "$hits/location", "$hits/quantity" , "$hits/@featured" )
  let $labels := ( "locations", "quantities", "featured" )
  return <facets>
    {for $facet at $p in $facets
      let $facetvalues := util:eval($facet)
      return <facet name="{$labels[$p]}">
        {for $a in distinct-values($facetvalues)
          return <value name="{$a}">
            {count($facetvalues[. eq $a])}
          </value>
        }
      </facet>
    }
  </facets>
};
```
Benchmark (4)

... to Rvdb Method
by Ron van den Branden
... to Rvdb Method
by Ron van den Branden

declare function local:cb($term, $data){
    <value name="{$term}">
    {$data[1]}
    </value>
};

declare function local:counts($hits){
    let $cb := util:function(xs:QName("local:cb"), 2)
    let $facets := ( "$hits/location", "$hits/quantity", "$hits/@featured" )
    let $labels := ( "locations", "quantities", "featured" )
    return
    <facets>
        {for $facet at $p in $facets
            let $vals := util:eval($facet)
            return
            <facet name="{$labels[$p]}">
            {util:index–keys($vals, "", $cb, 10000)}
            </facet>
        }
    </facets>
};
Benchmark (5)

- **Data**
  - Varying in size
- **Two** scenarios
  - Large (4 steps)
  - Tiny (4 steps)
- **Three** Methods
  - Bobo
  - Naive
  - RvdB
Results: **Large** Scenario

Figure 4: Processing time in seconds per step of Scenario A, for different datasizes, see Table 1 (p. 6). Figures are for the Naive Method, Rvdb Method and the Indexed Method. Note the log scale of the vertical axis.

The development of the faceted search micro-benchmark based on XMark and its application using XCheck proved useful. The plots give quick insights in the scalability of the algorithms, possibilities for improvements and anomalies in the code. A future version of the Benchmark could include a way to measure scalability over the number of facet-value pairs.

### 6.1 Future Work

There are two obvious starting points for improvements: find the reason for the outlier in our experiment, as mentioned in Section 5.2 (p. 10). Secondly, as mentioned in Section 4.3.4 (p. 10), we should implement a solution like Wang (2009) for mapping static from position-sorted to vector-identifier probably by integrating Zoie.

Besides improvements in our implementation of the Indexed Method, we would like to extend our experiments. It would be insightful to investigate the behavior of the three methods when we vary the number of facet-values. Also, a more precise measurement of where exactly in which method time is spent would give a clear indication of where we could still gain something. And lastly, we did not mention the behavior of the methods under (heavy) updates of the dataset; we assumed a static dataset where documents are never removed or added. It is very likely that for such a scenario another method should be prefered.

**Acknowledgements**

Maarten Marx acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh
Results: Large Scenario

Figure 4: Processing time in seconds per step of Scenario A, for different datasizes, see Table 1 (p. 6). Figures are for the Naive Method, RvdB Method and the Indexed Method. Note the log scale of the vertical axis.

The development of the faceted search micro-benchmark based on XMark and its application using XCheck proved useful. The plots give quick insights in the scalability of the algorithms, possibilities for improvements and anomalies in the code. A future version of the Benchmark could include a way to measure scalability over the number of facet-value pairs.

6.1 Future Work

There are two obvious starting points for improvements: find the reason for the outlier in our experiment, as mentioned in Section 5.2 (p. 10). Secondly, as mentioned in Section 4.3.4 (p. 10), we should implement a solution like Wang (2009) for facet mapping to an sfantom - vector - identifier probably by integrating Zoie.

Besides improvements in our implementation of the Indexed Method, we would like to extend our experiments. It would be insightful to investigate the behavior of the three methods when we vary the number of facet-values. Also, a more precise measurement of where exactly in which method time is spent would give a clear indication of where we could still gain something. And lastly, we did not mention the behavior of the methods under (heavy) updates of the dataset; we assumed a static dataset where documents are never removed or added. It is very likely that for such a scenario another method should be preferred.
Results: **Large Scenario**

Figure 4: Processing time in seconds per step of Scenario A, for different datasizes, see Table 1 (p. 6). Figures are for the Naive Method, RvdB Method and the Indexed Method. Note the log scale of the vertical axis.

The development of the faceted search micro-benchmark based on XMark and its application using XCheck proved useful. The plots give quick insights in the scalability of the algorithms, possibilities for improvements and anomalies in the code. A future version of the Benchmark could include a way to measure scalability over the number of facet-value pairs.

6.1 Future Work

There are two obvious starting points for improvements: find the reason for the outlier in our experiment, as mentioned in Section 5.2. Secondly, as mentioned in Section 4.3.4 (p. 10), we should implement a solution like Wang (2009) for facet mapping to an object-identifier probably by integrating Zoie.

Besides improvements in our implementation of the Indexed Method, we would like to extend our experiments. It would be insightful to investigate the behavior of the three methods when we vary the number of facet-values. Also, a more precise measurement of where exactly in which method time is spent would give a clear indication of where we could still gain something. And lastly, we did not mention the behavior of the methods under (heavy) updates of the dataset; we assumed a static dataset where documents are never removed or added. It is very likely that for such a scenario another method should be preferred.

Acknowledgements

Maarten Marx acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh
Results: Tiny Scenario

Figure 5: Processing time in seconds per step of Scenario B, for different datasizes, see Table 1 (p. 6). Figures are for the Naive Method, RvdB Method and the Indexed Method. Note the log scale of the vertical axis.

Framework Programme for Research of the European Commission under the FET-Open grant agreement FOX, number FP7-ICT-233599. This research was supported by the Netherlands organization for Scientific Research NWO under project number 380-52-005 (PoliticalMashup).

References


Afanasiev, L. and Manolescu, I. and Michiels, P. hsqqvio MemBeR: a micron benchmark repository for XQuery Database and XML Technologies.


Figure 3: The average processing time over all steps (from both scenarios) per method for a growing dataset. Note the log scale of the vertical axis. Looking at Figure 3, we see that the Indexed Method initially performs worse than both other methods but remains near constant while the time taken by both the Naive Method and Rvdb Method is polynomial. For datasets larger than about 400MB it is preferable to use the Indexed Method. It should also be noted that an average processing time of about 1 second per query is generally not acceptable in a user interface.

Both Figure 4 (following page) and 5 (p. 13) illustrate how the first drill-down step — step 2 in each scenario — for the Indexed Method consequently takes a lot of time. While all other timings for that method stay below or around 0.2 seconds, step 2 goes up to 8 seconds. We are not able to pinpoint the cause of this behavior yet. However, we do point out that when we solve this issue the method will become within a very acceptable range with respect to processing time.

If we leave our Indexed Method aside, it is interesting to see that the Rvdb Method outperforms the Naive Method method for the earlier steps and vice versa for the later steps.

Conclusions

We conclude that the bit-vector implementation is the only one which scales with the datasize; the processing time stays near constant. Such behavior is a very desirable one in many applications. Our average running time of around 1 second for all queries on all datasets (going up to almost 300K nodes or 1.5GB) is promising but not yet good enough.
Fast Faceted Search: Wrap up
Fast Faceted Search: **Wrap up**

- Expressive XQuery interface
Fast Faceted Search: Wrap up

- Expressive XQuery interface
- Extension to eXist database
Fast Faceted Search: Wrap up

- Expressive XQuery interface
- Extension to eXist database
- Intuitive Benchmark
Fast Faceted Search: Wrap up

• Expressive XQuery interface
• Extension to eXist database
• Intuitive Benchmark
• Scales with data size
Fast Faceted Search: Wrap up

- Expressive XQuery interface
- Extension to eXist database
- Intuitive Benchmark
- Scales with data size

Open Issues

- Normalization
- Caching
- Zoie: realtime indexing and search system
Outline

- **Motivation**
- **Fast** Faceted Search in XML
  - Implementation in eXist
  - Benchmarked
- **Structural** Faceted Search in XML
  - Work in progress
Explicit Metadata

for <title/> in this example
Implicit Metadata

for <title/> in this example
Implicit Metadata

for <title/> in this example
Structural FacetValues
... using this implicit metadata

Explicit FacetValues for <title/>

isbn:9789027439642
author:Anne Schuth
edition:3
Structural FacetValues
... using this implicit metadata

Explicit FacetValues for <title/>
isbn:9789027439642
author:Anne Schuth
dition:3

Structural FacetValues for <title/>
structure:/books/book/title
structure:/books/book/chapter/title
structure:/books/book/chapter/section/title
**Structural FacetValues**

... using this implicit metadata

**Explicit FacetValues** for `<title/>`

- isbn: 9789027439642
- author: Anne Schuth
- edition: 3

**Structural FacetValues** for `<title/>`

- structure://books/book/title
- structure://books/book/chapter/title
- structure://books/book/chapter/section/title
- structure://title
**Structural** **FacetValues**

... using this implicit metadata

**Explicit** **FacetValues** *for* `<title/>`

- isbn: 9789027439642
- author: Anne Schuth
- edition: 3

**Structural** **FacetValues** *for* `<title/>`

- structure://books/book/title
- structure://books/book/chapter/title
- structure://books/book/chapter/section/title
- structure://title
- structure://book/title
Structural FacetValues
... using this implicit metadata

Explicit FacetValues for <title/>
isbn: 9789027439642
author: Anne Schuth
description: 3

Structural FacetValues for <title/>
structure:/books/book/title
structure:/books/book/chapter/title
structure:/books/book/chapter/section/title
structure://title
structure://book/title
structure://chapter/title
Structural FacetValues
... using this implicit metadata

Explicit FacetValues for <title/>

isbn:9789027439642
author:Anne Schuth
dition:3

Structural FacetValues for <title/>

structure:/books/book/title
structure:/books/book/chapter/title
structure:/books/book/chapter/section/title
structure://title
structure://book/title
structure://chapter/title
...
Structural FacetValues
... using this implicit metadata

Explicit FacetValues for <title/>
isbn: 9789027439642
author: Anne Schuth
dition: 3

Structural FacetValues for <title/>
structure://books/book/title
structure://books/book/chapter/title
structure://books/book/chapter/section/title
structure://title
structure://book/title
structure://chapter/title
...
2^2
**Structural FacetValues**

... using this implicit metadata

**Explicit FacetValues for** `<title/>`

- `isbn:9789027439642`
- `author:Anne Schuth`
- `edition:3`

**Structural FacetValues for** `<title/>`

- `structure:/books/book/title`
- `structure:/books/book/chapter/title`
- `structure:/books/book/chapter/section/title`
- `structure://title`
- `structure://book/title`
- `structure://chapter/title`

...  

\[
2^2 + 2^3
\]
Explicit FacetValues for <title/>
isbn:9789027439642
author:Anne Schuth
dition:3

Structural FacetValues for <title/>
structure:/books/book/title
structure:/books/book/chapter/title
structure:/books/book/chapter/section/title
structure://title
structure://book//title
structure://chapter/title
...

$2^2 + 2^3 + 2^4 = 28$ Structural FacetValues for <title/> (at most)
Structural FacetValues

<books>
  <book>
    <meta>
      <isbn>9789027439642</isbn>
      <author>Anne Schuth</author>
      <edition>3</edition>
    </meta>
    <title>Everything you need to know on Faceted Search</title>
    <chapter>
      <title>From a Database Perspective</title>
      <section>
        <title>The Front-end</title>
        <text/>
      </section>
      <section>
        <title>The Back-end</title>
        <text/>
      </section>
    </chapter>
    <chapter>
      <title>From an Information Retrieval Perspective</title>
      <section>
        <title>Ranking Facetvalues</title>
        <text/>
      </section>
    </chapter>
  </book>
</books>
Structural FacetValues

<books>
  <book>
    <meta>
      <isbn>9789027439642</isbn>
      <author>Anne Schuth</author>
      <edition>3</edition>
    </meta>
    <title>Everything you need to know on Faceted Search</title>
    <chapter>
      <title>From a Database Perspective</title>
      <section>
        <title>The Front-end</title>
        <text/>
      </section>
      <section>
        <title>The Back-end</title>
        <text/>
      </section>
    </chapter>
    <chapter>
      <title>From an Information Retrieval Perspective</title>
      <section>
        <title>Ranking Facetvalues</title>
        <text/>
      </section>
    </chapter>
  </book>
</books>

Doc 1 <title>Everything you need to know on Faceted Search</title>
Doc 2 <title>From a Database Perspective</title>
Doc 3 <title>The Front-end</title>
Doc 4 <title>The Back-end</title>
Doc 5 <title>From an Information Retrieval Perspective</title>
Doc 6 <title>Ranking Facetvalues</title>
Structural FacetValues

| Doc 1 | <title>Everything you need to know on Faceted Search</title> |
| Doc 2 | <title>From a Database Perspective</title> |
| Doc 3 | <title>The Front-end</title> |
| Doc 4 | <title>The Back-end</title> |
| Doc 5 | <title>From an Information Retrieval Perspective</title> |
| Doc 6 | <title>Ranking Facetvalues</title> |
## Structural FacetValues

<table>
<thead>
<tr>
<th>Doc</th>
<th>Title</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</td>
<td>1</td>
</tr>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
</tr>
</tbody>
</table>
Structural FacetValues

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>Everything you need to know on Faceted Search</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 2</td>
<td>From a Database Perspective</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 3</td>
<td>The Front-end</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 4</td>
<td>The Back-end</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 5</td>
<td>From an Information Retrieval Perspective</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 6</td>
<td>Ranking Facetvalues</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# Structural Facet Values

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Structural FacetValues

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
# Structural FacetValues

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
# Structural Facet Values

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
# Structural Facet Values

## Document 1: Everything you need to know on Faceted Search

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

## Document 2: From a Database Perspective

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

## Document 3: The Front-end

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

## Document 4: The Back-end

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

## Document 5: From an Information Retrieval Perspective

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

## Document 6: Ranking Facet Values

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
## Structural FacetValues

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
## Structural facet values

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking facet values&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
# Structural FacetValues

<table>
<thead>
<tr>
<th>Doc 1</th>
<th>&lt;title&gt;Everything you need to know on Faceted Search&lt;/title&gt;</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 2</td>
<td>&lt;title&gt;From a Database Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 3</td>
<td>&lt;title&gt;The Front-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Doc 4</td>
<td>&lt;title&gt;The Back-end&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Doc 5</td>
<td>&lt;title&gt;From an Information Retrieval Perspective&lt;/title&gt;</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Doc 6</td>
<td>&lt;title&gt;Ranking Facetvalues&lt;/title&gt;</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
# Pruning Structural Facet Values

Step 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Pruning Structural FacetValues

step 1
## Pruning Structural FacetValues

### step 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Pruning Structural Facet Values

Step 1

Prune vectors of ones
# Pruning Structural Facet Values

**step 2**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
**Pruning Structural FacetValues**

step 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Pruning Structural Facet Values

step 2
Pruning Structural Facet Values

step 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Pruning Structural FacetValues

step 2

<table>
<thead>
<tr>
<th>Structure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>books/book/title</td>
<td>1</td>
</tr>
<tr>
<td>books/book/chapter/title</td>
<td>0</td>
</tr>
<tr>
<td>books/book/chapter/section/title</td>
<td>0</td>
</tr>
<tr>
<td>chapter/chapter/title</td>
<td>0</td>
</tr>
<tr>
<td>chapter/section/title</td>
<td>1</td>
</tr>
<tr>
<td>section/title</td>
<td>0</td>
</tr>
<tr>
<td>book/title</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>
Pruning Structural Facet Values

step 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
**Pruning Structural FacetValues**

step 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

...
Pruning Structural FacetValues

step 2

prune specific paths if there are duplicate vectors
Pruning Structural Facet Values

step 3

<table>
<thead>
<tr>
<th>//chapter/title</th>
<th>//section/title</th>
<th>//chapter/title</th>
<th>//book/title</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
## Pruning Structural FacetValues

### step 3

<table>
<thead>
<tr>
<th>//chapter/title</th>
<th>//section/title</th>
<th>//chapter/title</th>
<th>//book/title</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### Pruning Structural Facet Values

#### step 3

<table>
<thead>
<tr>
<th>//chapter/title</th>
<th>//section/title</th>
<th>//chapter/title</th>
<th>//book/title</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
# Pruning Structural Facet Values

*step 3*

<table>
<thead>
<tr>
<th>//chapter/title</th>
<th>//section/title</th>
<th>//chapter/title</th>
<th>//book/title</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0 1</td>
<td>1 0 1 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 0 1 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 0 1 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
<tr>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
<td>1 1 0 0</td>
</tr>
</tbody>
</table>
Pruning Structural FacetValues

step 3

generality vs specificity
Pruning Structural FacetValues

step 3

generality vs specificity
Pruning Structural FacetValues

• Step 1
  • prune vector of ones
Pruning Structural Facet Values

• Step 1
  • prune vector of ones

• Step 2
  • prune specific paths if there are duplicate vectors
Pruning Structural FacetValues

• Step 1
  • prune vector of ones

• Step 2
  • prune specific paths if there are duplicate vectors

• Step 3
  • ?
Explicit FacetValues for <title/>

isbn: 9789027439642
author: Anne Schuth
edition: 3
Explicit FacetValues for <title/>

isbn: 9789027439642
author: Anne Schuth
description: 3

Semantical Structural FacetValues for <title/>

structure://section//title
structure://chapter/title
structure://book/title
Semantical Structural Facet Values
Semantical Structural Facet Values

- Natural way of handling information in XML
Semantical Structural FacetValues

- Natural way of handling information in XML
- Intuitive way of navigating data
Semantical Structural Facet Values

- Natural way of handling information in XML
- Intuitive way of navigating data
- Lots to be done...
Semantical Structural Facet Values

- Natural way of handling information in XML
- Intuitive way of navigating data
- Lots to be done...
  - other axis
• Natural way of handling information in XML
• Intuitive way of navigating data
• Lots to be done...
  • other axis
  • (efficient) algorithm
Semantical Structural FacetValues

- Natural way of handling information in XML
- Intuitive way of navigating data
- Lots to be done...
  - other axis
  - (efficient) algorithm
  - implementation
Semantical Structural FacetValues

- Natural way of handling information in XML
- Intuitive way of navigating data
- Lots to be done...
  - other axis
  - (efficient) algorithm
  - implementation
  - ...

Thank you