An extension to the vector model for retrieving XML documents.
Slide show

- Information Retrieval
- INEX
  - An example
  - A simpler topic
- Our Aproch
  - Equations system
  - Topics used
- Results
  - Topic 128
  - Topic 141
  - Topic 145
- Conclusions
An extension to the vector model for retrieving XML documents. — Workshop OSWIR 2005 (September 19th 2005) — p. 3/15
Plan
• RI
• INEX
  • An example
  • A simpler topic
• Approch
• Results
• Conclusions

INEX

Initiative for the Evaluation of XML Retrieval

An extension to the vector model for retrieving XML documents. – Workshop OSWIR 2005 (September 19th 2005) – p. 4/15
Plan

• RI
• INEX
  • An example
  • A simpler topic
• Aproch
• Results
• Conclusions

An extension to the vector model for retrieving XML documents. — Workshop OSWIR 2005 (September 19th 2005) — p. 4/15
INEX

- Initiative for the Evaluation of XML Retrieval
- 2004
- 12 108 XML files

Plan
- RI
- INEX
  - An example
  - A simpler topic
- Approach
- Results
- Conclusions
INEX

- Initiative for the Evaluation of XML Retrieval
- 2004
- 12,108 XML files
- 74 Topics
INEX

- Initiative for the Evaluation of XML Retrieval
- 2004
- 12,108 XML files
- 74 Topics
  - 40 on content only
  - 34 on content and structure

Plan
- RI
- INEX
  - An example
  - A simpler topic
- Approach
- Results
- Conclusions
INEX

Initiative for the Evaluation of XML Retrieval
2004
12 108 XML files
74 Topics
  40 on content only
  34 on content and structure
Assessments on each topic

Plan
• RI
  INEX
  • An example
  • A simpler topic
• Aproch
• Results
• Conclusions
An example

Topic 139:

//article[
    (about(./bb//au//snm, Bertino)
    or about(./bb//au//snm, Jajodia))
    and about(./bb//atl, security model)
    and about(./bb//atl, -"indexing model"
             - "object oriented model")
]
An example

Topic 139:

```xml
//article[
  (about(.//bb//au//snm, Bertino)
  or about(.//bb//au//snm, Jajodia))
  and about(.//bb//atl, security model)
  and about(.//bb//atl, "indexing model"
  - "object oriented model")
]
```

too many operators to combine

- about
- and, or
- descendance
- proximity
- etc.
An example

• Topic 139:

  //article[
    (about(./bb//au//snm, Bertino)
    or about(./bb//au//snm, Jajodia))
    and about(./bb//atl, security model)
    and about(./bb//atl, "indexing model"
      - "object oriented model")
  ]

  too many operators to combine
  • about
  • and, or
  • descendance
  • proximity
  • etc.

  We’ll begin with a simpler topic
A simpler topic

Topic 141:
//article[about(., java)]//sec[about(., implementing threads)]

Plan
• RI
  ▶ INEX
    • An example
    → A simpler topic
    • Approch
    • Results
    • Conclusions
A simpler topic

Topic 141:
//article[about(., java)]//sec[about(., implementing threads)]
  about
  article and section association
Our Approach

- We can use classical IR methods to solve:

Plan

- RI
- INEX
- Approach
  - Equations system
  - Topics used
  - Results
  - Conclusions
Our Approach

we can use classical IR method to solve:

//article[about(., java)]
Our Aproch

we can use clasical IR method to solve:

\[ \text{article[about(., java)]:} \rightarrow \{(a_i, w_{a_i})\} \]
Our Approach

- we can use classical IR method to solve:
  - //article[about(., java)]
    → \{(a_i, w_{a_i})\}
  - //sec[about(., implementing threads)]
Our Approach

we can use classical IR method to solve:

- //article[about(., java)]
  → \{(a_i, w_{a_i})\}

- //sec[about(., implementing threads)]
  → \{(s_{i,j}, w_{s_{i,j}})\}
Our Approach

we can use classical IR method to solve:

- `//article[about(., java)]`
  → `{(a_i, w_{a_i})}`

- `//sec[about(., implementing threads)]`
  → `{(s_{i,j}, w_{s_{i,j}})}`

we now that `sec` must be descendant of an `article`
  → `{(s_{i,j}, w_{a_i}, w_{s_{i,j}})}`
Our Approach

- We can use classical IR method to solve:
  - `//article[about(. , java)]` → \{(a_i, w_{a_i})\}
  - `//sec[about(. , implementing threads)]` → \{(s_{i,j}, w_{s_{i,j}})\}

- We now that `sec` must be descendant of an `article` → \{(s_{i,j}, w_{a_i}, w_{s_{i,j}})\}

- How to combine $w_{a_i}$ and $w_{s_{i,j}}$?
Our Aproch

we can use clasical IR method to solve:

- `//article[about(., java)]`
  \[\rightarrow \{ (a_i, w_{a_i}) \}\]

- `//sec[about(., implementing threads)]`
  \[\rightarrow \{ (s_{i,j}, w_{s_{i,j}}) \}\]

we now that `sec` must be descendant of an `article`
\[\rightarrow \{(s_{i,j}, w_{a_i}, w_{s_{i,j}})\}\]

How to combine \(w_{a_i}\) and \(w_{s_{i,j}}\)?

We use:
\[\rightarrow R_{i,j} = \alpha * w_{a_i} + \beta * w_{s_{i,j}} + \gamma * w_{a_i} * w_{s_{i,j}} + \delta * \min(w_{a_i}, w_{s_{i,j}})\]
Equations system

We define the system:

\[
\begin{align*}
\alpha w_{a1} + \beta w_{s1,1} + \gamma w_{a1} w_{s1,1} + \delta \min(w_{a1}, w_{s1,1}) &= R_{1,1} \\
\alpha w_{a1} + \beta w_{s1,2} + \gamma w_{a1} w_{s1,2} + \delta \min(w_{a1}, w_{s1,2}) &= R_{1,2} \\
\alpha w_{a1} + \beta w_{s1,3} + \gamma w_{a1} w_{s1,3} + \delta \min(w_{a1}, w_{s1,3}) &= R_{1,3} \\
\vdots \\
\alpha w_{a1} + \beta w_{s1,m_1} + \gamma w_{a1} w_{s1,m_1} + \delta \min(w_{a1}, w_{s1,m_1}) &= R_{1,m_1} \\
\alpha w_{a2} + \beta w_{s2,1} + \gamma w_{a2} w_{s2,1} + \delta \min(w_{a2}, w_{s2,1}) &= R_{2,1} \\
\alpha w_{a2} + \beta w_{s2,2} + \gamma w_{a2} w_{s2,2} + \delta \min(w_{a2}, w_{s2,2}) &= R_{2,2} \\
\vdots \\
\alpha w_{a_i} + \beta w_{s_i,j} + \gamma w_{a_i} w_{s_i,j} + \delta \min(w_{a_i}, w_{s_i,j}) &= R_{i,j} \\
\vdots \\
\alpha w_{a_n} + \beta w_{s_n,m_n} + \gamma w_{a_n} w_{s_n,m_n} + \delta \min(w_{a_n}, w_{s_n,m_n}) &= R_{n,m_n}
\end{align*}
\]
Equations system

We define the system:

\[
\begin{align*}
\alpha w_{a1} + \beta w_{s1,1} + \gamma w_{a1} w_{s1,1} + \delta \min(w_{a1}, w_{s1,j}) &= R_{1,1} \\
\alpha w_{a1} + \beta w_{s1,2} + \gamma w_{a1} w_{s1,2} + \delta \min(w_{a1}, w_{s1,j}) &= R_{1,2} \\
\alpha w_{a1} + \beta w_{s1,3} + \gamma w_{a1} w_{s1,3} + \delta \min(w_{a1}, w_{s1,j}) &= R_{1,3} \\
\vdots & \\
\alpha w_{a1} + \beta w_{s1,m_{1}} + \gamma w_{a1} w_{s1,m_{1}} + \delta \min(w_{a1}, w_{s_{1,m_{1}}}) &= R_{1,m_{1}} \\
\alpha w_{a2} + \beta w_{s2,1} + \gamma w_{a2} w_{s2,1} + \delta \min(w_{a2}, w_{s2,1}) &= R_{2,1} \\
\alpha w_{a2} + \beta w_{s2,2} + \gamma w_{a2} w_{s2,2} + \delta \min(w_{a2}, w_{s2,2}) &= R_{2,2} \\
\vdots & \\
\alpha w_{ai} + \beta w_{si,j} + \gamma w_{ai} w_{si,j} + \delta \min(w_{ai}, w_{si,j}) &= R_{i,j} \\
\vdots & \\
\alpha w_{an} + \beta w_{sn,m_{n}} + \gamma w_{an} w_{sn,m_{n}} + \delta \min(w_{an}, w_{sn,m_{n}}) &= R_{n,m_{n}}
\end{align*}
\]

with the assessments:

<table>
<thead>
<tr>
<th>article</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>...</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>i</th>
<th>...</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>section</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>$m_{i}$</td>
<td>1</td>
<td>2</td>
<td>j</td>
<td>...</td>
<td>$m_{n}$</td>
</tr>
<tr>
<td>user relevance</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1/3</td>
<td>1</td>
<td>0</td>
<td>k</td>
<td>...</td>
<td>2/3</td>
</tr>
</tbody>
</table>
Equations system

We define the system:

\[
\begin{align*}
\alpha \cdot w_{11} + \beta \cdot w_{12} + \gamma \cdot w_{13} + \delta \cdot \min(w_{11}, w_{12}, w_{13}) &= 1/3 \\
\alpha \cdot w_{21} + \beta \cdot w_{22} + \gamma \cdot w_{23} + \delta \cdot \min(w_{21}, w_{22}, w_{23}) &= 0 \\
\alpha \cdot w_{31} + \beta \cdot w_{32} + \gamma \cdot w_{33} + \delta \cdot \min(w_{31}, w_{32}, w_{33}) &= 0 \\
\vdots \\
\alpha \cdot w_{i1} + \beta \cdot w_{i2} + \gamma \cdot w_{i3} + \delta \cdot \min(w_{i1}, w_{i2}, w_{i3}) &= k \\
\vdots \\
\alpha \cdot w_{n1} + \beta \cdot w_{n2} + \gamma \cdot w_{n3} + \delta \cdot \min(w_{n1}, w_{n2}, w_{n3}) &= 2/3
\end{align*}
\]

with the assements:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>...</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>...</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>article</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>section</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>mi</td>
<td>1</td>
<td>2</td>
<td>j</td>
<td></td>
<td>mn</td>
</tr>
<tr>
<td>user relevance</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1/3</td>
<td>1</td>
<td>0</td>
<td>k</td>
<td></td>
<td>2/3</td>
</tr>
</tbody>
</table>
Topics used

We used these three similar topics for testing this method:

- **Topic 128:**
  
  //article[about(. , intelligent transport systems)]//sec[about(. , on-board route planning navigation system for automobiles)]

- **Topic 141:**
  
  //article[about(. , java)]//sec[about(. , implementing threads)]

- **Topic 145:**
  
  //article[about(. , information retrieval)]//p[about(. , relevance feedback)]

- We use a simple least squares method to solve the three systems
Results

we obtain values for the unknown quantities:

<table>
<thead>
<tr>
<th>Topic</th>
<th>128</th>
<th>141</th>
<th>145</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0.051</td>
<td>0.014</td>
<td>0.009</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.085</td>
<td>0.059</td>
<td>0.123</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>-0.753</td>
<td>0.769</td>
<td>0.506</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.567</td>
<td>0.265</td>
<td>0.134</td>
</tr>
</tbody>
</table>

Now we can reintroduce these values into each system to draw the Precision-Recall graphics.
An extension to the vector model for retrieving XML documents. – Workshop OEWIR 2005 (September 19th 2005) – p. 11/15

Topic 128

//article[about(., intelligent transport systems)]//sec[about(., on-board route planning navigation system for automobiles)]

![Graph showing precision and recall for Topic 128's relevance.]
An extension to the vector model for retrieving XML documents. – Workshop OSEIR 2005 (September 19th 2005) – p. 12/15
An extension to the vector model for retrieving XML documents. – Workshop OSWIR 2005 (September 19th 2005) – p. 13/15

Topic 145

//article[about(., information retrieval)]//p[about(., relevance feedback)]

Plan
• RI
• INEX
• Aproch
⇒ Results
• Topic 128
• Topic 141
⇒ Topic 145
• Conclusions
Conclusions

- Experimental results only partially validate our hypothesis
Conclusions

- Experimental results only partially validate our hypothesis
- The relevance of the article has an impact
Conclusions

- Experimental results only partially validate our hypothesis
  - The relevance of the article has an impact
  - The multication tends to be the most important factor (AND operator ?)
Conclusions

- Experimental results only partially validate our hypothesis
  - The relevance of the article has an impact
  - the multication tends to be the most important factor (AND operator ?)
- The function that we use to evaluate the relevance is quite simple
Conclusions

- Experimental results only partially validate our hypothesis
  - The relevance of the article has an impact
  - the multication tends to be the most important factor (AND operator ?)
- The function that we use to evaluate the relevance is quite simple
  - we can take proximity of term in account for example
Conclusions

- Experimental results only partially validate our hypothesis
  - The relevance of the article has an impact
  - the multiplication tends to be the most important factor (AND operator ?)
- The function that we use to evaluate the relevance is quite simple
  - we can take proximity of term in account for example
- The future works may be to try to analyze each other nexi operators independently
  then try some "harder" topics
Thank you for your attention