Exploratory Search Pipes with Scoped Facets

Tim Gollub  Leon Hutans  Tanveer Al Jami  Benno Stein
Bauhaus-Universität Weimar

webis.de
Digital Humanities Project.

We develop distant reading devices.

First device is a faceted search system for content-level explorations of document collections.
Classical Facet Search Systems

Documents

D1  D2  D3  D4  D5  D6
Classical Facet Search Systems

Facets

Facets

Facet Terms

A1
A2

B1
B2

Documents

D1
D2
D3
D4
D5
D6
Issue: Documents as atomic context units are not flexible enough to accommodate for content-based explorations.
Span-Based Faceted Search
Annotation of Character Spans

Facets
A
B

Facet Terms
A1
A2
B1
B2

Documents
D1
D2
D3
D4
D5
D6
Benefit: Evaluation of facets within different contexts is now possible.
Span-Based Faceted Search

Three Facets Case

Facets

A

B

C

Facet Terms

A1

A2

B1

B2

C1

C2

Documents

D1

D2

D3

D4

D5

D6
Issue: A common use case is to search for facet terms (e.g. entities) that occur in a common context. This cannot be done with global intersections.
Benefit: Facets are evaluated sequentially one after the other. The result set for a facet is determined from intersections with the character spans of the facets terms from the facet to the left (if any).
Issue: A common use case is to filter the facet terms of facet B with respect to A2, but then to use all spans of the remaining facet terms in Facet B (=B1) for the intersections with Facet C.
Benefit: With the introduction of facet scopes, the user can decide which character spans of a facet term to take into account.
## Prototype Podascope

**Facet Navigation and Search**

### Facets (15)

<table>
<thead>
<tr>
<th>Facet</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>70601</td>
</tr>
<tr>
<td>Company</td>
<td>282</td>
</tr>
<tr>
<td>Conferences</td>
<td>47</td>
</tr>
<tr>
<td>Country</td>
<td>65</td>
</tr>
<tr>
<td>Database</td>
<td>5</td>
</tr>
<tr>
<td>Disease</td>
<td>56</td>
</tr>
<tr>
<td>Language</td>
<td>96</td>
</tr>
<tr>
<td>Location</td>
<td>325</td>
</tr>
<tr>
<td>Organisation</td>
<td>708</td>
</tr>
<tr>
<td>Person</td>
<td>209</td>
</tr>
<tr>
<td>Programming Language</td>
<td>39</td>
</tr>
</tbody>
</table>

Scope: 0

Total 15 results

Search

© webis.de 2019
Prototype Podascope
Facet Navigation and Search

Conferences (1/47)

Scope: 0

/ Conferences ictir

Total 1/47 results

ICTIR
Prototype Podascope
Pipe Construction

Conferences (1/47) Publications (314)

Scope: -1

Total 314 results

1 "A term is known by the company it keeps": On Selecting a Good Expansion Set in Pseudo-Relevance Feedback.
1 <i>Merge-Tie-Judge</i>: Low-Cost Preference Judgments with Ties.
1 A Belief Model of Query Difficulty That Uses Subjective Logic.
1 A Collaborative Ranking Model with Multiple Location-based Similarities for Venue Suggestion.
1 A Comparative Study of Pseudo Relevance Feedback for Ad-hoc Retrieval.
1 A Contextual Bandit Approach to Dynamic Search.
1 A Descriptive Approach to Classification.
1 A Diagnostic Study of Search Result Diversification Methods.
1 A Formal Account of Effectiveness Evaluation and Ranking Fusion.
## Prototype Podascope

**Pipe Construction**

### Authors (528)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Leif Azzopardi</td>
</tr>
<tr>
<td>10</td>
<td>Oren Kurland</td>
</tr>
<tr>
<td>10</td>
<td>W. Bruce Croft</td>
</tr>
<tr>
<td>9</td>
<td>Guido Zuccon</td>
</tr>
<tr>
<td>8</td>
<td>Haggai Roitman</td>
</tr>
<tr>
<td>7</td>
<td>Ben Carterette</td>
</tr>
<tr>
<td>7</td>
<td>ChengXiang Zhai</td>
</tr>
<tr>
<td>7</td>
<td>Christina Lioma</td>
</tr>
<tr>
<td>7</td>
<td>Craig Macdonald</td>
</tr>
<tr>
<td>7</td>
<td>Dawei Song 0001</td>
</tr>
<tr>
<td>7</td>
<td>Hui Fang 0001</td>
</tr>
</tbody>
</table>

Total 528 results

Scope: -2
## Prototype Podascope

### Scope Reduction

<table>
<thead>
<tr>
<th>Conferences (1/47)</th>
<th>Publications (314)</th>
<th>Authors (528)</th>
<th>Conferences (47)</th>
</tr>
</thead>
</table>

Scope: -1

```
Total 47 results
```

<table>
<thead>
<tr>
<th>528</th>
<th>ICTIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>340</td>
<td>SIGIR</td>
</tr>
<tr>
<td>293</td>
<td>CIKM</td>
</tr>
<tr>
<td>260</td>
<td>ECIR</td>
</tr>
<tr>
<td>208</td>
<td>TREC</td>
</tr>
<tr>
<td>131</td>
<td>WSDM</td>
</tr>
<tr>
<td>98</td>
<td>AIRS</td>
</tr>
<tr>
<td>76</td>
<td>JCDL</td>
</tr>
<tr>
<td>73</td>
<td>AAAI</td>
</tr>
<tr>
<td>72</td>
<td>CHIIR</td>
</tr>
<tr>
<td>62</td>
<td>IliX</td>
</tr>
</tbody>
</table>
User Study
Comparison with DBLP

- 14 participants with computer science background.
- Two phases: (P1) Simple search tasks. (P2) Complex search tasks.
- User Experience Questionnaire (UEQ)

Overall, our system PODA (red) is assessed equally to DBLP (black) after P1 (dashed), clearly superior to DBLP after P2 (solid).
Thank you!