Adaptively Modelling the Context of an Intranet Query

Flatlands ’08

Deirdre Lungley
dmlung@essex.ac.uk

University of Essex

June 6, 2008
Adaptively Modelling the Context of an Intranet Query

Search results

- Results for web pages and other online documents
- Results from the phonebook

Results for web pages and other online documents

You searched essex.ac.uk for ces
Results 1-10 of estimated 625 ordered by relevance:

Department of Computing and Electronic Systems :: I+I+ Departmental Newsletter
... AAMAS Award for CES Academic Maria Fasli has recently ...
http://www.essex.ac.uk/departments/news/newsletter/19_06_08.aspx

Department of Computing and Electronic Systems :: Home Page
... AAMAS Award for CES Academic | Computational Intelligence in ...
http://www.essex.ac.uk/depts/

Department of Computing and Electronic Systems :: I+I+ Departmental Newsletter
... 2 to view: CES Professor on government's "Blame ... http://www.essex.ac.uk/departments/news/newsletter/19_06_08.aspx

Department of Computing and Electronic Systems :: I+I+ Departmental Newsletter
... sample tracks are shown below: CES Research featured on ITV Local ...
http://www.essex.ac.uk/departments/news/newsletter/16_02_08.aspx

Figure: University of Essex Intranet Search
Adaptively Modelling the Context of an Intranet Query

Motivation

- Users do like some help!
  - Kruschwitz and Al-Bakour, 2005
  - White and Ruthven, 2006
    - known-item search - query suggestions
    - exploratory search - query destinations

- Web examples
  - Clusty - Vivisimo Ltd.
  - CREDO - FUB, Italy

- Intranet example
  - Aquabrowser - Medialab Solutions, The Netherlands

- Analysis of UoE intranet search modifications
  - Dominated by single-term queries
  - Many of these queries met by documents in top 5 results
  - However, how about?
    - Multi-context terms - sport, parking, printing
    - Ambiguous terms - CES
### Query Table

<table>
<thead>
<tr>
<th></th>
<th>Query Term(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>library</td>
</tr>
<tr>
<td>2</td>
<td>accommodation</td>
</tr>
<tr>
<td>3</td>
<td>exam timetable</td>
</tr>
<tr>
<td>4</td>
<td>timetable</td>
</tr>
<tr>
<td>5</td>
<td>courses</td>
</tr>
<tr>
<td>6</td>
<td>accommodation</td>
</tr>
<tr>
<td>7</td>
<td>fees</td>
</tr>
<tr>
<td>8</td>
<td>moodle</td>
</tr>
<tr>
<td>9</td>
<td>mba</td>
</tr>
<tr>
<td>10</td>
<td>graduation</td>
</tr>
</tbody>
</table>

**Table:** Prominent Modified Queries
Research Focus

- All require a domain model
- Non-trivial task
  - Relying on appropriate document annotation
- Our answer!
  - Automatically adapt our domain model - let it learn from implicit user feedback (clickthrough data)
- Current uses of clickthrough data
  - Re-ranking of results
  - Query refinement
Why Intranet Search?

- Controlled environments
  - Often imposed annotation standards
  - Less spam, making inlinks and metadata more reliable
- Relatively cohesive community of users
  - Similar search needs aid the viability of harnessing user population feedback
Components

- Underlying Search Engine
  - Lucene’s Nutch

- Natural Language Processing
  - QTag
  - Collocations (Justeson and Katz)
    - AN, NN, AAN, ANN, NAN, NNN, NPN

- Context Model
  - Formal Concept Analysis (FCA)

- Machine Learning
  - SVM-Light (Joachims)
Adaptively Modelling the Context of an Intranet Query

Figure: System Architecture
### Figure: Classical Lattice Example - Hasse Table

<table>
<thead>
<tr>
<th></th>
<th>horse</th>
<th>male</th>
<th>female</th>
<th>adult</th>
<th>young</th>
</tr>
</thead>
<tbody>
<tr>
<td>horse</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stallion</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>mare</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>foal</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>filly</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>colt</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Adaptively Modelling the Context of an Intranet Query

Figure: Classical Lattice Example - Concept Lattice
Automatically Maintained Domain Knowledge

University of Essex: travel information
... campus and secure cycle parking facilities. There are nine separate cycle routes and parking spaces are available on the...
http://www.essex.ac.uk/about/find.html

Travel Directions
... it on downloadable directions. Parking Visitor parking is available in the ... a 300m walk away
Parking Charges...
http://www.essex.ac.uk/labs/co-lister/directions.hmt

Campus Guide: finding your way around
... campus and secure cycle parking facilities. There are nine separate cycle routes and parking spaces are available on the...
http://www.essex.ac.uk/about/grade.html

Venue Hire
... computer. We offer free parking. on call technical support and ... of 12 guest:rooms here
Parking permits for guests...
http://www.essex.ac.uk/labs/co-lister/hallway_charges.html

Lab
... delegate rates including all refreshments, parking, venue hire, and the ...
http://www.essex.ac.uk/labs/co-lister/booking_faq.html

Venue Essex: Colchester campus conference and accommodation facilities at The University of Essex, UK
... to 1000. July - September FREE parking To find out...
http://www.essex.ac.uk/venue%20Guides/ccc.html

Venue Essex: Southend campus facilities at University of Essex, UK
... car links and parking is a brand ...
http://www.essex.ac.uk/venue%20Guides/ccc.html

Wyvern: Research, May 2006
... road access and parking issues, and there ...
http://www.essex.ac.uk/venue%20Guides/ccc.html

Wyvern: People: April 2007
... About 10 minutes driving, but parking could be a problem ...
http://www.essex.ac.uk/venue%20Guides/ccc.html

Sports Resources: Details - University of Essex Sports Centre
... the Sports Centre. Assistance with parking on campus. A place...

Figure: Automade Screenshot
Adaptively Modelling the Context of an Intranet Query

**Figure:** Example Adapted Lattice
Adaptively Modelling the Context of an Intranet Query

SVM-Light

- Machine learning tool developed by Thorsten Joachims
- Particularly suitable for Information Retrieval - developed to surmount the problem of sparsity in document/term matrices
- Default linear kernel
- Lattice-based kernel to optimize the lattice structure?
Clickthrough Data

- Questions have been raised regarding the accuracy of using clickthrough data as an indicator of relevance.
- Radlinski and Joachims promote relative relevance as against absolute relevance.
  - A document clicked on for a query is deemed more relevant to that particular query than documents above and below.
Adaptively Modelling the Context of an Intranet Query

**Adaptation Steps**

- **Record Log Data**
  - Log initial query term
  - Log subsequent query terms either entered in the textbox or chosen by clicking on the lattice node
  - Log clicked URL plus subsequent browser clicked URLs (possibly not within result list)

- **Adaptive Element.** Before creation of query lattice apply SVM-Light Model. This should:
  - Associate query terms positively with the clicked URLs and negatively with skipped URLs (i.e., increase/decrease document/term weight)
  - Decrease weight of document terms not in query terms
  - If query term does not exist in document terms, add with positive weight
  - Apply threshold to delete terms within documents and entire documents where all terms deleted
Thank You!