Adaptive Search at Essex

Udo Kruschwitz

School of Computer Science and Electronic Engineering
University of Essex
udo@essex.ac.uk

7th October 2011
THINGS ON THE FRONT PAGE OF A UNIVERSITY WEBSITE

- CAMPUS PHOTO SLIDESHOW
- PROMOTIONS FOR CAMPUS EVENTS
- PRESS RELEASES
- STATEMENT OF THE SCHOOL'S PHILOSOPHY
- LETTER FROM THE PRESIDENT
- VIRTUAL TOUR

THINGS PEOPLE GO TO THE SITE LOOKING FOR

- LIST OF FACULTY PHONE NUMBERS AND EMAILS
- CAMPUS ADDRESS
- ACADEMIC CALENDAR
- CAMPUS POLICE PHONE NUMBER
- DEPARTMENT/ COURSE LISTS
- PARKING INFORMATION
- USABLE CAMPUS MAP

FULL NAME OF SCHOOL

(Source: http://xkcd.com/773)
Our Approach

- Search system that makes suggestions using automatically extracted domain knowledge
- But ...
  - Domain knowledge is noisy and incomplete
  - System suggestions not always useful/helpful
  - Document collection is changing
- Learn from the users’ interactions
- Improve system over time by adapting to the users’ search behaviour
- No single user profile but “community profile”
Motivation

Adaptive Search

Evaluation

Current Work

Conclusions

Adaptive Search - LAC Day - 7 October 2011
Partial Domain Knowledge (Web Site)
Partial Domain Knowledge (Different Web Site)

(visualisation: http://moritz.stefaner.eu/projects/relation-browser/)
Partial Domain Knowledge (Digital Library)
Using Log Data to Acquire a Domain Model

- Queries submitted by users
- Identify sessions
- Associate related queries (many possible ways of doing so)
- Result is a query association graph (of some sort)
Our Log Data

- We use query logs collected on different collections, e.g.
  - University of Essex intranet search engine:
    more than 2 million queries (since Nov 2007)
  - The European Library:
    1.8 million interactions (Jan 2007 - Jun 2008)
- Query log analysis (not discussed here)
- Bootstrap (adaptive) domain models
Towards Adaptive Search

- Start by employing initially extracted domain knowledge
- Observe user interaction with the system
- Incorporate clickthrough trails
- Use this *implicit relevance feedback* to adjust domain knowledge accordingly
- Do this fully automatically
- Aim: evolving domain knowledge that adjusts to the users’ search behaviour
- Should learn common patterns over time, e.g. “map” → “campus map”
- Should deal with seasonal terms appropriately, e.g. “registration”

This should improve search ...
... and Navigation
Automatic Domain Model Adaptation

Variety of adaptation models, including:

- Exploiting Maximum Likelihood Estimates (MLE)
- Formal Concept Analysis (FCA)
- Ant Colony Optimization analogy (ACO)
- Enhanced Query Flow Graph
- Adaptive Intranet Navigation

... let’s quickly look at the first three approaches.
### MLE: Domain Model derived from Query Logs

<table>
<thead>
<tr>
<th>q1</th>
<th>q2</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>registration</td>
<td>online registration</td>
<td>0.045</td>
</tr>
<tr>
<td>registration</td>
<td>registration office</td>
<td>0.035</td>
</tr>
<tr>
<td>registration</td>
<td>timetable</td>
<td>0.025</td>
</tr>
<tr>
<td>registration</td>
<td>enrol</td>
<td>0.020</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>online registration</td>
<td>registration</td>
<td>0.211</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>registration office</td>
<td>careers centre</td>
<td>0.053</td>
</tr>
<tr>
<td>registration office</td>
<td>albert sloman library</td>
<td>0.053</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>enrol</td>
<td>course enrolment</td>
<td>0.050</td>
</tr>
</tbody>
</table>
MLE: Domain Model derived from Query Logs II
MLE: Reminder - Original Domain Knowledge

- registration
  - dates
  - card
  - essex
  - students
    - office
    - regulations
    - jobshop
    - undergraduate
MLE: Results

- Studied 3-year log file of Essex search engine (about 1.6 million queries)
- Sampled frequent and less frequent queries
- User study to assess quality of derived term suggestions
- Compared these terms to alternatives, e.g. snippet approach, association rules approach, query flow graph, n-grams etc.
- Simple log analysis using MLE beats alternatives
- Session-based approach very powerful, dialogue-based even better
FCA Approach to Adaptation

- Lattice structure representing terms and corresponding documents
- Concept in lattice defined by objects (URLs) and attributes (terms)
FCA Approach to Adaptation II

- Learn from past user queries (implicit relevance judgements) using relative judgements (Radlinski & Joachims, 2005)
- Train a classifier (SVM) that associates terms with documents
- Rerun lattice construction

... for more details ask Deirdre
ACO: Adaptive Domain Model

- Simple approach using the idea of ant colony optimization:
  - Update domain model in daily batches
  - Add weight to query pairs observed that day, e.g. “library” → “library”
  - Normalise the weights so that all outgoing graphs in a node sum up to 1
  - And so on ...
- Idea: learn associations as they become popular, allow for forgetting relations as well!
ACO: Results

- Again: user studies to assess quality of derived term suggestions
- Two studies: Essex university logs (Essex), European Library logs (TEL)
- Compared these terms to alternatives, e.g. Google suggestions, association rule approach, snippet processing
- Essex: ACO beats all alternatives and suggestions improve over time
- TEL: ACO better than association rule approach but not snippet baseline
- Suggestions derived using different methods can be complementary (TEL)
**AutoEval**: Evaluate Adaptive Search

- Limitations of user studies
- Evaluate suggestions *without* recruiting subjects
- Compare different models automatically
- Idea: use log files and exploit past user interactions

![Diagram](image)

**Score Calculation**

\[
score = \frac{1/2 + 1/4 + 1/1}{3} = 0.583
\]
AutoEval Results (Web Site)
AutoEval Results (Digital Library)
Current Work

- Prototype went live at the University of Essex last weekend
- Work with industrial partners
- TREC 2011 (session track) and CLEF 2011 (LogCLEF)
- Ongoing EPSRC project (Essex, Robert Gordon University Aberdeen & Open University): AutoAdapt (November 2008 - May 2012)
Conclusions

- Adaptive search by exploiting query logs
- Adaptive domain models can be learned, experiments with different approaches demonstrate this
- Have to deal with noisy data
- Data sparsity
- Navigation support as a suitable alternative to query suggestions
Acknowledgements

- Dyaa Albakour, Nikolaos Nanas, Jinzhong Niu, Dawei Song, Anne De Roeck, Maria Fasli, Stephen Dignum (EPSRC-funded AutoAdapt research project: http://autoadaptproject.org/
- Deirdre Lungley, Sharhida Saad
- Johannes Leveling (DCU)
References (ACO, QFG)


References (FCA, Adaptive Navigation)

References (Evaluation Methodology, Survey Paper)


References (Essex at TREC and CLEF)


For your Diary:

- Search Solutions 2011 in London
- Dedicated to the latest innovations in web and enterprise search
- Mix of research and application
- Excellent list of speakers
- Highly interactive and collegial, with attendance limited to 60-80 delegates
- Date: 16th November
That’s it ...