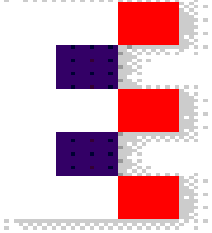


Adaptively Modelling the Context of an Intranet Query

Motivation
Research Plan
FCA
Machine Learning

Flatlands '08

Deirdre Lungley
dmlung@essex.ac.uk

University of Essex

June 6, 2008

Adaptively Modelling the Context of an Intranet Query

Motivation
Starting Point
Users Like Help
Query Table
Research Focus
Intranet Search
Research Plan
FCA
Machine Learning

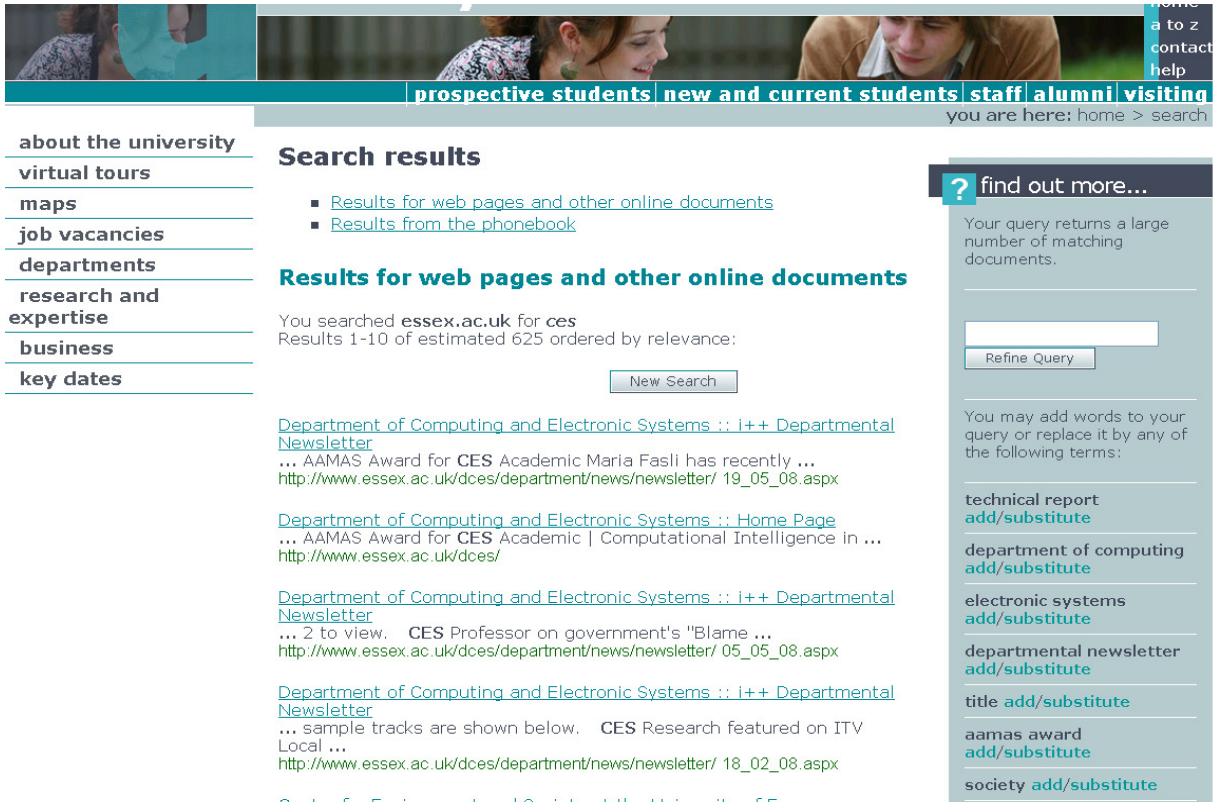
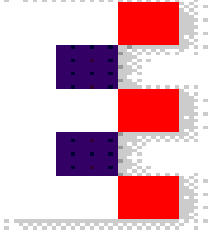


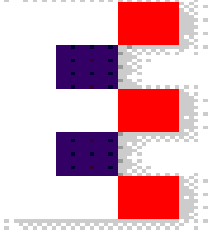
Figure: University of Essex Intranet Search



Motivation
Starting Point
Users Like Help
Query Table
Research Focus
Intranet Search
Research Plan
FCA
Machine Learning

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

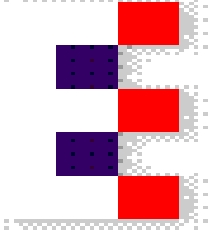


Adaptively Modelling the Context of an Intranet Query

Motivation
Starting Point
Users Like Help
Query Table
Research Focus
Intranet Search
Research Plan
FCA
Machine Learning

	Query Term(s)
1.	library
2.	accomodation
3.	exam timetable
4.	timetable
5.	courses
6.	accommodation
7.	fees
8.	moodle
9.	mba
10.	graduation

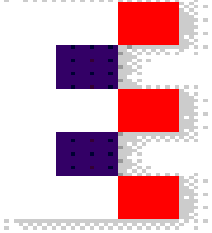
Table: Prominent Modified Queries



Motivation
Starting Point
Users Like Help
Query Table
Research Focus
Intranet Search
Research Plan
FCA
Machine Learning

Research Focus

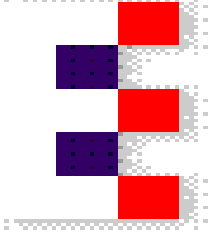
- All require a domain model
- Non-trivial task
  - Relying on appropriate document annotation
- Our answer!
  - Automatically adapt our domain model - let it learn from implicit user feedback (clickthrough data)
- Current uses of clickthrough data
  - Re-ranking of results
  - Query refinement



Motivation
Starting Point
Users Like Help
Query Table
Research Focus
Intranet Search
Research Plan
FCA
Machine Learning

Why Intranet Search?

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



Motivation
Research Plan
Components
System Architecture
FCA
Machine Learning

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

Motivation
Research Plan
Components
System Architecture
FCA
Machine Learning

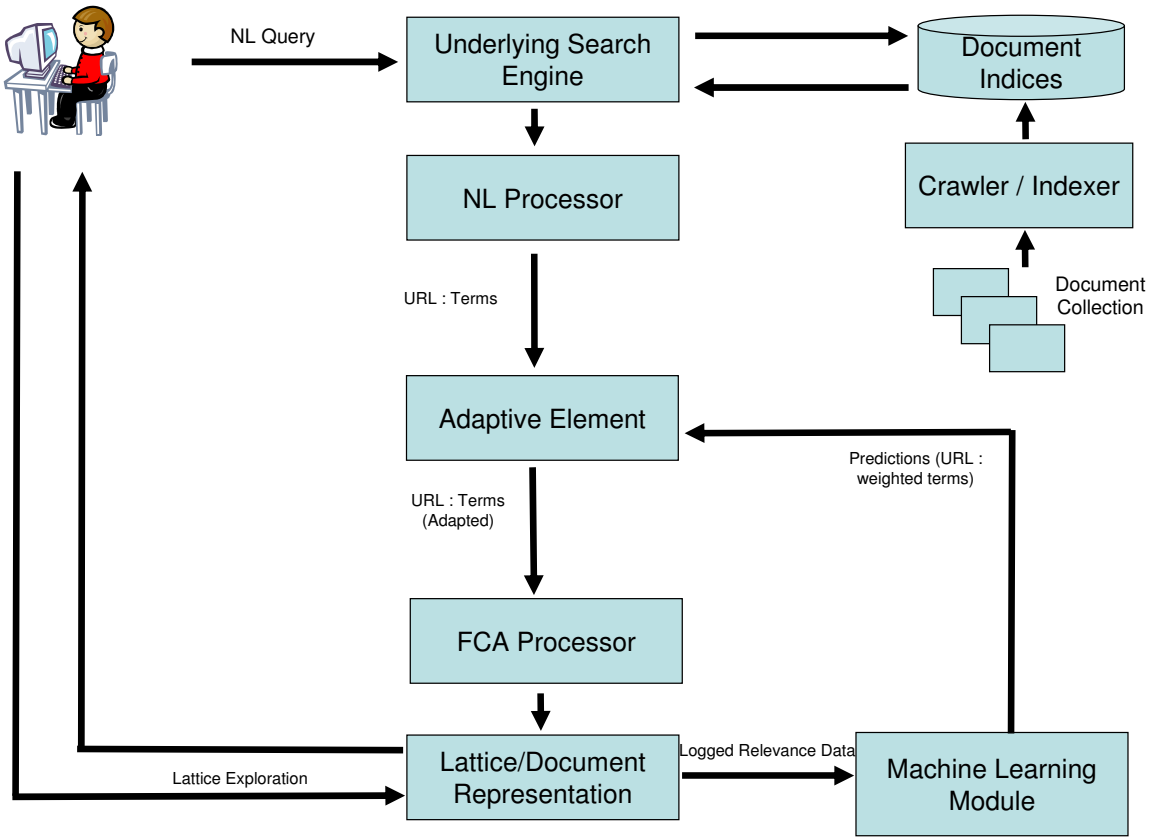
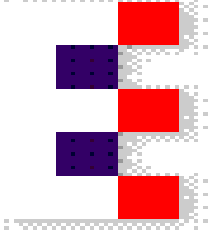


Figure: System Architecture



Adaptively Modelling the Context of an Intranet Query

Motivation
Research Plan
FCA
Hasse Table
Related Concept Lattice
Automade
Adapted Lattice
Machine Learning

	horse	**male**	**female**	**adult**	**young**
horse	X				
stallion	X	X		X	
mare	X		X	X	
foal	X				X
filly	X		X		X
colt	X	X			X

Figure: Classical Lattice Example - Hasse Table

Motivation
Research Plan
FCA
Hasse Table
Related Concept Lattice
Automade
Adapted Lattice
Machine Learning

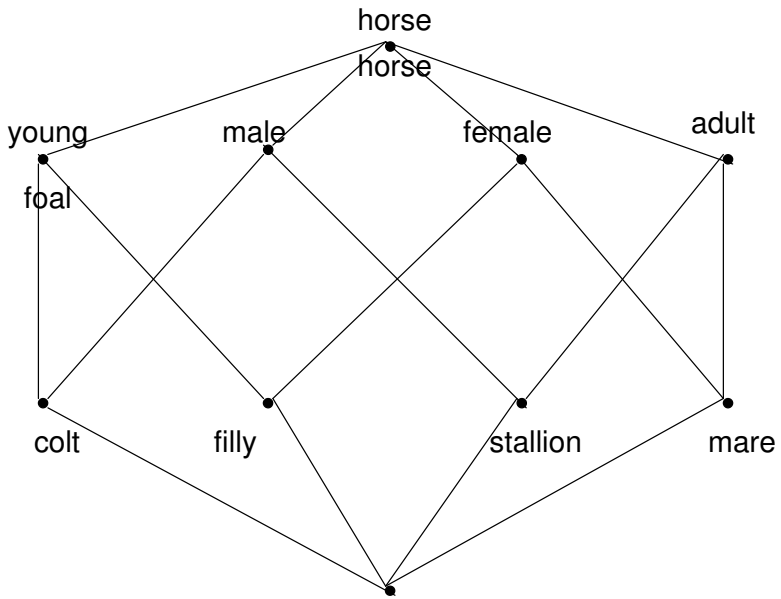


Figure: Classical Lattice Example - Concept Lattice

Adaptively Modelling the Context of an Intranet Query

Motivation

Research Plan

FCA

Hasse Table

Related Concept Lattice

Automade

Adapted Lattice

Machine Learning

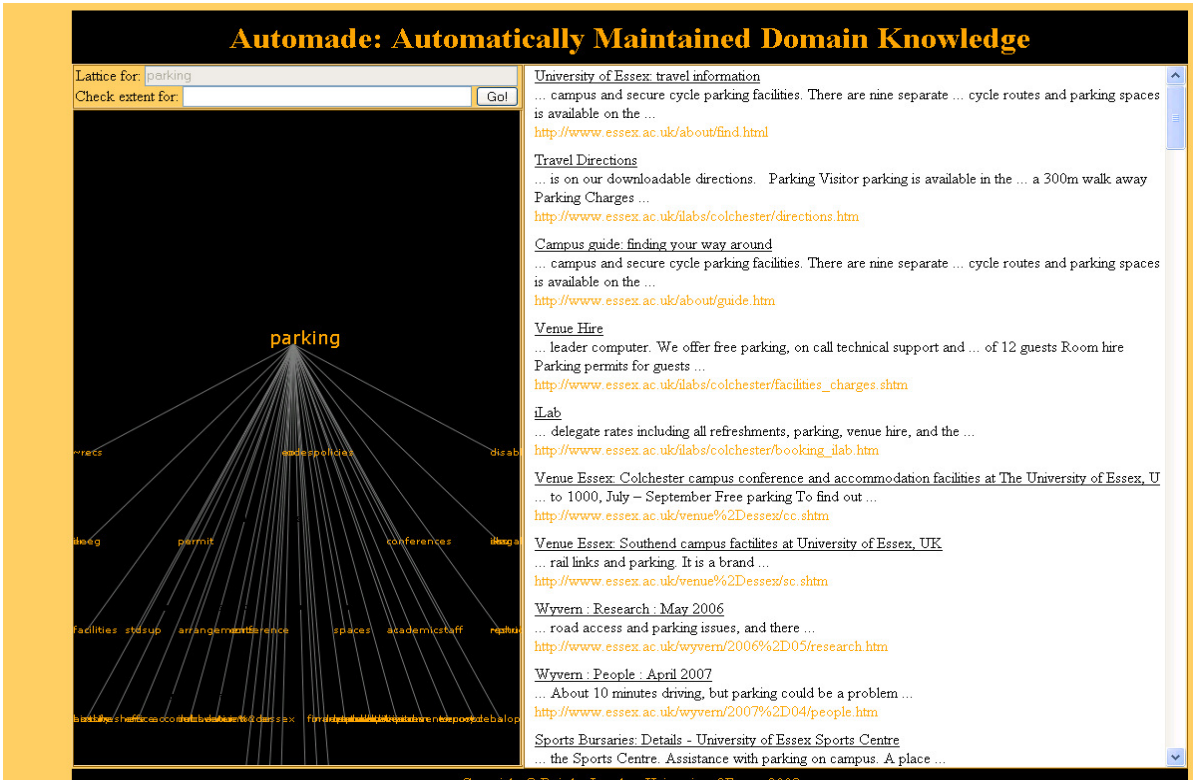


Figure: Automade Screenshot

Adaptively Modelling the Context of an Intranet Query

Motivation
Research Plan
FCA
Hasse Table
Related Concept Lattice
Automade
Adapted Lattice
Machine Learning

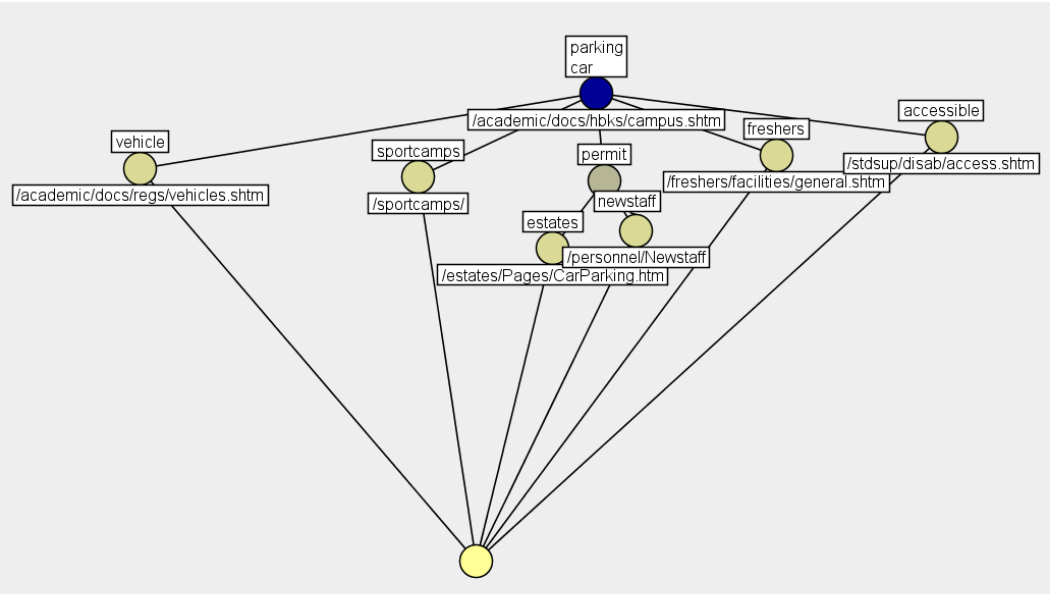
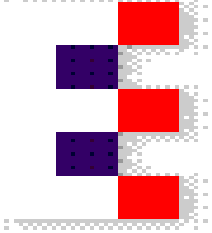


Figure: Example Adapted Lattice



Adaptively Modelling the Context of an Intranet Query

Motivation
Research Plan
FCA
Machine Learning
SVM-Light
Clickthrough Data
Adaptation Steps

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?

Motivation
Research Plan
FCA
Machine Learning
SVM-Light
Clickthrough Data
Adaptation Steps

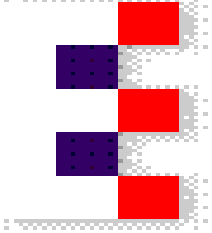
Clickthrough Data

- 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

Motivation
Research Plan
FCA
Machine Learning
SVM-Light
Clickthrough Data
Adaptation Steps

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



Motivation
Research Plan
FCA
Machine Learning
SVM-Light
Clickthrough Data
Adaptation Steps

Thank You!