

Explorations in Bootstrapping Guided Search

8th Language and Computation Day

Deirdre Lungley
dmlung@essex.ac.uk

October 8, 2009

Explorations in Bootstrapping Guided Search

Research Contribution

- 1 Automatically acquire a domain model for a document collection

Explorations in Bootstrapping Guided Search

Research Contribution

- 1 Automatically acquire a domain model for a document collection
- 2 Allow for user adaptation through the incorporation of log data

Explorations in Bootstrapping Guided Search

Research Contribution

- 1 Automatically acquire a domain model for a document collection
- 2 Allow for user adaptation through the incorporation of log data
- 3 Provide an insight into the different nature of general search, e.g., WWW search versus intranet search

Explorations in Bootstrapping Guided Search

Methodology

- Formal Concept Analysis (FCA) lattice based domain model
 - Navigational qualities
 - Coatoms provide initial query refinement suggestions

Explorations in Bootstrapping Guided Search

Methodology

- Formal Concept Analysis (FCA) lattice based domain model
 - Navigational qualities
 - Coatoms provide initial query refinement suggestions
- Deriving lattice document descriptors (index terms)
 - Lattice structure dependant on good document descriptors
 - Use combination of NLP and mining of query logs

Explorations in Bootstrapping Guided Search

Methodology

- Formal Concept Analysis (FCA) lattice based domain model
 - Navigational qualities
 - Coatoms provide initial query refinement suggestions
- Deriving lattice document descriptors (index terms)
 - Lattice structure dependant on good document descriptors
 - Use combination of NLP and mining of query logs
 - NLP techniques:
 - Noun phrase terms which occur in at least 2 contexts are included.
 - Also extract terms which co-occur with query term(s)

Explorations in Bootstrapping Guided Search

Methodology

- Formal Concept Analysis (FCA) lattice based domain model
 - Navigational qualities
 - Coatoms provide initial query refinement suggestions
- Deriving lattice document descriptors (index terms)
 - Lattice structure dependant on good document descriptors
 - Use combination of NLP and mining of query logs
 - NLP techniques:
 - Noun phrase terms which occur in at least 2 contexts are included.
 - Also extract terms which co-occur with query term(s)
 - Query log mining:
 - Machine learning through relative relevance
 - Learn the URLs relevant to a query term(s)
 - Attach query term(s) to these URLs

Explorations in Bootstrapping Guided Search

Early Interactive Intranet Experiment¹

- Simulate log data transactions for some frequent queries

¹Lungley, D. and Kruschwitz, U., Automatically Maintained Domain Knowledge: Initial Findings. In proceedings of the 31st European Conference on IR Research, ECIR 2009

Explorations in Bootstrapping Guided Search

Early Interactive Intranet Experiment¹

- Simulate log data transactions for some frequent queries
- Evaluate generated query refinement suggestions over two baselines:
 - Lattice based solely on text processing of documents
 - Frequent terms

¹Lungley, D. and Kruschwitz, U., Automatically Maintained Domain Knowledge: Initial Findings. In proceedings of the 31st European Conference on IR Research, ECIR 2009

Explorations in Bootstrapping Guided Search

Early Interactive Intranet Experiment¹

- Simulate log data transactions for some frequent queries
- Evaluate generated query refinement suggestions over two baselines:
 - Lattice based solely on text processing of documents
 - Frequent terms
- Results:

	Adapted Lattice	B1:Unadapted Lattice	B2:Frequent Terms
% suggestions judged relevant	73%	32%	42%

¹Lungley, D. and Kruschwitz, U., Automatically Maintained Domain Knowledge: Initial Findings. In proceedings of the 31st European Conference on IR Research, ECIR 2009

Explorations in Bootstrapping Guided Search

Early Interactive Intranet Experiment¹

- Simulate log data transactions for some frequent queries
- Evaluate generated query refinement suggestions over two baselines:
 - Lattice based solely on text processing of documents
 - Frequent terms
- Results:

	Adapted Lattice	B1:Unadapted Lattice	B2:Frequent Terms
% suggestions judged relevant	73%	32%	42%

- Results confirm our assumption that users would prefer query refinement suggestions learnt from user queries over content generated terms

¹Lungley, D. and Kruschwitz, U., Automatically Maintained Domain Knowledge: Initial Findings. In proceedings of the 31st European Conference on IR Research, ECIR 2009

Explorations in Bootstrapping Guided Search

World Wide Web Bootstrapping Experiment

- MSN Search Asset Data Collection
- 15 million queries and related clicks

Explorations in Bootstrapping Guided Search

World Wide Web Bootstrapping Experiment

- MSN Search Asset Data Collection
- 15 million queries and related clicks
- TREC topics, 1 low frequency, 3 medium and 6 high

Explorations in Bootstrapping Guided Search

World Wide Web Bootstrapping Experiment

- MSN Search Asset Data Collection
- 15 million queries and related clicks
- TREC topics, 1 low frequency, 3 medium and 6 high
- Results of UK evaluation:

	Adapted Lattice	B1:Unadapted Lattice	B2:Noun Count
% suggestions judged relevant	61%	63%	59%

Explorations in Bootstrapping Guided Search

World Wide Web Bootstrapping Experiment

- MSN Search Asset Data Collection
- 15 million queries and related clicks
- TREC topics, 1 low frequency, 3 medium and 6 high
- Results of UK evaluation:

	Adapted Lattice	B1:Unadapted Lattice	B2:Noun Count
% suggestions judged relevant	61%	63%	59%

- Results of Mechanical Turk evaluation:

	Adapted Lattice	B1:Unadapted Lattice	B2:Noun Count
% suggestions judged relevant	67%	69%	64%

Explorations in Bootstrapping Guided Search

Observations

- Can we say deriving suggestions from logs works better on intranet data?

Explorations in Bootstrapping Guided Search

Observations

- Can we say deriving suggestions from logs works better on intranet data? Influencing factors:
 - Limitation to simple term pair evaluation - WWW requires more context
 - Temporal dimension - log data dated May 2006

Explorations in Bootstrapping Guided Search

Observations

- Can we say deriving suggestions from logs works better on intranet data? Influencing factors:
 - Limitation to simple term pair evaluation - WWW requires more context
 - Temporal dimension - log data dated May 2006
- Can we say deriving suggestions from historic queries works better than from historic queries and clicks?

Explorations in Bootstrapping Guided Search

Observations

- Can we say deriving suggestions from logs works better on intranet data? Influencing factors:
 - Limitation to simple term pair evaluation - WWW requires more context
 - Temporal dimension - log data dated May 2006
- Can we say deriving suggestions from historic queries works better than from historic queries and clicks? Useful since:
 - Query data more readily available
 - Sensitive nature of click data

Explorations in Bootstrapping Guided Search

Observations

- Can we say deriving suggestions from logs works better on intranet data? Influencing factors:
 - Limitation to simple term pair evaluation - WWW requires more context
 - Temporal dimension - log data dated May 2006
- Can we say deriving suggestions from historic queries works better than from historic queries and clicks? Useful since:
 - Query data more readily available
 - Sensitive nature of click data
- Suggests evaluation of query-only adaptation
 - Intranet experiment
 - Adapt relative relevance learning
 - Highly dependant on good precision (P@1/P@2/P@5)
 - Nutch (VSM) to Lucene (BM25F)

Explorations in Bootstrapping Guided Search

Deriving query suggestions from Intranet Query Logs using MLE

- Research Questions:
 - Usefulness of dialogue log component
 - Suitability of Web derived suggestions for domain-specific search
 - General Web user perception of "usefulness" of extracted suggestions

Explorations in Bootstrapping Guided Search

Deriving query suggestions from Intranet Query Logs using MLE

- Research Questions:
 - Usefulness of dialogue log component
 - Suitability of Web derived suggestions for domain-specific search
 - General Web user perception of "usefulness" of extracted suggestions
- Query bigram MLE – $\max P(q_{n+1}|q)$ over (q, q_{n+1})

Explorations in Bootstrapping Guided Search

Deriving query suggestions from Intranet Query Logs using MLE

- Research Questions:
 - Usefulness of dialogue log component
 - Suitability of Web derived suggestions for domain-specific search
 - General Web user perception of "usefulness" of extracted suggestions
- Query bigram MLE – $\max P(q_{n+1}|q)$ over (q, q_{n+1})
- Experimental setup:
 - Suggestions generated for top 25 most frequently submitted queries
 - 67 participants for both evaluations

Explorations in Bootstrapping Guided Search

Deriving query suggestions from Intranet Query Logs using MLE

- Research Questions:
 - Usefulness of dialogue log component
 - Suitability of Web derived suggestions for domain-specific search
 - General Web user perception of "usefulness" of extracted suggestions
- Query bigram MLE – $\max P(q_{n+1}|q)$ over (q, q_{n+1})
- Experimental setup:
 - Suggestions generated for top 25 most frequently submitted queries
 - 67 participants for both evaluations

Method	Relevant – Local	Relevant – MT
MLE-Session	71.0%	63.6%
MLE-Dialogue	75.7%	68.9%
MLE-Dialogue-Add	72.1%	63.6%
MLE-Dialogue-Replace	75.2%	73.1%
Baseline-Snippets	54.9%	51.3%
Baseline-Google	35.6%	58.3%

Explorations in Bootstrapping Guided Search

Going Forward

- Revisit lattice document descriptors
 - Move from "Related searches" to "concepts"
 - Conceptual representation to map a specific URL into some space
 - Latent Semantic Analysis (LSA) kernel

Explorations in Bootstrapping Guided Search

Questions?

Explorations in Bootstrapping Guided Search

Automade: Automatically Maintained Domain Knowledge

Letter for:

Check extent for:

parking

- permit
- car
- events
- vehicle

Calendar of the University of Essex
... of the University P Parking and Driving of ... the Driving and Parking of Vehicles within ...
<http://www.essex.ac.uk/academic/docs/itemap.shtml>

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

Information for new staff
... obtain a car parking permit from the Security ...
<http://www.essex.ac.uk/personel/newstaff/default.htm>

Freshers 2007: Travel Information - Getting around once you've arrived
... committed to decreasing car parking demand on campus and ... To register for parking, please go to the
<http://www.essex.ac.uk/freshers/travel/gettingaround.shtml>

Calendar of the University of Essex - Student Handbook - The Campus
... top of page | Car Parking Students resident in campus ... The Quays, are prohibited from parking a vehicle on ...
<http://www.essex.ac.uk/academic/docs/hbbs/campus.shtml>

University of Essex: Conference Office: Frequently asked questions
... to departure. ? Is car parking free of charge ... supplied with a parking permit for the ...
<http://www.essex.ac.uk/conference/faq.shtml>

Calendar of the University of Essex - Rules Governing the Driving and Parking of Vehicles within Uni
... when the car parking charge is paid and ... does not guarantee that car parking places, in accordance ...
<http://www.essex.ac.uk/academic/docs/regsvehicles.shtml>

Calendar of the University of Essex - Academic Regulations, Regulations relating to Conduct
... the Driving and Parking of Vehicles within ...
<http://www.essex.ac.uk/academic/docs/regsvconduct.shtml>

Communications Office at the University of Essex - Organising an event
... Dumbrell to discuss special parking arrangements. You will also need ...
<http://www.essex.ac.uk/comm/events/organising.htm>

Figure: Automade - UoE Intranet