Recommending Experts and Scientific Articles

Toine Bogers

Research talk @ RSLIS, Copenhagen

June 12, 2008

Tilburg University

Outline

- About me
- Expert search & recommendation
- Recommending scientific articles

About me

- Education
  - **1997-2001** Master's degree in Information Management & Technology
  - **2002-2004** Master's degree in Computational Linguistics & Artificial Intelligence
- Employment
  - **2005-now** PhD student in the A Propos project about pro-active document recommendation
- Teaching
  - **2006-now** various guest lectures about search engines and IR
  - **2007** Information Search, Retrieval, and Recommendation
  - **2008** Information Search, Retrieval, and Recommendation

About me

authoritative re-ranking blogs citation analysis clustering collaborative filtering collaborative search context data fusion deduplication distributed IR document recommenders evaluation expertise expert search geographic IR group dynamics humor image retrieval IMAs index size estimation information extraction information filtering information interaction information retrieval information seeking interactive IR language modeling machine learning manuals metrics networks NLP passage retrieval predictive text entry psychology query log analysis recommender systems retrieval smoothing social bookmarking social tagging sociology spam statistics test collections text categorization user interface design user modeling web crawling web search

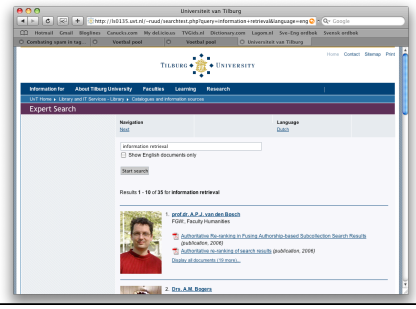
Tilburg University

Outline

- About me
- Expert search & recommendation
  - Definition & history
  - Tasks & approaches
  - A look at evaluation & test collections
  - Expertise seeking
  - A university-wide expert search engine
- Recommending scientific articles

What is it?

- Basically: searching for experts instead of documents



History of expert search

- In 80's and 90's
  - Implemented as large-scale databases containing employee skills
  - Problems
    - Puts the workload on employees
    - 'Unnatural' approach
    - Easily out-of-date
- TREC 2005 Enterprise Track introduced the *Expert Search Task*
  - Large-scale evaluation effort of expert finding
    - 2005 & 2006: W3C collection
    - 2007 & 2008: CSIRO collection
  - Huge boost in research into automatic approaches
  - Usually co-occurrence of people and topics is seen as *evidence* of expertise

Evidence of expertise

- Content-based evidence
  - Documents
  - E-mails
  - Homepages
- Evidence from social networks
  - Organizational structure
  - E-mail networks
  - Bibliographic information
- Activity-based evidence
  - Project time
  - Search history
  - Publication history

Tasks and approaches

- Different tasks
  - Expert finding
    - Find the experts on a specific topic
  - Expert profiling
    - Find out what one expert knows about different topics
  - Recommending similar experts
    - Find experts who share the same profiles

Evaluation

- Majority of work is evaluated using TREC collections
  - W3C collection
    - 5.7 GB and 331,037 documents (Web pages, mailing lists, project pages)
    - Topics are group names
    - Relevance judgments
      - 2005: group members are experts
      - 2006: TREC participants judge expertise themselves
  - CSIRO collection
    - 4.2 GB and 370,715 documents (similar diversity as W3C)
    - Work tasks created by actual CSIRO science communicators
      - Goal is to create an overview page on a certain topic
    - Relevance judgments done by science communicators in 2007 and 2008

UvT Expert Collection

- Problems with TREC collections
  - Expertise is never self-assessed
  - Only one specific type of organization
  - Only in English
- We therefore created the UvT Expert Collection
  - Crawl of a medium-sized Dutch university
  - Based on Webwijs ("*Webwise*"), our online expert profiling database
    - 1168 experts
    - 1400 *self-assessed* expertise topics
    - Bilingual (Dutch and English)
  - Documents include publications, course pages, research descriptions, and homepages
  - Information about organizational structure and topic hierarchy
  - See SIGIR '07 paper for more information

Expertise seeking

- All expert finding work so far has been from an IR perspective
  - What is missing is an IS perspective: *expertise seeking*
- What we did to remedy this
  - Focused on the task of *recommending similar experts*
    - Scenario sketch: "*The media wishes to communicate with the top expert, but he is unavailable for a while. Who would you recommend to take their place?*"
  - Got 6 of our university's communication advisors to participate in our study
  - Two-fold purpose of our questionnaire
    - Investigate expertise seeking behavior
    - Get realistic relevance judgments for the 'similar experts'-task
      - Had to judge 10 recommended experts for 10 familiar 'focus' experts
  - See SIGIR '08 workshop paper for more information

Expertise seeking

- Investigate expertise seeking behavior
  - Inspired by 2007 IP&M paper by Woudstra and Van den Hooff
    - Identified 11 important factors for source selection (*topic of knowledge*, *familiarity*, *reliability*, *availability*, *perspective*, *up-to-dateness*, *approachability*, *cognitive effort*, *contacts*, *physical proximity*, *saves time*)
  - Asked participants to describe
    - Typical requests for expertise
    - Reasons for picking and *not* picking specific experts
    - How important each factor was for their decisions
- Some findings
  - *Topic of knowledge* was most important in recommending someone
  - *Familiarity* with the expert was also important
  - New factors we identified
    - Organizational structure (professors and project leaders are preferred)
    - Media experience ("*one of them is not suitable for talking to the media*")

Expertise seeking

- Get realistic relevance judgments
  - Used 44 unique focus experts divided over the 6 PR advisors (10 each)
  - First, participants were asked for their own suggestions
  - Generated 10 recommended experts for each using system pooling
  - Participants then ranked these suggested experts on a 10-point scale
- Integrated the factors into expert finding models
  - Evaluated using MRR and NDCG@10
- Some findings
  - Best baseline approach combined terms from documents with the self-assessed expertise areas
  - Integrated the following factors into retrieval models: *organizational structure*, *media experience*, *reliability*, *up-to-dateness*, quality of *contacts*
  - Significant improvements using *reliability*, *up-to-dateness*, and *organizational structure*

A university-wide expert search engine

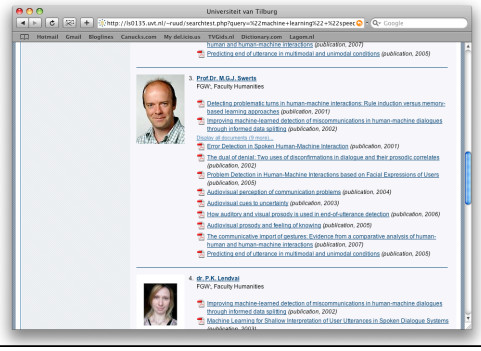
- Work in progress by Master's student Ruud Liebregts
  - Designing and evaluating a university-wide expert search engine
- Design
  - Data sources include publications, theses, course descriptions, research descriptions, self-assessed expertise areas
  - Allows for filtering on language and faculty
  - Shows collaboration networks for papers and thesis supervision
- Evaluation
  - System-based evaluation
    - 240 test topics
      - 120 Dutch and 120 English
      - 120 based on thesis supervisors and 120 based on paper authors
    - Gold standard judgments from user-based evaluation (see next slide)

A university-wide expert search engine

- Evaluation (cont'd)
  - User-based evaluation
    - 30 employees asked to
      - Describe one of their expertise areas
      - List and rank 5 possible experts
      - Formulate a query based on the topic
      - Judge the top 10 search engine results
    - ±30 UvT students will be randomly assigned
      - 3 out of 5 possible expert finding work tasks
      - 3 out of 5 possible supervisor finding work tasks
      - Comparing the new search engine vs. everything else the university has to offer
      - Using the baseline 3 times and the new search engine 3 times
    - Possibly
      - ±30 people external to Tilburg University
      - Two classes of ±50 Dutch high school seniors

A university-wide expert search engine

A university-wide expert search engine

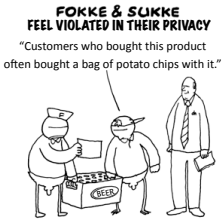


Outline

- About me
- Expert search & recommendation
- Recommending scientific articles
  - What is it?
  - Approaches
  - Social bookmarking
  - Recommending using CiteULike
  - At the library school
  - Future work

What is it?

- Formal definition
  - A recommender system tries to identify sets of items that are likely to be of *interest* to a certain user given some information from that user's profile.
- More casual definition



Approaches to recommendation

- Some popular approaches
  - Most popular item (non-personalized recommendation)
  - Demographic recommendation (uses user features)
  - *Content-based* recommendation (using IR models)
    - Use item features to find items similar to past items
    - Good content match between items, but no quality control
  - *Collaborative filtering* (mining usage patterns)
    - Uses user-item preferences (e.g. explicit ratings data, purchase data)
    - Good for areas where content analysis is hard (e.g. movies, music)
    - Two types
      - User-based filtering
      - Item-based filtering

Approaches to recommendation

- User-based filtering (find similar users)
- Item-based filtering (find similar items)

items

users	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
U_1	5		8		4		10		7	8
U_2	10				1		6			
U_3	2	3		10		4			9	9
U_4			7		3				9	10
U_5	1	5								
U_x	2			9		5			10	?

Social bookmarking

- Way of storing, organizing, and managing bookmarks of Web pages, scientific articles, books, etc.
  - Users can add bookmarks
  - Can be made public or kept private
  - Often allow users to tag/describe their item
  - Lots of social bookmarking services available

Research so far

- There are golden opportunities here!
  - Tons of free, useful data
    - Large amounts of content described using tags and other metadata
    - Users reveal information about themselves by adding and tagging items
    - Treasure trove of user-item preferences
  - Can be used to predict new items
- However, research still in its infancy
  - Mostly exploratory and theoretical
  - Some scattered attempts at improving IR using tags
  - Recommendation for social bookmarking
    - Mostly tag recommendation (easy to evaluate)
    - And ofcourse there's StumbleUpon

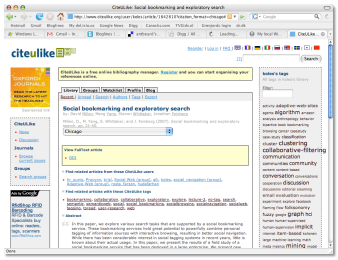
Main focus

- My main focus
  - Recommending interesting bookmarks based on user profiles from social bookmarking websites
  - Experiment with different
    - Algorithms
    - Contextual representations
    - Aspects (temporal, growth curves, spam, duplicates)
    - Combinations of approaches (data fusion)
  - Evaluation
    - System-based evaluation
    - User-based evaluation
  - Preferably for two different areas
    - Scientific articles (CiteULike, Bibsonomy)
    - Web pages (Delicious)

BibSonomy citeulike del.icio.us

citeulike

- Social bookmarking for scientific papers
  - According to their website, it is "*a free service to help you to store, organise, and share the scholarly papers you are reading*"
  - Some features
    - Article metadata
    - Tagging
    - Groups
    - Comments
    - Reading priorities
    - Batch importing



citeulike

- Creating a collection
  - Daily database dumps available
    - Contain user-item-tag triples with timestamps
    - But none of the additional information available on the website
  - Used the November 2, 2007 dump as a starting point
  - Crawled the rest of the website
    - Article and user metadata
    - Group information
    - Reading priorities
- Some statistics
  - 803,521 items (metadata available for 67%)
  - 25,375 users (29% spam profiles)
  - 232,937 tags

Experimental setup & evaluation

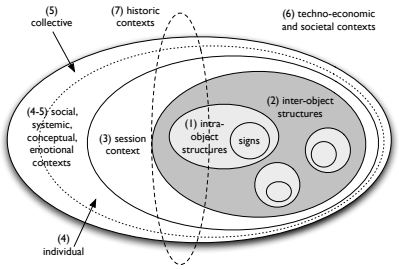
- System-based evaluation
  - We know what papers a user liked from his profile
    - How well can we predict what we already know?
    - User profiles we have are user-item pairs
  - Formal setup
    - Take out 10 items from each user profile
    - Train on remaining profile, predict missing items
    - Users with ≥ 20 items and articles added at least twice
    - 10-fold cross-validation to prevent overfitting
  - Evaluation
    - If we recommend the missing items, that's good!
    - MAP, MRR, Precision @ 10, user coverage
  - We can use this same setup for all experiments

At the library school

- First experiments using collaborative filtering
  - Best model has a MAP of 0.2478 and similar P@10
  - User-based filtering performed best
    - Optimal number of neighbors was 5
  - User coverage is high at 99.6%
    - For how many users can we predict something?
    - Some users too new or eclectic
  - Difficult task because of high sparsity (99.98%)
    - MAP of 1.0 not necessarily achievable (or realistic)
  - Performance okay, but room for improvement

At the library school

- What context do we have in CiteULike?



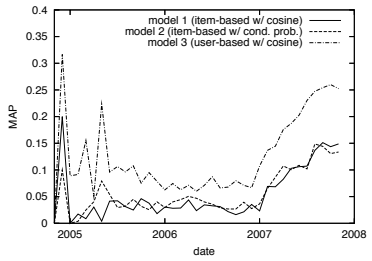
Ingwersen and Järvelin (2006)

At the library school

- What context do we have in CiteULike?
  - (1) **Intra-object structures** Properties of the documents themselves, such as article metadata and the abstract (available 33% of the time)
  - (2) **Inter-object structures** Relations between documents, such as those available through authorship information, assigned tags, and inclusion by users.
  - (4-5) **Social, systemic, conceptual, and emotional contexts** The folksonomy can represent social, conceptual, and emotial context. The information about the groups and the usage patterns are all social context for the recommender.
  - (7) **Historical contexts** Activity logs allow for, for instance, temporal analysis.

At the library school

- Birth of a recommender
  - When is a social bookmarking website big enough for recommendations?



At the library school

- No de-duplication by CiteULike upon entry
- Many duplicates
  - Early estimates of around 10% (on manually annotated testset)
    - Mismatches on title, year, authors, etc.
  - With 20% of those articles having over 20 duplicates

```
Collective dynamics of "small-world" networks     Collective dynamics of 'small world' networks
Collective dynamics of 'small-world' networks     Collective dynamics of 'small-world' networks
Collective dynamics of 'small-world' networks     Collective dynamics of 'small-world' networks
Collective dynamics of 'small-world' networks     Collective dynamics of 'small-world' networks.
Collective dynamics of 'small-world' networks.    Collective dynamics of 'small-world' networks.
Collective dynamics of 'small-world' networks.    Collective dynamics of `small-world'' networks
Collective dynamics of ``small-world'' networks   Collective dynamics of `small-world' networks
Collective dynamics of `small-world' networks     Collective dynamics of `small-world' networks
Collective dynamics of `small-world' networks     Collective dynamics of `small-world' networks.
Collective dynamics of small-world networks       Collective dynamics of small-world networks
Collective dynamics of small-world networks       Collective dynamics of small-world networks
Collective dynamics of small-world networks.      Collective dynamics of small-world networks.
Collective dynamics ofsmall-worldnetworks.        Collective dynamics ofsmall-worldnetworks.
```

Future work

- Experiment with different
  - Algorithms
  - Contextual representations
  - Aspects (temporal, growth curves, spam, duplicates)
  - Combinations of approaches (data fusion)
  - Datasets
- User-based evaluation
  - Pick one or two tasks in paper recommendation defined by McNee (2006)
    - Maintain awareness
    - Explore research interest
    - Find more like this
  - Evaluate algorithms using users
    - Based on their actual profile
    - By simulating one of these recommender tasks

Questions? Comments? Suggestions?