Something about IR in the age of the web and social media

Jaap Kamps
University of Amsterdam

16 October 2014
Fall EARIA School, Notre Dame de Parmenie, France
Outline

- IR in a changing world
- Classic IR and revolutionary changes
- Challenges to IR and how to tackle them
- Case study: Social Book Search
- Changing collections, tasks, users, …
- Toward rich task/user context
IR at age 50-something...

Part I
“Modern search” is pervasive but has a long history
AMERICAN DOCUMENTATION 1955

A QUARTERLY REVIEW OF IDEAS, TECHNIQUES, PROBLEMS AND ACHIEVEMENTS IN DOCUMENTATION

VOL. VI, NOS. 1-4 JANUARY - OCTOBER 1955
Interpreted literally this time-honored oath of the courts confronts the witness with a humanly impossible task. No one could be expected to perceive all the elements—all the minute detail, important or otherwise—that comprise a given incident, event, or situation. The best that one may expect is that the really important will be recognized, recorded in the memory, and recalled with accuracy when the occasion demands. As every lawyer knows, even the most honest and scrupulous witnesses will swear to contradictory accounts and disagree among themselves concerning relatively obvious detail.

So too, honest differences of opinion respecting the relative importance of the components of subject content have initiated extensive and bitter controversy among those who have devised or espoused classification schemes that follow a traditional pattern. The division and subdivision of areas of knowledge into rigid compartmentation is possible only by the identification of lines of demarcation which must, in their definition and location, be, in the last analysis, little more than arbitrary. Failure to recognize this basic fact has led logicians into error and classifiers into strife.

But classification unencumbered by the arbitrary delineation of knowledge into rigid compartmentation is now possible through the use of automatic, or semi-automatic mechanisms—machines which give to classification a new flexibility and elasticity through the ease with which entirely new categories, classes, or composite groupings may be generated whenever a particular situation or need so demands. Selection based on combinations, any combinations, of characteristic attributes is a fundamental property of documentation systems utilizing automatic or semi-automatic mechanisms (mechanical aids). Thus with properly designed equipment the full range of combinations permitted by the logical theory of class definitions becomes possible in making selections.

Nevertheless, documentation systems predicated upon logical definitions—

Jesse H. Shera comments on the indexing debate

the attributes and properties (characteristics) of the subject content of a given document or store of documents might well outweigh any anticipated future advantage. Thus does the documentalist become as much economist as logician, and the laws of logical division yield before the law of diminishing returns.
which give to classification a new flexibility and elasticity through the ease with which entirely new categories, classes, or composite groupings may be generated whenever a particular situation or need so demands. Selection based on combinations, any combinations, of characteristic attributes is a fundamental property of documentation systems utilizing automatic or semi-automatic mechanisms (mechanical aids). Thus with properly designed equipment the full range of combinations permitted by the logical theory of class definitions becomes possible in making selections.

Nevertheless, documentation systems employing such mechanical aids are also subject to the limitations imposed by human inability to tell “the whole truth.” But even were the machines able to transcend this human limitation the expense involved in recording for subsequent searching all the attributes and properties (characteristics) of the subject content of a given document or store of documents might well outweigh any anticipated future advantage. Thus does the documentalist become as much economist as logician, and the laws of logical division yield before the law of diminishing returns.

One example may make our point clear. A mechanized system may prove entirely satisfactory when applied to a small collection of documents but break down completely under the burden of a more extensive and complex load. Conversely, a more elaborate system, quite capable of handling efficiently large masses of diverse materials, may be disappointingly ineffective when tested upon a small collection. One does not need a wheel to break a butterfly! Thus the most exact and careful of measurements can prove misleading when the efficiency of a system is strongly influenced, as it so often is, by the volume of documents that it is called upon to handle. Cautious and searching evaluation of all experimental results is essential in rating the efficiency of documentation systems.

May the age-old controversies that arose from the conventional concepts of classification not be reborn in the mechanized searching systems of the future. There is hope for the avoidance of such error if we will but regard documentation systems as useful devices the benefits of which must be determined, not by polemics but by the intelligent measurement of such benefits in relation to needs and costs. The machines of the future can make us free, but only if we are willing to subject them, and ourselves, to the most rigid intellectual discipline.

Shera asks for proper scientific evaluation (1955)
Picked up by a Librarian at Cranfield
Cleverdon develops his evaluation method (1960-)

REPORT ON THE TESTING AND ANALYSIS OF AN INVESTIGATION INTO THE COMPARATIVE EFFICIENCY OF INDEXING SYSTEMS

by

Cyril W. Cleverdon

An investigation supported by a grant from The National Science Foundation Washington
1960s: Automatic Indexing can be as effective as Manual
Since 1960 much development of search technology

Evaluation stayed the same...
Just need a corpus, a bunch of topics, some judgements

Text REtrieval Conference (TREC)
...to encourage research in information retrieval from large text collections.

Overview
Publications
Information for Active Participants
Tracks
Past TREC Results
Data
Contact Information

Other Evaluations
Frequently Asked Questions

Still dominates academia and industry
What’s the problem?
Underlying “User Model” did not change

Still *library consultation*: static collection, dynamic requests
Predominantly one-step-search (batch, no interaction)

Isolated search ignoring the actual use of the information
Is this the right model in the “always online” world?
Our infrastructure changed in a revolutionary way.
Our technology changed in a revolutionary way
How radical did information access methods change?
Information → User
Daily content published, by number of pieces

Facebook content
1.5 billion pieces

Tweets
140 million

Tumblr posts
10 million

Blog posts
1.6 million

Images (Flickr)
5 million

New websites
60,000

Videos (YouTube)
2 million
Classic IR: Query → Doc
Evaluation Criteria:

A: Recall, precision, efficiency, quality of information/process

B: Usability, quality of information/process

C: Quality of info & work process/result

D: Socio-cognitive relevance; quality of work task result

Work task context

Seeking context

Socio-organizational & cultural context

Query++ → Doc
Query++ → Hitlist → Doc++

Open Source Software Development: An Overview
Ming-Wei Wu, Ying-Dar Lin

article[1]
  bdy[1]
  sec[2] (OPEN SOURCE BACKGROUND)
  ... for work on the GNU Project (http://www.gnu.org/gnu/thegnu.html). GNU is a recursive acronym for "GNU's Not Unix" and a homophone for "new." The GNU Project seeks to develop ... principal author of the GNU C Compiler (GCC), GNU symbolic debugger (GDB), GNU ... in essence, copyrights with GPL regulations. Open source software, essentially ...

article[1]
  bdy[1]
  sec[4] (LICENSING MODELS)
  p[5]
  ... code falls under the GPL model, nVidia had to remove all GPL code, then re-release the ... Softway Systems, makers of GPL-regulated software, and repackaged its ... work, thereby skirting the GPL regulation.

article[1]
  bdy[1]
  sec[5] (BUSINESS MODELS)
  p[6]
  ... multilicensing models to avoid GPL violations. For example, Sun's ... adapts three licensing models: GPL, LGPL, and SISSL. Ordinary users who can fulfill GPL regulations can use StarOffice under GPL. Proprietary developers and companies can use LGPL or SISSL, which both ...
Query++ → Query Suggest → Hitlist → Doc++
Modern Web data is highly structured resulting in a massive multidimensional graph

- Entities from massive knowledge resources, geo-temporal references, social network structure, ...

- E.g. Facebook: entities, relations between entities, locations and timestamps, social network, https://www.facebook.com/about/graphsearch

Related Work: Facebook Graph Search
Graph Search

- Unifies the *searcher and the information resources* that played fundamentally different roles in classic IR
- Searching information from a personal point of view — *you are the query!* — extreme personalization
- Highly structured and curated information space
- Beyond 2-term queries, beyond 10 blue links
- More interactive session
SERP vs IQEP: Incremental Query Exploration Page
Interaction Stages (1)

- **Incremental Structured Query Formulation**
  - Dynamic incremental query suggestion
  - Previews and surrogates from SERP moving to query suggestion
  - Generic entity types and instances (based on your activity)
Interaction Stages (II)

- Dynamic Structured Result Set Exploration
- Unique for you and a point in time
- Not a top-10 anymore; faceted exploration based on your interests
- Far more dynamic structure than classic facet hierarchies
Evaluation

• How to evaluate the overall process? How strict are both stages divided?

• How to evaluate the first stage (query formulation)? How well the query expresses the need? How well the system promotes verbose queries? …

• How to evaluate the second stage (explore and exploit the result set)? How many filters are used? What is the engagement? …
Wrap Up (I)

• IR in a changing world
  • Core IR problems as relevant as ever!
  • Classic IR has value, but didn’t keep up…

• Exciting new opportunities!
  • User/task context; Structured/heterogeneous data; Novel access tools (search stage sensitive)
  • Classic dichotomy of user and information broken
  • User-centered and system-centered IR reunited
Richer theories, richer experiments

Stephen Robertson
Microsoft Research Cambridge and
City University
ser@microsoft.com
A caricature

On the one hand we have the Cranfield / TREC tradition of experimental evaluation in IR
  – a powerful paradigm for laboratory experimentation, but of limited scope
On the other hand, we have observational studies with real users
  – realistic but of limited scale
[please do not take this dichotomy too literally!]
Richer models, richer experiments

Why develop richer models?
– because we want richer understanding of the phenomena
– as well as other useful predictions

Why design richer experiments?
– because we want to believe in our models
– and to enrich them further

A rich theory should have something to say *both* to lab experiments in the Cranfield/TREC tradition, *and* to observational studies
Social Book Search

Part II
Motivation

• Web gives access to a wealth of information different in **quantity** but also in **character**
  • Traditional IR focuses on ad hoc search & topical relevance
  • Social media have a different character:
    • different **data** (volume, subjective, opiniated)
    • different **tasks** (views/interpretations rather than facts)
    • different notions of **relevance**?
• What are such new **tasks**? and do these require new **IR models**?
Social Book Search

Professional metadata

Fall Reading
What's hot in Books this fall

Social content
UGC allows new search requests now more than traditional metadata. Yet book search ignores reviews and tags!
Introduction

• Directly compare standard IR & social search
• Context: LibraryThing (LT) discussion forums
  • book requests and suggestions from LT forums
  • compare with familiar search tasks (ad hoc, known-item)
• Relevance in book search is very complex
  • readers have personal preferences, knowledge/reading level
  • want books that are engaging, funny, well-written, recent, educational, comprehensive, ...
Research Questions

• How does social book search compare to traditional search tasks?
  • Can we use forum suggestions for evaluation?
  • How is social book search related to traditional IR tasks?
  • How effective are professional metadata and UGC for book suggestion?
  • How do standard IR models cope with UGC?
  • Do users/judges prefer professional metadata or UGC for determining relevance and recommendation?
Social Book Search

• INEX Social Book Search Track
  • user query is NL statement, system returns ranked list
• Collection: 2.8 million book descriptions
  • ISBN is document ID
  • Mix of Amazon and LT data, fiction & non-fiction
  • crawled by Uni. Duisburg-Essen (for INEX iTrack)
Book Descriptions

- **Professional metadata** is **balanced**, but low quantity (except Amazon browsing categories)

- **UGC** is **skewed** (popularity), higher mean (max cutoff for tag & review)
SBS 2011
LT Recommendations

- **LT members** discuss and ask for books on the LT discussion forums
  - groups for specific genres/subjects
  - member starts topic to ask for recommendations, others suggest books
  - requests are real needs and often very detailed
- Topics covers **broad range of needs**
  - subject, genre, author, known-item, `similar to book X`
Politics of Multiculturalism Recommendations?

Political Philosophy

I'm new, and would appreciate any recommended reading on the politics of multiculturalism. Parekh's Rethinking Multiculturalism: Cultural Diversity and Political Theory (which I just finished) in the end left me unconvincing, though I did find much of value I thought he depended way too much on being able to talk out the details later. It may be that I found his writing style really irritating so adopted a defiant skepticism, but still...

Anyway, I've read Sen, Rawls, Habermas, and Nussbaum, still don't feel like I've wrapped my little brain around the issue very well and would appreciate any suggestions for further anyone might offer.

Reply | More

Will Kymlicka's Multicultural Citizenship is one of the key works within this literature, and his later work has built on but also modified his argument there. See his author page here. I think his latest ones are Multicultural Odysseys and Politics in the Vernacular.
Rethinking Multiculturalism: Cultural Diversity and Political Theory

by Bhikhu Parekh

Book suggestion => LT work ID => ISBN
Creating Topics

- Can we use forum threads for evaluation?
  - Crawled 18,000 threads
  - Manually Selected ~1,000 threads with request for books
  - Used title of thread as topic/query on full-text index
  - Selected 211 threads with 50% recall of suggestions in top 1,000 (to filter non-descriptive titles)

- 211 topics used at INEX’11 SBS track
  - another 43 topics for training to get good pools
LT Forum Suggestions

• Most topic have few relevant books (suggestions): median 7, mean 11

• Incomplete judgments?
  • depends on number of members of discussion group
  • members only suggest books they know or have read
  • but small number may also indicate only best books are suggested (=recommendation rather than relevance?)
Mechanical Turk

- Judge top 10 results of submissions, at least on topical relevance
  - we ran MTurk experiment, selected 24 topics (12 fiction, 12 non-fiction)
  - workers read topic narrative, get 10 book descriptions, answer few questions per book
  - 272 HITs, 3 workers per HIT, $0.50 per HIT ($408 total)
Instructions

For this HIT you're shown a message from the LibraryThing discussions forums, of someone requesting recommendations for books in a certain genre or on a certain topic. We ask you to look at book descriptions of 10 books and decide whether you would recommend these books to this person. Each book description has 2 parts: 1) official description and 2) user-generated description.

1. the official description has information like title, author, publisher and subject classification information.
2. the user-generated description has user reviews and tags from Amazon and LibraryThing

We want to find out which part of the book description is helpful in determining whether you recommend a book or not.

Book request

Would you recommend the books below to a person with the following request:

**Topic:** Help: WWII pacific subs

**LibraryThing group:** Second World War History

**Request:** Can anyone recommend a good strategic level study of us sub campaign in pacific? All I seem to scare up is exploits of individual subs. I have ordered clay blairs big study but I would like something from this decade if it exists

*Important: Please note that some of the pages below have been pre-judged by experts. Your answers will need to match at least 60% of the experts' answers to qualify for payment.*

How familiar are you with the topic of the request? **Very unfamiliar** ○ ○ ○ ○ ○ **Very familiar**
MTurk Quality Control

- We checked MTurk judgments for quality
  - LT Agreement: at least one forum suggestion in each HIT
  - Relevance contradiction: answering book is rel. for Q1 then say it’s non-rel. for Q3 (suggests random clicking)
  - Type contradiction: preferring UGC when there is none
  - Qualification: only US workers, at least 50 assignments, 95% approval rate (previous MTurk work of high quality)

- 7 HITs rejected
  - 3 low agreement, 4 for skipping questions
MTurk Stats

• Quality check:
  • 816 assignments, 133 workers (30 did 3 or more, half of assignments by 7 workers)
  • **LT Agreement**: over workers 0.52, over assignments 0.84, workers who did many have high agreement
  • 18 relevance contradictions (judgments were discarded)
  • no type contradictions

• Suggests workers were conscientious
System Centered Evaluation

• INEX’11 SBS: 4 teams submitted 22 runs

• Submissions were deduplicated
  • ISBN mapped to LT work ID
  • multiple ISBN can map to single LT work ID
  • we keep highest ranked and ignore the rest

• Official evaluation measure is nDCG@10
System Rank Correlation

- **3 sets of judgments**
  - **LT-211**: all 211 forum topics, forum suggestions as judgments
  - **LT-24**: 24 selected topics, same judgments
  - **AMT-24**: 24 selected topics, MTurk topical relevance judgments

- **MTurk judgments give different rankings**
  - **LT suggestions** stable across topic selection
  - So different judgments? Or incomplete suggestions?

---

**Table 2: Kendall’s \( \tau \) and \( \tau_{AP} \) system ranking correlations on nDCG@10 between the three sets of judgements (\( \tau / \tau_{AP} \))**

<table>
<thead>
<tr>
<th></th>
<th>LT-211</th>
<th>AMT-24-Rel</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT-24</td>
<td>0.90/0.83</td>
<td>0.39/0.20</td>
</tr>
<tr>
<td>LT-24</td>
<td>–</td>
<td>0.36/0.19</td>
</tr>
</tbody>
</table>
Incompleteness & MRR

• Best score of 0.481 on MRR for LT suggestions

• if suggestions are incomplete, chance of getting high MRR is low

• high MRR for a few topics could be accidental

• over 211 topics, high MRR with incomplete judgments is unlikely

• Indicates that suggestions are relatively complete

• LT suggestions can be used for evaluation, resulting in a test collection with high fidelity to real-world applications
Importance of Fields

- We created **indexes** of metadata fields:
  - Title (book title only)
  - Dewey (Dewey classification code)
  - Subject (LoC Subject heading)
  - BrowseNode (Amazon browse categories)
  - Review (all Amazon reviews for a book)
  - Tag set (all distinct LT tags)
  - Tag bag (LT tags, tag count = term frequency)
- Indri LM, dirichlet (mu=2,500), Krovetz, stopwords removed
Suggestions vs. MTurk

- **Reviews** most effective for all types of judgments (and currently ignored when searching Amazon or LT)
- **Titles** work for AMT relevance/rec. not for LT suggestions (title bias?)
- **Standard LM** is effective for UCG -- both for AMT and LT
- **AMT-Rel&Rec** closest to LT-Sug, but not identical judgment

Table 4: MTurk and LT Forum evaluation of runs over different index fields

<table>
<thead>
<tr>
<th>Field</th>
<th>AMT-Rel</th>
<th>AMT-Rec</th>
<th>AMT-Rel&amp;Rec</th>
<th>LT-Sug</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nDCG10</td>
<td>MAP R@1000</td>
<td>nDCG10</td>
<td>MAP R@1000</td>
</tr>
<tr>
<td>Title (field)</td>
<td>0.212</td>
<td>0.105</td>
<td>0.601</td>
<td>0.260</td>
</tr>
<tr>
<td>BrowseNode</td>
<td>0.096</td>
<td>0.052</td>
<td>0.322</td>
<td>0.142</td>
</tr>
<tr>
<td>Dewey</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>0.003</td>
</tr>
<tr>
<td>Subject</td>
<td>0.016</td>
<td>0.002</td>
<td>0.008</td>
<td>0.021</td>
</tr>
<tr>
<td>Review</td>
<td><strong>0.579</strong></td>
<td><strong>0.309</strong></td>
<td><strong>0.720</strong></td>
<td><strong>0.786</strong></td>
</tr>
<tr>
<td>Tag (set)</td>
<td>0.337</td>
<td>0.173</td>
<td><strong>0.744</strong></td>
<td>0.422</td>
</tr>
<tr>
<td>Tag (bag)</td>
<td>0.368</td>
<td>0.182</td>
<td>0.694</td>
<td>0.435</td>
</tr>
</tbody>
</table>
Suggestions vs. Known-Item

We created 50 Known-Item topics based on sampled suggestions from 211 topics.

- For **KI** topics, **Review** and **Title** indexes are effective.
- For **LT forum** suggestions: **Title** index is much less effective.

### Table 3: Known-item and forum suggestion evaluation of runs over different index fields

<table>
<thead>
<tr>
<th>Field</th>
<th>Known-item</th>
<th></th>
<th>Forum suggestions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRR</td>
<td>R@10</td>
<td>R@1000</td>
<td>MRR</td>
</tr>
<tr>
<td>Title</td>
<td>0.414</td>
<td>0.540</td>
<td><strong>0.820</strong></td>
<td>0.118</td>
</tr>
<tr>
<td>BrowseNode</td>
<td>0.004</td>
<td>0.000</td>
<td>0.240</td>
<td>0.083</td>
</tr>
<tr>
<td>Dewey</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>Subject</td>
<td>0.010</td>
<td>0.020</td>
<td>0.020</td>
<td>0.012</td>
</tr>
<tr>
<td>Review</td>
<td><strong>0.480</strong></td>
<td><strong>0.680</strong></td>
<td><strong>0.800</strong></td>
<td><strong>0.382</strong></td>
</tr>
<tr>
<td>Tag (set)</td>
<td>0.118</td>
<td>0.220</td>
<td>0.540</td>
<td>0.213</td>
</tr>
<tr>
<td>Tag (bag)</td>
<td>0.227</td>
<td>0.400</td>
<td>0.560</td>
<td>0.342</td>
</tr>
</tbody>
</table>
User Centered Analysis

- MTurk workers need at least one review, otherwise they don’t have enough info for judgment
- Reviews are important for both relevance and recommendation
- Presence of tags has little impact (similar to subject headings?)

Table 5: Impact of presence of reviews and tags on judgements

<table>
<thead>
<tr>
<th></th>
<th>Reviews 0 rev.</th>
<th>≥1 rev.</th>
<th>Tags 0 tags</th>
<th>≥10 tags</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top. Rel. (Q1)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enough info.</td>
<td>0.37</td>
<td>0.01</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>Relevant</td>
<td>0.30</td>
<td>0.54</td>
<td>0.49</td>
<td>0.48</td>
</tr>
<tr>
<td><strong>Recommend. (Q3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enough info.</td>
<td>0.53</td>
<td>0.01</td>
<td>0.14</td>
<td>0.12</td>
</tr>
<tr>
<td>Rel. + Rec.</td>
<td>0.22</td>
<td>0.51</td>
<td>0.46</td>
<td>0.45</td>
</tr>
</tbody>
</table>
When there are no reviews, workers use professional metadata for relevance, but can’t make a recommendation.

With at least one review, they prefer UGC and can make a recommendation.

Table 6: Impact of the presence of reviews on metadata preference

<table>
<thead>
<tr>
<th></th>
<th>Q2. Relevance</th>
<th>Q4. Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all</td>
<td>0 rev.</td>
</tr>
<tr>
<td>Prof.</td>
<td>0.29</td>
<td>0.51</td>
</tr>
<tr>
<td>Equal</td>
<td>0.27</td>
<td>0.40</td>
</tr>
<tr>
<td>UGC</td>
<td>0.43</td>
<td>0.06</td>
</tr>
<tr>
<td>Skip</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

With at least one review, they prefer UGC and can make a recommendation.
Wrap Up SBS’11

- **Social Book Search** is a new area of research
- Book search over **professional metadata** and **UGC**
- **LT forums** as window on real book search needs and suggestions
- Suggestions are **complete enough** for evaluation!
  - ...but different from **topical rel. judgments** or **known-item**
- **Standard retrieval models** deal well with skewed distribution of **UGC** across descriptions
  - Although currently Amazon and LT ignore **UGC** in their search, it improves both **ad hoc search** and **book suggestion**
  - **Reviews** are important for both **relevance** and **recommendation**: provide **information of a different nature** than other metadata
Continue with SBS

• Recall the motivation: **Web data** is different in **quantity** but also in **character**

• Understand the differences in **task** and **relevance judgments**

• Differentiate **suggestions** based on personal **catalogue of topic creator**

• Analyze **relevance dimensions**: recommendation versus **relevance**
SBS Changes

- **Collection**: 2.8M book records
- Extended with library records from BL and LOC
- **Topics**: forum requests from LibraryThing (LT)
- **User Profiles**: LT catalogues of topic creators
- **Judgments**: suggestions from forum threads
  - graded by what’s in catalogue of topic creator
- **Task**: submit ranked list of books that user wants to catalogue
Creating Topic Set

• We use the forum topics for evaluation
  • Crawled 60,000 topics, discarded topics without suggestions
  • Ran remaining topic titles as queries on full-text index
  • Discarded topics with no suggestion in top 1,000
  • Labelled 89 new topics that have request for books
  • Added them to 211 topics of last year

• That’s this year’s topic set (300 topics)
  • suggestions are used as relevance judgments
Suggestion? Relevant?

• Each suggestion is a human judgment, so relevant in some sense...

• Topic creator may not like every suggestion

• We compare suggestions to what topic creator has in her catalogue
User Profile/Catalogue

- Book selection is very personal
- others cannot make good judgments (so traditional topical rel. judgments are out)
- requires knowledge of user preference, likes/dislikes, mood, reading level, etc.
- Can we use profile information to reflect personal preference in judgments?
Politics of Multiculturalism Recommendations?

Political Philosophy

I'm new, and would appreciate any recommended reading on the politics of multiculturalism. Parekh's Rethinking Multiculturalism: Cultural Diversity and Political Theory (which I just finished) in the end left me unconvincing, though I did find much of value I thought he depended way too much on being able to talk out the details later. It may be that I found his writing style really irritating so adopted a defiant skepticism, but still...

Anyway, I've read Sen, Rawls, Habermas, and Nussbaum, still don't feel like I've wrapped my little brain around the issue very well and would appreciate any suggestions for further anyone might offer.

Reply | More

2 rsterling
Edited: Sep 27, 2010, 1:31am

Will Kymlicka's Multicultural Citizenship is one of the key works within this literature, and his later work has built on but also modified his argument there. See his author page here. I think his latest ones are Multicultural Odysseys and Politics in the Vernacular.
<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Year</th>
<th>Subjects</th>
<th>Rating</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics of Piety: The Islamic Revival and the Feminist Subject</td>
<td>Saba Mahmood</td>
<td>2005</td>
<td>feminist theory, feminism, political theory, multiculturalism</td>
<td>3 stars</td>
<td>Aug 20, 2012</td>
</tr>
<tr>
<td>Constitutional Patriotism</td>
<td>Jan-Werner Muller</td>
<td>2007</td>
<td>political theory, multiculturalism, european union, germany</td>
<td>3 stars</td>
<td>Oct 31, 2010</td>
</tr>
<tr>
<td>Rethinking Multiculturalism: Cultural Diversity and Political Theory</td>
<td>Bhikhu Parekh</td>
<td>2002</td>
<td>political theory, immigration, multiculturalism</td>
<td>3 stars</td>
<td>Sep 1, 2010</td>
</tr>
</tbody>
</table>
Post-Cataloguing

- 94 out of 300 topics have at least one Post-Catalogued Suggestion (PCS)
  - these are used for evaluation (graded judgments)

- Forum suggestions relevant (rv=1) except:
  - post-catalogued suggestions (rv=4)
  - creator-provided suggestions (rv=0)
  - avg. 16.2 suggestions per topic, 2 provided by creator, 1.7 post-catalogued, 12.5 have rv=1
Table 5. Evaluation results for the official submissions. Best scores are in bold.

<table>
<thead>
<tr>
<th>Run</th>
<th>MRR</th>
<th>nDCG@10</th>
<th>P@10</th>
<th>R@10</th>
<th>R@1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>p54.run2.all-topic-fields.all-doc-fields</td>
<td>0.3069</td>
<td>0.1492</td>
<td>0.1198</td>
<td>0.1527</td>
<td>0.5736</td>
</tr>
<tr>
<td>p54.run3.all-topic-fields.QIT.alpha0.99</td>
<td>0.3066</td>
<td>0.1488</td>
<td>0.1198</td>
<td>0.1527</td>
<td>0.5736</td>
</tr>
<tr>
<td>p4.inex2012SBS.xml_social.fb.10.50</td>
<td>0.3616</td>
<td>0.1437</td>
<td>0.1219</td>
<td>0.1494</td>
<td>0.5775</td>
</tr>
<tr>
<td>p62.B_IT30.30</td>
<td>0.3410</td>
<td>0.1339</td>
<td></td>
<td>0.1260</td>
<td>0.1659</td>
</tr>
<tr>
<td>p4.inex2012SBS.xml_social</td>
<td>0.3256</td>
<td>0.1297</td>
<td>0.1135</td>
<td>0.1476</td>
<td>0.5588</td>
</tr>
<tr>
<td>p62.mrf-booklike</td>
<td>0.3584</td>
<td>0.1295</td>
<td>0.1250</td>
<td>0.1514</td>
<td>0.5242</td>
</tr>
<tr>
<td>p54.run5.title.II.alpha0.94</td>
<td>0.2558</td>
<td>0.1173</td>
<td>0.1073</td>
<td>0.1289</td>
<td>0.4891</td>
</tr>
<tr>
<td>p62.IOT30</td>
<td>0.2933</td>
<td>0.1141</td>
<td>0.1240</td>
<td>0.1503</td>
<td>0.5864</td>
</tr>
<tr>
<td>p62.IT30</td>
<td>0.2999</td>
<td>0.1082</td>
<td>0.1187</td>
<td>0.1426</td>
<td>0.5864</td>
</tr>
<tr>
<td>p54.run6.title.II.alpha0.97</td>
<td>0.2392</td>
<td>0.0958</td>
<td>0.0823</td>
<td>0.0941</td>
<td>0.4891</td>
</tr>
<tr>
<td>p62.lcm-2</td>
<td>0.2149</td>
<td>0.0901</td>
<td>0.0667</td>
<td>0.1026</td>
<td>0.5054</td>
</tr>
<tr>
<td>p100.sb_g0</td>
<td>0.2394</td>
<td>0.0884</td>
<td>0.0844</td>
<td>0.1145</td>
<td>0.5524</td>
</tr>
<tr>
<td>p54.run4.title.QIT.alpha0.65</td>
<td>0.1762</td>
<td>0.0875</td>
<td>0.0719</td>
<td>0.0949</td>
<td>0.4891</td>
</tr>
<tr>
<td>p100.sb_g_ttl_nar0</td>
<td>0.1581</td>
<td>0.0740</td>
<td>0.0594</td>
<td>0.0939</td>
<td>0.4634</td>
</tr>
<tr>
<td>p54.run1.title.all-doc-fields</td>
<td>0.1341</td>
<td>0.0678</td>
<td>0.0583</td>
<td>0.0729</td>
<td>0.4891</td>
</tr>
<tr>
<td>p100.sb_2xsh_ttl_nar0</td>
<td>0.0157</td>
<td>0.0057</td>
<td>0.0021</td>
<td>0.0022</td>
<td>0.0393</td>
</tr>
<tr>
<td>p100.sb_2xsh0</td>
<td>0.0199</td>
<td>0.0042</td>
<td>0.0021</td>
<td>0.0020</td>
<td>0.0647</td>
</tr>
</tbody>
</table>
Results (LT Forums)

- **RSLIS** Copenhagen has best run: all topic fields as query, all document fields in index

- **RSLIS** Copenhagen & LIA Avignon used **user profile information** for some submissions...

- ... but not for their best runs

- Official submissions do not show added value of user profiles (yet)
Mechanical Turk

- LT suggestions are relatively complete
  - high MRR scores over last year’s 211 topics means suggestions can’t be highly incomplete

- Analyze suggestions with MTurk
  - complex relevance judgments for top 10 results of submissions, covering different aspects of relevance
  - workers read topic narrative, get 10 book descriptions, answer few questions per book
MTurk Judgments (1/2)

- Judge top-k results of submitted runs
  - Add at least 1 suggestion from forum (so one known good book)

- We want to know:
  - is book relevant/recommendable?
  - what information is used for the judgment? (metadata fields: title, reviews, ...)
  - what relevance aspects are important? (subject, genre, recency, comprehensiveness, engagement, etc.)
MTurk Judgments (2/2)

• 37 runs from 5 groups (2 yrs), top 10 pool, 25 topics
• 10 books/HIT for $0.50
• 326 HITs, 3 workers/HIT (US only, 95% approval)
• Cost: 326 x 3 x $0.50 = $489 + 10% = $537
• Mean time/HIT: 11:59 (median 9:17)
• 55% of workers spent 6-15 minutes
Topic: Tudor Fiction

LibraryThing group: English History - Tudor through Edwardian

Request: Any Tudor fiction suggestions? I'm already completely wrapped up in Philippa Gregory's books about the women of Henry VIII's court: The Constant Princess, The Other Boleyn Girl, and The Boleyn Inheritance. What others do people recommend??

Important: Please note that some of the books have been pre-judged by experts. Your answers will need to match at least 60% of the experts' answers for payment.

How familiar are you with the topic of the request? Very unfamiliar ○ ○ ○ ○ ○ Very familiar

Which aspects of books are relevant to this request:

- Subject
- Genre
- Recency
- Comprehensiveness
- Reading level (difficulty)
- Objectivity
- Engagement/fun factor
- Novelty / familiarity
Topic: Tudor Fiction

LibraryThing group: English History - Tudor through Edwardian

Request: Any Tudor fiction suggestions? I'm already completely wrapped up in Philippa Gregory's books about the women of Henry VIII's court: The Constant Princess, The Other Boleyn Girl, and The Boleyn Inheritance. What others do people recommend??

With the forum request in mind, please look at the different parts of the book description on the right and answer the questions below.

Q1. Which of the requested aspects does this book satisfy?
   - Subject
   - Genre
   - Recency
   - Comprehensiveness
   - Reading level
   - Objectivity
   - Engaging/fun
   - Novel/familiar

Q2. What book information did you find useful to determine this?

<table>
<thead>
<tr>
<th>Type</th>
<th>Very</th>
<th>Somewhat</th>
<th>Not</th>
<th>Didn't use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catalogue information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User reviews</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User tags</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q3. Would you recommend this book to the requestor?
Yes, it perfectly fits the request and looks like a great book.
## Worker Judgments

<table>
<thead>
<tr>
<th>Choice</th>
<th>Rel</th>
<th>Rec</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes, it perfectly fits the request and looks like a great book</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probably, it doesn't fit the request completely, but still looks worth reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No, it fits the request, but doesn't look like it's worth reading</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No, it doesn't match the request</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No, the requester already has (read) this book</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I don't know, the book description is too incomplete or too vague</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Quality Control

• Some quality control measures:
  • **Agreement**: 1 forum suggestion/HIT (should be relevant). Agreement $> 60\%$ once worker did 3 HITs or more.
  • **Consistency**: 1) if book aspect is relevant (Q1), some description parts must be relevant (Q2), 2) if topic rel. (Q3) $\rightarrow$ subject or genre aspects must be ticked (Q1)

• Unfortunately, over 600 HITs approved automatically (too late to check work)

• But in first 350 assignments only 8 had to be rejected
Agreement

Q3: would you recommend this book?

1) rec+rel, 2) rec+nonrel, 3) nonrec+rel, 4) nonrec+nonrel

pairwise per HIT (3 workers = 3 pairs)

<table>
<thead>
<tr>
<th></th>
<th>min</th>
<th>max</th>
<th>med.</th>
<th>mean</th>
<th>std.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>rel+rec</td>
<td>0</td>
<td>1</td>
<td>0,6</td>
<td>0,56</td>
<td>0,21</td>
</tr>
<tr>
<td>rel</td>
<td>0,1</td>
<td>1</td>
<td>0,7</td>
<td>0,72</td>
<td>0,19</td>
</tr>
<tr>
<td>rec</td>
<td>0,2</td>
<td>1</td>
<td>0,7</td>
<td>0,73</td>
<td>0,17</td>
</tr>
</tbody>
</table>
## Agreement with Suggestions

<table>
<thead>
<tr>
<th></th>
<th>All books</th>
<th>LT Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>total</strong></td>
<td>9780</td>
<td>1247</td>
</tr>
<tr>
<td><strong>unknown</strong></td>
<td>522</td>
<td>24</td>
</tr>
<tr>
<td><strong>unrelated</strong></td>
<td>3491</td>
<td>68</td>
</tr>
<tr>
<td><strong>related</strong></td>
<td>1952</td>
<td>213</td>
</tr>
<tr>
<td><strong>relevant</strong></td>
<td>3865</td>
<td>942</td>
</tr>
<tr>
<td><strong>recommended</strong></td>
<td>5212</td>
<td>976</td>
</tr>
</tbody>
</table>

- Agreement with suggestions is high
- Suggestions more often relevant and recommended than other books
Relevance Aspects

- **Subject** is relevant for almost all requests (93%)
  - and a *necessary condition* for a book’s relevance.

- **Other important aspects**
  - genre (71%), comprehensiveness (39%), engagement (21%)
  - these aspects are *usually ignored* in trad. IR evaluation

- **Workers select relevance aspects for books**
  that they don’t select for the forum request
  - perhaps to point out *recommendable qualities of a book*,
    regardless of whether requester asked for it or not
Relevant or Recommended?

- Difference between relevance and recommendation:
  - 22% of recommended books is not relevant but related (recommended on other qualities)
  - 5% is relevant but not recommended
  - 92% of LT suggestions is relevant or related, 76% is topically relevant. Request could be about genre, not subject.
Browsing Behavior

• We log clicks and track which parts of description are browsed by workers
  • title information always shown, other parts require click.

• How often is book part clicked?
  • catalogue (76%), reviews (72%), and tags (63%)
  • For 1 in 4 books judges only use title information
### Usefulness of Metadata

<table>
<thead>
<tr>
<th></th>
<th>Very</th>
<th>Some</th>
<th>Not</th>
<th>Not used</th>
<th>Not present</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>0.65</td>
<td>0.25</td>
<td>0.07</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td>Catalogue</td>
<td>0.37</td>
<td>0.27</td>
<td>0.18</td>
<td>0.18</td>
<td>0.07</td>
</tr>
<tr>
<td>Reviews</td>
<td>0.29</td>
<td>0.15</td>
<td>0.24</td>
<td>0.32</td>
<td>0.4</td>
</tr>
<tr>
<td>Tags</td>
<td>0.26</td>
<td>0.24</td>
<td>0.18</td>
<td>0.32</td>
<td>0.07</td>
</tr>
</tbody>
</table>

- How useful is **metadata type** for judgment?
- **Title information** is very important (book judged by cover)
- **No reviews** for 40% of books. When present they’re useful.
Conclusions

- **Social Book Search** is a new area of research
- Social media integration raises many open questions
- Ideal playground for studying professional metadata and UGC
- LT forums provide real book search needs and suggestions

- Studied the **character** of topic-based recommendation
  - **Relevance aspects**: Subject (93%), genre (73%), comprehensiveness (39%), engagement (21%)
  - **Information**: Title, catalogue (76%), reviews (72%), and tags (63%)
  - **Relevant/Recommended**: 5% is rel+not rec. 22% is rec+not rel.
Social Book Search (SBS)

• Book Search with social data:
  • book information from trained professionals and users
  • user-generated content
    • almost always ignored in search index
    • uncontrolled, inconsistent, unlike library catalogues (well...)
    • unbalanced, skewed towards popular books
  • but:
    • “The user-generated review was so enthusiastic, I would recommend it just based on that.”
Search or Recommendation?

- **Two paradigms:**
  - **Retrieval** (trad. catalogue)
  - **Recommendation** (user data)

- **Amazon, GoodReads, LibraryThing**
  - Offer both paradigms for search and discovery
  - **Retrieval** ignores user profile and user-gen. content
  - **Recommendation** ignores specific information need
  - Book search requires both?
Suggestions and Relevance?

- Discussion participants form crowd of recommenders
  - Thread is discussion of “best” books?
  - Each suggestion is a human judgement, so relevant in some sense (2011)
  - Topic creator may not like every suggestion: check if topic creator adds to catalogue (2012)
  - Not all books mentioned may be intended as suggestions: analyse and label suggestions (2013/2014)
Creating Topic Set

• We use the forum topics for evaluation
  • we built a new, simple interface to gather annotations
  • 8 students from RSLIS, OUC & AAU annotated topics and suggestions
  • many thanks to Toine Bogers, Birgir Larsen and Michael Preminger for recruiting and paying them!

• 2,646 topics annotated
  • 944 are book search topics, 680 with suggestions
  • annotated suggestions are used as relevance judgements
Example Topic

<title>Politics of Multiculturalism Recommendations?</title>
<query>Politics of Multiculturalism</query>
<group>Political Philosophy</group>
<narrative>I’m new, and would appreciate any recommended …</narrative>
<catalog>
  <book><LT_id>9036</LT_id><entry_date>2007-09</entry_date>…</book>…
</catalog>
Distribution of Aspects

- Content is most important relevance aspect (topical relevance!)
- Familiarity is important (recommendation!)
- Engagement is hard to express in query (reason for social search?)
### Number of Aspects

<table>
<thead>
<tr>
<th># aspects</th>
<th># topics</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>191</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>260</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>183</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

- Topics have multiple relevance aspects
- Suggests multiple metadata sources are needed
From Suggestions To Relevance Judgements

• We used a complex decision tree to map annotated suggestions to relevance values
  • opinion of topic creator overrules others
  • opinions of users who read book overrule opinions of those who didn’t
  • opinion of multiple people weights stronger than opinion of single person
User Profile/Catalogue

- Book selection is very personal
  - others cannot make good judgements for user,
  - traditional topical rel. judgements are problematic (Koolen & Kamps, CIKM 2012)

- Requires knowledge of user
  - preference, knowledge, reading level, ...
  - Can we use personal catalogue to reflect personal preference in suggestions?
Topic Creator Profiles

• We distributed user profiles for 680 topics
  • plus anonymised profiles of 93,976 other users

• Allows
  • content based recommendation (topical, temporal, ...)
  • collaborative filtering
## Example Profiles

<table>
<thead>
<tr>
<th>user-id</th>
<th>book-id</th>
<th>entry-date</th>
<th>rating</th>
<th>tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>u8218518</td>
<td>952822</td>
<td>2012-02</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>3344349</td>
<td>2012-02</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>2317257</td>
<td>2012-02</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>6415999</td>
<td>2012-02</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>5525956</td>
<td>2012-09</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>842432</td>
<td>2012-02</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>u8218518</td>
<td>3171103</td>
<td>2012-02</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>u8780837</td>
<td>542201</td>
<td>2009-05</td>
<td>10.0</td>
<td>tarot</td>
</tr>
<tr>
<td>u9054475</td>
<td>5403381</td>
<td>2010-11</td>
<td>10.0</td>
<td>adventure, potter, philosopher, Harry, stone</td>
</tr>
</tbody>
</table>

...
Table 4. User profile statistics of the topic creators and all other users.

<table>
<thead>
<tr>
<th>Type</th>
<th>N</th>
<th>total</th>
<th>min</th>
<th>max</th>
<th>median</th>
<th>mean</th>
<th>stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic Creators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-catalogued</td>
<td>680</td>
<td>399,147</td>
<td>1</td>
<td>5884</td>
<td>239</td>
<td>587</td>
<td>927</td>
</tr>
<tr>
<td>Post-catalogued</td>
<td>680</td>
<td>209,289</td>
<td>1</td>
<td>5619</td>
<td>114</td>
<td>308</td>
<td>499</td>
</tr>
<tr>
<td>Total catalogue</td>
<td>680</td>
<td>608,436</td>
<td>2</td>
<td>8563</td>
<td>432</td>
<td>895</td>
<td>1202</td>
</tr>
<tr>
<td>All users</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>93,976</td>
<td>33,503,999</td>
<td>1</td>
<td>41,792</td>
<td>134</td>
<td>357</td>
<td>704</td>
</tr>
<tr>
<td>Total</td>
<td>94,656</td>
<td>34,112,435</td>
<td>1</td>
<td>41,792</td>
<td>135</td>
<td>360</td>
<td>710</td>
</tr>
</tbody>
</table>
## Evaluation Results

<table>
<thead>
<tr>
<th>Group</th>
<th>Run</th>
<th>nDCG@10</th>
<th>P@10</th>
<th>MRR</th>
<th>MAP</th>
<th>Profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>USTB</td>
<td>run6.SimQuery1000.rerank_all.L2R.RandomForest</td>
<td>0.303</td>
<td>0.464</td>
<td>0.232</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>run4.newXml.rerank_all.L2R.RandomForest</td>
<td>0.142</td>
<td>0.258</td>
<td>0.102</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>326</td>
<td>0.142</td>
<td>0.275</td>
<td>0.107</td>
<td>0.426</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>run3.newXml.rerank_all.L2R.Coordinate</td>
<td>0.138</td>
<td>0.256</td>
<td>0.101</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>run5.newXml.rerank_all.L2R.RankNet</td>
<td>0.133</td>
<td>0.246</td>
<td>0.098</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>run2.newXml.rerank_T</td>
<td>0.131</td>
<td>0.246</td>
<td>0.096</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>USTB</td>
<td>run1.newXml.feedback</td>
<td>0.128</td>
<td>0.246</td>
<td>0.095</td>
<td>0.390</td>
<td>No</td>
</tr>
<tr>
<td>LSIS</td>
<td>InL2</td>
<td>0.128</td>
<td>0.236</td>
<td>0.101</td>
<td>0.441</td>
<td>No</td>
</tr>
<tr>
<td>AAU</td>
<td>run1.all-plus-query.all-doc-fields</td>
<td>0.127</td>
<td>0.239</td>
<td>0.097</td>
<td><strong>0.444</strong></td>
<td>No</td>
</tr>
<tr>
<td>AAU</td>
<td>run3.all-plus-query.all-doc-fields</td>
<td>0.120</td>
<td>0.227</td>
<td>0.090</td>
<td>0.425</td>
<td>No</td>
</tr>
<tr>
<td>CYUT</td>
<td>Type2QTGN</td>
<td>0.119</td>
<td>0.246</td>
<td>0.086</td>
<td>0.340</td>
<td>No</td>
</tr>
<tr>
<td>CYUT</td>
<td>0.95AverageType2QTGN</td>
<td>0.119</td>
<td>0.243</td>
<td>0.085</td>
<td>0.332</td>
<td>No</td>
</tr>
<tr>
<td>UJM</td>
<td>328</td>
<td>0.117</td>
<td>0.226</td>
<td>0.088</td>
<td>0.392</td>
<td>Yes</td>
</tr>
<tr>
<td>UJM</td>
<td>329</td>
<td>0.116</td>
<td>0.217</td>
<td>0.087</td>
<td>0.392</td>
<td>Yes</td>
</tr>
<tr>
<td>UJM</td>
<td>325</td>
<td>0.115</td>
<td>0.214</td>
<td>0.087</td>
<td>0.392</td>
<td>Yes</td>
</tr>
<tr>
<td>LSIS</td>
<td>InL2Feedback</td>
<td>0.114</td>
<td>0.230</td>
<td>0.094</td>
<td>0.434</td>
<td>No</td>
</tr>
<tr>
<td>UJM</td>
<td>324</td>
<td>0.112</td>
<td>0.214</td>
<td>0.086</td>
<td>0.392</td>
<td>No</td>
</tr>
<tr>
<td>LSIS</td>
<td>InL2tagFeedback</td>
<td>0.102</td>
<td>0.212</td>
<td>0.075</td>
<td>0.388</td>
<td>No</td>
</tr>
<tr>
<td>UvA</td>
<td>inex14.ti_qu.fb.10.50.5000</td>
<td>0.097</td>
<td>0.179</td>
<td>0.073</td>
<td>0.421</td>
<td>No</td>
</tr>
<tr>
<td>UMD</td>
<td>Full_TQG_fb.10.50_0.0000227_50</td>
<td>0.097</td>
<td>0.188</td>
<td>0.069</td>
<td>0.328</td>
<td>Yes</td>
</tr>
<tr>
<td>UMD</td>
<td>Social_TQG_fb.10.50_0.0000222_50</td>
<td>0.096</td>
<td>0.184</td>
<td>0.067</td>
<td>0.327</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Analysis

• All top runs use User-Generated Content
  • best run: USTB (Beijing), hybrid learning to rank
  • no run uses prof. metadata different from user content

• MRR score are low
  • Many hard topics, all systems score zero

• Exploiting profile information
  • Some runs use profile information, but none in top ranks
Tasks & Relevance
Aspects

• 8 aspects:
  • accessibility, content, engagement, familiarity, known-item, metadata, novelty, socio-cultural

• Aspects reveal relation between retrieval and recommendation?
  • Content aspect typical for search, familiarity for retrieval?
  • many topics contain both aspects
Topic Categories

- Known-item (KI) contains all known-item topics (202 topics)
- Search (S): contains topics with content but not familiarity (338 topics)
- Search and Recommendation (SR) contains topics with both content and familiarity topics (260 topics)
- Recommendation (R) contains topics with familiarity, but not content (66 topics)
- Context (C) contains all topics without content, familiarity, and known-item (78 topics)
3 cooks baking a cake

Librarians who LibraryThing

3 messages | ★ Star this topic | × Ignore topic | ¶ Jump to first unread (3 unread)

This topic is currently marked as "dormant"—the last message is more than 90 days old. You can revive it by posting a reply.

1 TheNovelWorld Dec 16, 2011, 7:15pm

That's all I have to go on for this picture book request. Any ideas on the title or author?

Thanks so much!
Recommendations for novels about the build-up to World War II

**Historical Fiction**

6 messages | ★ Star this topic | × Ignore topic | ✂ Jump to bottom (0 unread)

This topic is currently marked as "dormant"—the last message is more than 90 days old. You can revive it by posting a reply.

1 sturlington

Dec 9, 2011, 10:05am

I am interested in reading some fiction set immediately before WWII depicting the build-up to the war, particularly from a German perspective. Can anyone recommend a book? I know this is fairly specific...

Reply | More
Catholic Children's Books

Catholic Tradition

7 messages | ★ Star this topic | × Ignore topic | ⬇ Jump to first unread (7 unread)

This topic is currently marked as "dormant"—the last message is more than 90 days old. You can revive it by posting a reply.

1 CaleeMLee Oct 18, 2011, 12:14am ⬇

I always get my kids (preschool age) a few books for Christmas and would love your suggestions. My daughter is just getting into listening to chapter books so I think we'll start the Narnia books. Any other books out there you'd recommend? She loved *The Holy Twins* by Kathleen Norris.

More
Merchant/Trader SF

Science Fiction Fans

33 messages | ★ Star this topic | × Ignore topic | ▼ Jump to bottom (0 unread)

This topic is currently marked as "dormant"—the last message is more than 90 days old. You can revive it by posting a reply.

1 cosmicdolphin

Jan 1, 2012, 8:52pm

I'd like suggestions for good Merchant/Trader SF? I've read Cherryh's Merchanter books, and I'm currently reading the Andre Norton Solar Queen books (both of which are good in their own ways)

Any other SF with Traders/Merchants as the primary focus of the story?

Thanks

Reply | More
Great books for year 8 students

Librarians who LibraryThing

This topic is currently marked as "dormant"—the last message is more than 90 days old. You can revive it by posting a reply.

1 jothom Oct 12, 2011, 9:12pm

I have been asked to recommend a class text for year 8 English classes. I would really like to have something new that kids will love. Does anyone have any suggestions? Not too long, and fairly accessible as we have lots of kids with literacy problems.
## Types & Performance

<table>
<thead>
<tr>
<th>Type</th>
<th># topics</th>
<th>nDCG@10 with aspect</th>
<th>nDCG@10 w/o aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>56</td>
<td>0.120</td>
<td>0.100</td>
</tr>
<tr>
<td>Known-item</td>
<td>97</td>
<td>0.208</td>
<td>0.084</td>
</tr>
<tr>
<td>Recommend</td>
<td>53</td>
<td>0.126</td>
<td>0.100</td>
</tr>
<tr>
<td>Search</td>
<td>274</td>
<td>0.094</td>
<td>0.107</td>
</tr>
<tr>
<td>Search+Rec</td>
<td>200</td>
<td>0.050</td>
<td>0.123</td>
</tr>
</tbody>
</table>

*Note: some categories harder than others, search+rec hardest*
User Catalogues

• For 589 of 680 topics, user has catalogue
  • For 91, we don’t know (private vs. empty)

• Profiles give us:
  • catalogue size: how many books catalogued before posting request on forum
  • popularity of books in catalogue: some read only popular books, others obscure books or mix
**Catalogue Size**

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Known-Item</th>
<th>Context</th>
<th>S+R</th>
<th>Rec</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Topic</td>
<td>84</td>
<td>0</td>
<td>38</td>
<td>100</td>
<td>104</td>
<td>177</td>
</tr>
<tr>
<td>Post-Topic</td>
<td>65</td>
<td>4</td>
<td>80</td>
<td>65</td>
<td>81</td>
<td>108</td>
</tr>
<tr>
<td>Pre+Post</td>
<td>197</td>
<td>16</td>
<td>155</td>
<td>201</td>
<td>195</td>
<td>415</td>
</tr>
</tbody>
</table>

- Median catalogue size related to type of request
- Heavy readers can explain more precisely what they want?
- Specific aspects instead of “similar to X, Y and Z”
- Categories represent different stages in search process?
Impact of Catalogue Size

<table>
<thead>
<tr>
<th></th>
<th>#tpcs</th>
<th>small cat.</th>
<th>big cat.</th>
<th>no cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>24 / 24 / 8</td>
<td>0.127</td>
<td>0.147</td>
<td>0.016</td>
</tr>
<tr>
<td>Known-item</td>
<td>24 / 24 / 9</td>
<td>0.258</td>
<td>0.146</td>
<td>0.262</td>
</tr>
<tr>
<td>Recommend</td>
<td>23 / 24 / 6</td>
<td>0.176</td>
<td>0.068</td>
<td>0.168</td>
</tr>
<tr>
<td>Search</td>
<td>119 / 119 / 36</td>
<td>0.068</td>
<td>0.120</td>
<td>0.094</td>
</tr>
<tr>
<td>Search+Rec</td>
<td>84 / 84 / 32</td>
<td>0.032</td>
<td>0.068</td>
<td>0.053</td>
</tr>
<tr>
<td>All</td>
<td>294 / 295 / 91</td>
<td>0.100</td>
<td>0.106</td>
<td>0.094</td>
</tr>
</tbody>
</table>

- Search topics of heavy readers better formulated? More specific?
Book Popularity

- Profiles provide popularity information
  - Some users catalogue only popular books
  - Others only obscure books or a mix
  - Popular books have more UGC
  - Do readers of obscure books have different needs or express them differently than readers of popular books?
  - We rank searchers by the median popularity of books in their catalogue
# Median Book Popularity

<table>
<thead>
<tr>
<th></th>
<th>nDCG@10</th>
<th>#tpcs</th>
<th>least pop.</th>
<th>most pop.</th>
<th>no cat.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td></td>
<td>24 / 24 / 8</td>
<td>0.119</td>
<td>0.155</td>
<td>0.016</td>
</tr>
<tr>
<td><strong>Known-item</strong></td>
<td></td>
<td>44 / 44 / 9</td>
<td>0.199</td>
<td>0.205</td>
<td>0.262</td>
</tr>
<tr>
<td><strong>Recommend</strong></td>
<td></td>
<td>23 / 24 / 6</td>
<td>0.094</td>
<td>0.146</td>
<td>0.168</td>
</tr>
<tr>
<td><strong>Search</strong></td>
<td></td>
<td>119 / 119 / 36</td>
<td>0.116</td>
<td>0.071</td>
<td>0.094</td>
</tr>
<tr>
<td><strong>Search+Rec</strong></td>
<td></td>
<td>84 / 84 / 32</td>
<td>0.046</td>
<td>0.053</td>
<td>0.053</td>
</tr>
<tr>
<td><strong>All</strong></td>
<td></td>
<td>294 / 295 / 91</td>
<td>0.111</td>
<td>0.095</td>
<td>0.094</td>
</tr>
</tbody>
</table>

- Note: popularity affects context, recommend and search categories
SBS’14 Conclusions

• Topic set covers many request types
  • mix of tasks: topic search, recommendation, known-item
• User profiles provide another perspective on requests
  • request type related to catalogue size and popularity of books
• Catalogue size and popularity of books affect performance on topics
  • Related to how specific/concrete or well-expressed it is?
Interactive Social Book Search
Information Seeking Stage Aware Systems

- Elastic Search Index
- PyIRE workbench
- iSBS UI components
- Search requests
- Session storage
Multistage: Focus

- Search box
- Search history
- Filter options
- Search results
- Book-bag
User Study with goal/non-goal oriented tasks

- Welcome
- Informed Consent
- Background
- Pre-Task Information
- Task
- Post-Task Questions
- Experience
- Thank you
### Table 2: Statistics over systems and tasks

<table>
<thead>
<tr>
<th>Session Length</th>
<th>Goal-oriented</th>
<th>Non-goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>6:25</td>
<td>(3:42) 3:42 (3:45)</td>
</tr>
<tr>
<td>Multi-Stage</td>
<td>3:35</td>
<td>(4:24) 2:40 (6:21)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Queries</th>
<th>Goal-oriented</th>
<th>Non-goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>4 (5.5)</td>
<td>2 (4.5)</td>
</tr>
<tr>
<td>Multi-Stage</td>
<td>3 (2.75)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Books Viewed</th>
<th>Goal-oriented</th>
<th>Non-goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>4 (5.5)</td>
<td>2 (4.5)</td>
</tr>
<tr>
<td>Multi-Stage</td>
<td>3 (2.75)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Books Collected</th>
<th>Goal-oriented</th>
<th>Non-goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>3 (3)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Multi-Stage</td>
<td>3.5 (3)</td>
<td>2 (3)</td>
</tr>
</tbody>
</table>

Overall: less queries, more exploration, more books collected in *multi-stage UI*
Wrap Up (II)

- SBS as playground to study (aspects of) non-classic IR
  - Traditional vs. UGC
  - Rich context of user and request
  - Mixing search and recommendation
- Bridge system-centric and user-centric research (iSBS)
  - Running at CLEF, ECIR WS, RecSys WS in 2015 (tbc)
Take Home Messages

- Information access problems are more relevant than ever
  - Classic abstraction important, but limited
  - Current systems framed by past: changes have just begun!
  - Many new opportunities: best time ever to do a PhD in IR!
- IR revolution with the impact of ‘cranfield’ is happening
- Not “anything goes”
  - Scientific understanding requires generalization of results
  - Abstraction to research task, focus on 1 aspect crucial
  - Science should lead and not follow industry…
SBS/iSBS continues @ CLEF in Toulouse
Nov 7: Join the discussion @ ESAIR/CIKM

**ESAIR’14: Seventh International Workshop on Exploiting Semantic Annotations in Information Retrieval**

<table>
<thead>
<tr>
<th>Workshop Homepage</th>
<th>Call for Papers</th>
<th>Program</th>
<th>Organizers</th>
</tr>
</thead>
</table>

**Introduction**

There is an increasing amount of structure on the Web as a result of modern Web languages, user tagging and annotation, emerging robust NLP tools, and an ever growing volume of linked data. These meaningful, semantic, annotations hold the promise to significantly enhance information access, by enhancing the depth of analysis of today's systems. Currently, we have only started exploring the possibilities and only begin to understand how these valuable semantic cues can be put to fruitful use. To complicate matters, standard text search excels at shallow information needs expressed by short keyword queries, and here semantic annotation contributes very little, if anything.

**Articulate Queries and Query Auto Suggest**

The goal of the ESAIR’14 is to advance the general research agenda on this core problem, with an explicit focus on two of the most challenging aspects to address in the coming years.

- First, there is a need to explore more articulate queries, with concepts and relations linking their statement of request to existing semantic models as offered by emerging knowledge bases (DBpedia, Freebase).
- Second, there is a need to extend the query suggestion paradigm to dynamically negotiate longer queries exploring powerful new aspects or facets of the underlying information need.

**We Need Help!**

The Workshop will bring together researchers working with semantic annotations, its use cases, its sources (authoring to NLP tools), its users, and its use in DB, IR, KM, or Web research, and work together on one of the greatest challenges in the years to come. We envision a lively and interactive workshop, with the explicit aim to push the boundaries and think outside the box.

**ESAIR history**

Previous ESAIR editions are found at:
As part of the National Institute of Standards and Technology (NIST) Text REtrieval Conference (TREC), the Contextual Suggestion Track deals with complex information needs which are highly dependent on context and user interests.

Track Organizers

- Charles L A Clarke, Waterloo
- Adriel Dean-Hall, Waterloo
- Jaap Kamps, Amsterdam
- Paul Thomas, CSIRO

TREC 2014 Contextual Suggestion Track Guidelines

Submission Validation and Evaluation Scripts

Contexts, Profiles, and Example Suggestions
Become ‘Friend’ event of ACM SIGIR
Looking for a Postdoc!