Recommender systems: enhancing the end-user’s discovery experience

Gerry Grenier
Staff Director, Publishing Technologies
IEEE, Inc.
The mission of online retrieval systems

- Repositories of information that will help users find answers to their questions
- First we provided the content
  - Our work to this point has conquered any temporal-spatial barriers to content
- Second we are now providing tools to enhance and retrieve that content
  - Links, supplementary content etc.
First, a brief look back, in the spirit of the Herman Skolnik Symposium
I was wrong when I thought that the web started it all

- **1963** – Stanford Research Institute (SRI) demonstrates the first online bibliographic search system and subsequently the first online full-text search system
  - A modest number of records – 7!
- **1964** GE-Valley Forge online search system uses Boolean Logic to combine terms in searching for bibliographic records
- **1964** MIT’s TIP (Technical Information Project) introduces stem searching and left-truncation searching
  - Stem searching: COMPUT to retrieve eg Computer, Computational
  - Left truncation: eg: *ase retrieves Amylase, Typosinase
- **1965** TEXTIR (Text Indexing and Retrieval) incorporates synonyms in a system developed for the LAPD to search robbery reports
  - Featured structured report form, fixed field and full text narrative
- **1965** David Lefkowitz and Clarence Van Meter announce CIDS – Chemical Information and Data Systems – designed to input, store and retrieve chemical compounds
Fielded browse, abstract search … and finally a commercial service

- **1965-1966** BOLD (Bibliographic Organization for Library Display)
  - First online search system to retrieve records by classification codes
- **1966** – Lockheed begins research on a system designed to search abstracts and names the project Dialog
  - **1972** – Dialog becomes a commercial service
  - **1988** – Lockheed sells Dialog to Knight-Ridder for $353 million
A proliferation of information services

- SMART (Harvard/Cornell)
- SOLAR (USDA)
- MEDUSA (Univ of Newcastle)
- ICS (North American Rockwell)
- ADAM (Mitre Corp)
- AESOP (Mitre Corp)
- QUOBIRD ((Queens University of Belfast)
- AUTONOTE (U Michigan)
- TIPS (Tulane University)
- BIDAP (Northwestern U.)
Standards however came to the rescue

- **1965 – 1968** The Library of Congress began the MARC I project, followed quickly by MARC II.
  - MARC was designed as way of "tagging" bibliographic records using 3-digit numbers to identify fields. For example, a tag might indicate "ISBN," while another tag indicates "publication date," and yet another indicates "Library of Congress subject headings."
- **1974**, the MARC II format became the basis of a standard incorporated by NISO (National Information Standards Organization)
- **1974** – Charles Goldfarb invents Standard Generalized Markup Language
- **1994** -- ISO releases 12083 SGML DTD for publishers
- **1997** – the W3C releases the XML standard
History teaches us this:

Information science has evolved constantly to provide answers to research questions as the sea of literature grows.
A major challenge then as it is now is the concept of relevance

- Relevance = effectiveness of communication
  - a best match between the search criteria and a set of documents
- Relevance currently is a machine-phenomenon. Search engines analyze documents and assign a score
Relevancy ranking is a cornerstone of any information retrieval system

- Every search engine has its own relevancy ranking algorithm, but generally speaking relevancy is measured by:
  - **Density**: the number of times the word occurs relative to the size of the document.
  - **Proximity**: The occurrence of the terms near one another
Today – machines tell us what is relevant – what role, though do people play?

- Recommendation is a powerful aid in every day decision making
  - People often help each other in various ways to find needed information
- Recommendation engines are not as prevalent in digital libraries as they are in consumer e-commerce
  - Amazon.com pioneered the online recommendation services
- Two categories of recommendations:
  - Content based filtering (a system making a recommendation based upon document analysis)
  - Collaborative filtering (a person or group of persons making recommendations)
Content-based filtering (CBF)

- Content based filtering is commonly used today as recommender systems
  - Essentially. CBF translated means document similarity:
    - Recommendations based on the current document, which is rather common
    - Recommendations based on set of past-read documents by that end user
  - Shortcoming
    - CBF recommends items similar in content to items rated in the past.
    - Most CBF in use today uses implicit ratings
Content-based filtering based on past purchases. Provided by Amazon’s search engine.
Content based filtering. Provided by Amazon’s machine. An analysis of metadata similarities.
Developing Information Tools that Help Solve Engineering Problems

IEEE e-Workshop provides sample views of IEEE Xplore®-related projects in development. Your feedback is very important to help us improve these enhancements before we make them available to our users. Select the Feedback link to provide your comments.

Similar Articles

Enter search criteria and select “Run Search”. Your results will include a link for similar articles.

Enter keywords or phrases, select fields, and select operators

1. dvdm in All Fields
2. AND
3. in All Fields
4. AND
5. in Full Text & All Fields

Run Search  Reset

Publications

Selected publications:
- IEEE Periodicals
- IEEE Standards
- IEEE Conference Proceedings
- IEEE Conference Proceedings
- IEEE Standards

Select database:
- From year: All
- To: Present

Organize results:
- Maximum: 100
- Display: 25 results per page
- Sort by: Relevance
- Order: Descending

Feedback  Privacy  Terms of Use

© Copyright 2005 IEEE - All Rights Reserved
Similar Articles

1. Dispersion effects of optical filter in DWDM millimeter-wave fiber radio systems
   Issue., Volume, 11-13 Sept. 2000
   Pages(2): 133-136
   Abstract | Full Text | PDF | Similar Articles

2. A novel channel selection scheme with optical heterodyne detection using dual-mode local light for DWDM millimeter-wave band radio-on-fiber systems
   Kuri, T.; Killanen, K.; Microwave Photonics, 2002. International Topical Meeting on
   Issue., Volume, 5-8 Nov. 2002
   Pages(2): 29-32
   Abstract | Full Text | PDF | Similar Articles

3. Error-free DWDM transmission of 60-GHz band fiber-radio signals with square-like response fiber Bragg gratings
   Kuri, T.; Killanen, K.; Murashima, K.; Microwave Photonics, 2001. MWPa '01. 2001 International Topical Meeting on
   Issue., Volume, 7-9 Jan. 2002
   Pages(2): 69-72
   Abstract | Full Text | PDF | Similar Articles

4. A DWDM microwave-fiber-radio system by optical frequency interleaving for high spectral efficiency
   Issue., Volume, 7-9 Jan. 2002
   Pages(2): 85-88
   Abstract | Full Text | PDF | Similar Articles

5. Optimized filtering for 40-Gbps-chip-based DWDM transmission systems over standard single-mode fiber
   Issue 7, Volume 15, July 2003
   Pages(2): 1002-1004
   Abstract | Full Text | PDF | Similar Articles
Developing Information Tools that Help Solve Engineering Problems

IEEE e-Workshop provides sample views of IEEE Xplore® related projects in development. Your feedback is very important to help us improve these enhancements before we make them available to our users. Select the Feedback link to provide your comments.

### Similar Articles Search Results


Results for "(ldwrdin-in-metadata)"

Your search matched 211 of 343590 documents. A maximum of 100 results are displayed, 25 to a page, sorted by Relevance, and in Descending order.

#### Article Information

<table>
<thead>
<tr>
<th>View: 1-25 26-50 51-75 76-100</th>
</tr>
</thead>
</table>

1. Dispersion effects of FGB filter and optical SSB filtering in DWDM millimeter-wave fiber-radio systems
   - Kihayama, K., Kuri, T., Oroshara, K., Kambayashi, M., Murakami, K.
   - Lightwave Technology, Journal of
   - [Abstract](#) | [Full Text](#) | [PDF](#) | [Similar Articles](#) | [IEEE JNL](#)

2. Metropolitan WDM: a dynamically configurable ring for the KomNet field trial in Berlin
   - Stoll, D.; Leisching, P.; Bock, H.; Richter, A.
   - Communications Magazine, IEEE
   - Volume 39, Issue 2, Date: Feb. 2001, Pages: 108-113
   - [Abstract](#) | [Full Text](#) | [PDF](#) | [Similar Articles](#) | [IEEE JNL](#)

3. A differentiated optical services model for WDM networks
   - Communications Magazine, IEEE
   - Volume 38, Issue 2, Date: Feb. 2000, Pages: 88-93
   - [Abstract](#) | [Full Text](#) | [PDF](#) | [Similar Articles](#) | [IEEE JNL](#)

4. An approach to DWDM for real-time applications
   - Alosio, A.; Cerinelli, F.; Izzo, V.
   - Nuclear Science, IEEE Transactions on
   - Volume 51, Issue 3, Date: June 2004, Pages: 526-531
   - [Abstract](#) | [Full Text](#) | [PDF](#) | [Similar Articles](#) | [IEEE JNL](#)

5. Ultra-long-haul 40-Gbit/s-based DWDM transmission using optically prefiltered CS-RZ signals
   - Tsutsumi, T.; Agata, A.; Morita, I.; Edozawa, N.; Akiba, S.
   - Selected Topics in Quantum Electronics, IEEE Journal of
   - Volume 10, Issue 2, Date: March-April 2004, Pages: 488-491
   - [Abstract](#) | [Full Text](#) | [PDF](#) | [Similar Articles](#) | [IEEE JNL](#)
Example of document similarity.

Research

The relevance of recall and precision in user evaluation

Louise T. Su *
School of Library and Information Science, University of Pittsburgh, Pittsburgh, PA 15260

*Correspondence to Louise T. Su, School of Library and Information Science, University of Pittsburgh, Pittsburgh, PA 15260

INDEX TERMS
precision • end users • recall • end user searching • evaluation • user satisfaction • relevance • information retrieval • judgment • user expectations • effectiveness

ABSTRACT

The appropriateness of evaluation criteria and measures have been a subject of debate and a vital concern in the information retrieval evaluation literature. A study was conducted to investigate the appropriateness of 20 measures for evaluating interactive information retrieval performance, representing four major evaluation criteria. Among the 20 measures studied were the two most well-known relevance-based measures of effectiveness, recall and precision. The users' judgment of information retrieval success was used as the desired criterion measure with which all other 20 measures were to be correlated. A sample of 46 end-users with individual information problems from an academic environment were observed, interacting with six professional intermediaries searching on their behalf in large operational systems. Quantitative data consisting of values for all measures studied and verbal data containing users' reasons for assigning certain values to selected measures were collected. Statistical analysis of the quantitative data showed that precision, one of the most important traditional measures of effectiveness, is not significantly correlated with the user's judgment of success. Users appear to be more concerned with absolute recall than with precision, although absolute recall was not directly tested in the study. Four related measures of recall and precision are found to be significantly correlated with success. Among these are user's satisfaction with completeness of search results and user's satisfaction with precision of the search. This article explores the possible explanations for this outcome through content.
Sometimes it is more effective than others …
Collaborative Filtering (CF)

- CF combines some content analysis with user opinions about the items
- In essence, CF is the sharing of relevance judgments, or consensus within the field
  - Shortcomings
    - First-rater problem: items need to be rated to be recommended
    - Sparsity problem: a user is likely to rate only a small percentage of available items
Collaborative Filtering

- **Implicit feedback**
  - Popular ranking (adaptive ranking). System associates search terms with followed links, boosting the ranking of the item for that term

- **Explicit feedback**
  - User rates the relevance of a document returned in a search results list
A form of adaptive ranking. Search results based on current search terms. “Other users who searched on a similar term”
A recommendation where people now play a role.
User profiles

- Session profiling
  - A temporary profile enabling context sensitive results for a particular session
    - Results based on users most recent actions
    - A highly relevant experience

- User profiling – creating communities of trusted expert users.
  - One advantage that member societies have is their large base of expert users
User profiles

- Over time a user’s interest file can be informed by:
  - List of interests given via a form
  - Queries submitted to the system
  - Documents rated or viewed
- The recommendation engine can match user profiles – and display the items read by those users
Recommendation Engine Import Interface

In order to harness the various modes of seeding and updating profiles in the system, the Recommendation Engine supports an XML-based Import Interface. In addition to allowing the transactions described in the earlier section as XML elements, the XML Import Interface is extensible to allow enterprise-specific data to be imported into the engine. Thus, the import interface can be used to import customer data from a Customer Relationship Management (CRM) system, employee data from human resource databases, employee profiles from corporate LDAP directories, data submitted via employee forms, job descriptions from organization charts, product metadata from product databases, document relationships from clustering algorithms, query substitutes from natural language processors, etc.

The following code shows how the XML Import Interface represents a sample transaction in the system. The XML shown below was generated by the system due to a user transaction. This transaction has two sub transactions, one targeting the user, using the document and the query as sources; the other sub transaction targets the document with the query as the source.

```xml
<action type="updateEntity">
  <user id="tom">
  <query parser="freetext"> <![CDATA[tree]]> </query>
  <doc id="../doc/htmldoc/k2entgs/k2gsov4.html8samplecoll1"/>
  </user>

  <doc id="../doc/htmldoc/k2entgs/k2gsov4.html8samplecoll1">
  <query parser="freetext"> <![CDATA[tree]]> </query>
  </doc>
</action>
```
Conclusions

- Both content based filtering and collaborative filtering have their place.
- Content based filtering is simply a variant of current relevancy ranking.
- Collaborative filtering leverages the power of the social networks – let’s exploit that power.
  - Privacy issues may limit temper our excitement.
IEEE and Rutgers U collaborate on CF study

- IEEE and the Rutgers School of Information and Library Science will work together to develop a collaborative filtering feature for IEEE Xplore

- Two goals:
  - Give users an alternative to traditional info retrieval tools
  - Inform scholarship on the subject
Further reading

- Charles P. Bourne and Trudi Bellardo Hahn. A history of online information services.