Publications and Research   New York City College of Technology

2007

Recommender Systems in the Online Catalog: The Cornell Experience

Monica Berger
CUNY New York City College of Technology

[How does access to this work benefit you? Let us know!](#)

Recommender Systems in the Online Catalog: The Cornell Experience
Monica Berger, NYC College of Technology

Leading off the afternoon panel of LACUNY Institute 2007 with her presentation, ''This is so Manual!,''Kornelia Tancheva, Director of Collections, Reference, Instruction, and Outreach, Cornell University Library, described how Cornell experimented with incorporating social navigation, specifically an Amazon.com-like recommender interface, into the online catalog. A recommender system, created by data-mining circulation data, allows for more organic resource discovery than the traditional catalog. The title of the talk comes from the following anecdote: a student, not sure how to handle a citation, was asked by a librarian to copy and paste a citation into the catalog. The student's response was, ''This is so manual.'' Students expect library systems to easily lead them to results.

Using a photograph of footprints in the snow, Tancheva described social navigation, a concept originally from urban design and sociology, as a deliberate path where users ''follow other people's trails in space.'' Users also intuit that by following other people trails, the result might be a good outcome; the process of following is also fun. Recommender systems, which are direct or indirect, enable social navigation. Examples of libraries that have used recommender systems are the University of Karlsruhe and University of California, Berkeley's MELVYL Project.

The Cornell project started out with careful consideration of external data including research from the Pew Internet and American Life Project. Many users have utilized Web 2.0 technology such as social tagging and have had social navigation experiences using recommendations from Amazon.com. Amazon's recommendations are based on the preferences of other users whose profiles are similar and ranks results based on popularity. This type of recommender system is used in e-commerce and entertainment. Does this type of social navigation work for scholarship and libraries? Tancheva asked the audience to consider if voting and popularity are appropriate in this context: social navigation theory questions the objectivity of information.

Cornell's project was limited to five years of historical circulation data for monographs current circulation data. Privacy of subjects was a major concern. The study was limited to graduate students because undergraduates are far less likely to choose monographs on their own (monographs largely as required reading) and faculty are more likely to borrow monographs of lesser interest since they tend to own key books in their area. Graduate field of study was a data element and cross-disciplinary use was examined. To tag the monographs by subject, The Hierarchical Interface to Library of Congress Class (HILCC) was used. The HILCC has fewer and larger categories than Library of Congress Subject Headings. The circulation data was ranked by HILCC category and the results illustrate the principle of the ''long tail'': most level one (or broad category) book tags from the HILCC correlated to top 20 lists. Most level two (or more specific) book tags from HILCC did not correlate to top 20 lists and the tags were numerous and scattered.

In sum, most books had low circulation and few books were found to be popular. The quantity of correlations between books was insufficient to create a robust recommender system and Cornell will need to reconsider its approach in the future. Tancheva suggested using data from peer-institutions to expand the project in the future as well as using more descriptive data about Cornell students which would also include undergraduates. Privacy issues would be considered but would be treated less as a concern. User descriptive tagging (folksonomy) and rating would also enrich the data as well as engage users.