Pundit: Intelligent Recommender of Courses

Ankit Ranka[1], Faisal Anwar[1], Hui Soo Chae[1]

[1] EdLab, Russell Hall 5th Floor
525 West 120th Street, New York
NY 10027, USA
ar2816@columbia.edu, fa2227@columbia.edu, hsc2001@tc.edu

Abstract. In this paper the authors describe Pundit, a course recommendation and search tool at Teachers College, Columbia University. The alpha prototype employs a novel combination of data retrieval and data mining approaches to recommend courses to users based on a match between their profiles and course contents. We utilize course management system and library e-reserves data to collect information about course content that is indexed and then matched with user profiles. In conclusion, we define an evaluation function that was used to determine the quality of recommendations.

1 Introduction

Pundit was developed as a course search and recommendation system for students at Teachers College (TC), Columbia University. The following data sources were used to design the Pundit database: 1) the TC Library e-reserve system (DocDel), 2) the TC course management system (Classweb), and 3) the TC Directory. We utilized a content-based approach [5] and information retrieval techniques [4] to build the recommendation system. A ranked list of recommended courses was generated for each user based on: 1) user profiles (e.g., degree status, program of study), 2) user behavior (e.g., courses taken), and 3) an inverted index [2] created for all TC courses.

2 Methodology

Using key words from user profiles [3], we queried the inverted index of all TC courses. Profiles include user's scholarly interests, academic background, employment history, resume, publications, and additional documents provided by users. In this alpha version of Pundit, we used over 150 user profiles from Netposse [1], an academic social networking tool at Teachers College, that helps users to identify student, faculty, and staff who share similar interests and backgrounds.

Pundit first takes extracted keywords from user's Netposse profile. Then the application generates bi-grams that correspond to the extracted keywords. Afterwards, Pundit eliminates those words that are part of bi-grams from the keyword list. Then it issues an OR query containing all the bi-grams and remaining keywords to the index. Apache Lucene's ranking function was used to calculate a score for each course. Lastly, Pundit removes all courses that were previously taken or taught by the user.

3 Evaluation

Our evaluation strategy links every user in the system with every course, s/he has taken at Teachers College. We compared this information with recommendations from Pundit to determine the number of intersections. Before doing this, we removed transcripts and/or other course related information from user profiles to eliminate any bias towards the final results. The data used for evaluating the Pundit methodology consisted of 45,713 course files, 3,403 courses, 35,215 student & faculty, 40 user profiles, 74 degree programs, and 247 department. Figure 1 below shows the results we obtained when 200 recommended courses for 40 faculty member profiles were compared against the courses faculty previously taught.
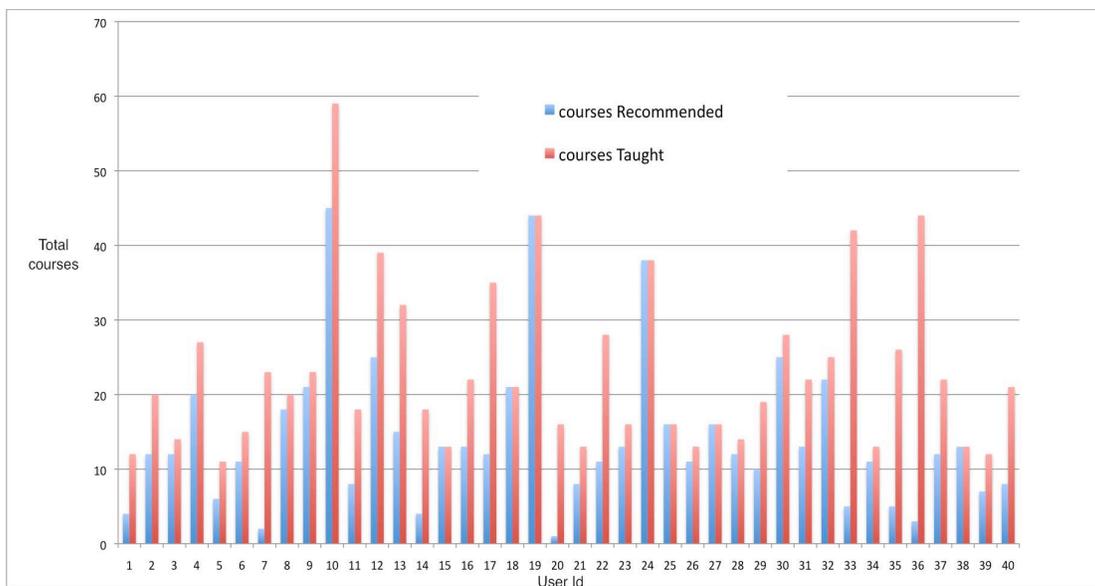

Figure 1. Courses Recommended Vs. Courses Taught

4 References

[1] Anwar, F., Chae, H., Natriello, G. Netposse: A Tool for Connecting Users in Virtual Communities. *Journal of Systems, Cybernetics, and Informatics*, 2008, 8, p. 67-71.
[2] Baeza-Yates, R., Ribeiro-Neto, B. *Modern Information Retrieval*, 1999. New York, NY: Addison Wesley.
[3] Basu, C., Hirsh, H., Cohen W. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *Proceedings of the 15th National Conference on Artificial Intelligence,* 1998, p. 714-720
[4] Manning, C. D., Raghavan, P., Schütze, H., *Introduction to Information Retrieval.* 2008. New York, NY: Cambridge University Press.
[5] Pazzani, M. J., Billsus, D. Content-Based Recommendation Systems. *Lecture Notes in Computer Science*, 2007, 4321, p. 325-341.