

A WEB BASED MANAGEMENT OF REFERENCES

*S. O'Shea, M. Saraee and S. Vadera, School of Computing, Science and Engineering,
The University of Salford, Manchester M5 4WT, UK
Email: Saraee@salford.ac.uk*

ABSTRACT

During the evolution of research from the beginning of a project to the end a large amount of information is accumulated from books, journals, articles, manuals and the internet. Managing all this information is a complex and crucial part, especially since it is important to keep track of and acknowledge the works of those who helped to formulate their ideas and also to show that their work is original. Often these references are made available to groups of researchers from websites. Unfortunately on-line bibliography databases often contain hundreds if not thousands of records. When a user visits these sites they are often overwhelmed with inappropriate information. This means Academic Staff have to consciously search through the references. Academic Staff would prefer to have documents and references automatically recommended to them by the application. The main focus of this work is to design and implement a suitable Recommender System to filter and promote the information that users really wanted, helping like researchers to collaborate together more effectively to make a further contribution to world knowledge.

A fully functional reference management system that incorporated both BibTex upload, download, and restricted security facilities has been implemented. By employing recommendation techniques that had been successfully used by commercial and publication websites like Amazon and CiteSeer we designed and implemented several suggestion systems, based on non personalized, and personalized short-term and long term user interests. The Web Based Reference Management System suggests references to the user through the Statistical Summarization, average user (ratings, reviews, and alternate) recommendations. The final recommendation system implemented provides recommendations for references based on a learned user profile, making use of both hotList and coldList user profiles using a form of incremental Hebbian and Anti Hebbian learning rule to incrementally adapt a feature vector providing a fully functional Content Based Filtering System.

1. INTRODUCTION

An important role of the academics in the computer science discipline in addition to their core responsibilities of teaching is to contribute to world knowledge. They achieve this by carrying out their own or collaborating on various computing projects. The basis for all academic projects involves a substantial amount of research. This initially takes the form of a literature survey to build an understanding over a broad range of subjects and then consequently refining their search to identify more in-depth topics to focus on. During the evolution of their research from the beginning of a project to the end a large amount of information is accumulated from books, journals, articles, manuals, company reports, and the internet. Managing all this information is a complex and crucial part, especially since it is important to keep track of and acknowledge the works of those who helped to formulate their ideas and also to show that their work is original. Often these references are made available to groups of researchers from websites. Unfortunately on-line bibliography databases often contain hundreds if not thousands of records. When a user visits these sites they are often overwhelmed with inappropriate information. This means Academic Staff have to consciously search through the references.

Academic Staff would prefer to have documents and references automatically recommended to them by the application. The traditional methods to achieve this previously might have been to store details about the book or journal that they have read using a manual card index system. Nowadays a more efficient method is to store these entries (called references) on their computer using software known as reference management software (RMS). This software allows them to maintain a personal list of references, which they can quickly search, sort, and insert into their own word documents in the appropriate style for creating bibliographies. The most popular Reference Management Software used by academic institutions is called *EndNote* [21] although on the computer science network a competing application called Papyrus [23]

is promoted. There are many other developers of RMS but they all generally have the same functionality, one of which includes the ability to import references from various formats the most standard being *BibTex* format [24]-[25]. In addition to managing the various sources of information research involves the effective dissemination of ideas and results to others. On-line bibliography/library databases allow academic staff to publish papers in order that their peers, who have access to the World Wide Web, can review and include this knowledge in their work. One of the biggest on-line databases used by the computer science division is called ISI Web of Science [6]. On-line bibliographies allow the user to search for documents or references by providing their search criteria. Once the appropriate reference has been found can the user then download the BibTex format of the reference for importing into their Personal Reference Management Software.

The main focus of this work is to design and implement a suitable Recommender System to filter and promote the information that users really want, helping like researchers to collaborate together more effectively to make a further contribution to world knowledge. This includes

- Design and implement a basic database of references that can be made available over the web.
- Provide a suitable mechanism to allow academic staff to manage the papers they wish to make available.
- Provide a simple way for researchers and academic staff to effectively find documents that they are interested in.
- Implement functionality to allow references to be effectively exchanged between the system and the different user types.
- Research and evaluate the different Methods used by websites to cope with the problem of information overload.
- Identify and implement suitable recommendation systems.

Testing and Evaluation

Due to the incremental nature of the Software Development strategy that we adopted, it was decided to undertake the testing of the system in stages following each successful increment of the system or program component implemented. The first stage of testing involved verifying that the validation mechanism within the Reference Management System would accurately capture input errors made by the user and following this that each of the reference type could be added , deleted and modified from the database correctly. The second phase of testing was carried out after the personal forms of the recommendation systems were implemented to ensure that they worked correctly. This involved in particular ensuring that the similarity recommendation system worked based on a previously downloaded reference and on a learned user profile. This second phase of testing involved selecting references under various subjects to ensure that only the references for a particular subject would be recommended.

Conclusion and Future Work

We successfully developed a web enable reference management system together with providing several types of recommendation techniques identified to help with the problem of information overload. We identified the most effective forms and recommendation systems for the suggestion of references. The decision not to opt for a collaborative based recommendation system was based purely on the fact that it would have required a huge amount of complicated time consuming testing involving perhaps many more users. We also believe that the choice of a simple incremental learning was also justified.

The system could have been made more scalable by building the system on top of the Corba Architecture since processing could be distributed across several machines in the event of future expansion. Naturally the inclusion of a collaborative filtering technique would provide added benefits of being able to recommend references based on the similarity of users' preferences for references. To accompany the inclusion of this additional recommendation technique the use of a more hybrid recommendation algorithm would serve to

better integrate the different sources of recommendation into a single interface. Another good enhancement to the recommendation techniques of the Web Based Reference Management System would allow academic users to share their learned user profile or to provide any profile within the system for selection. This would allow the academic users to access the same recommendations as their colleagues. But this would have the drawback of not protecting the privacy of the users preferences

REFERENCES

- [1] Sommerville, I. (1998), Software Engineering 5th Edition, Addison Wesley, Essex, UK.
- [2] Skidmore, S. (1994), Introducing Systems Analysis, Macmillan Computer Science Series, Hampshire, England
- [3] Pressman, R.S. (1997), Software Engineering A Practitioner's Approach Fourth Edition, McGraw-Hill, USA.
- [4] Bigus, J. (2001), Constructing Intelligent Agents Using Java, Wiley Computer Publishing, New York, USA.
- [5] Flanagan, D. (1996), Java In A NutShell – A Desktop Quick Reference for Java Programmers, O'Reilly & Associates Inc, USA
- [6] Rashid, A.M. Albert, I. Cosley, D., Lam, S.K. McNee, S.M. Konstan, J.A., Riedl, J. (2002), 'Getting to know you: Learning New Users Preferences in Recommender Systems', ,pp. 1-2
- [7] Schafer, J.B. Konstan, J.A. Riedl, J. (2001), 'E-Commerce Recommendation Applications', pp. 5-16
- [8] Good, N. Schafer, J.B. Konstan, J.A. Borchers, A. Sarwar, B. Herlocker. J and Riedl, J. (1999), 'Combining Collaborative Filtering with Personal Agents for Better Recommendations', AAAI, USA,
- [9] Sarwar, B. Karypics, G. Konstan, J. and Riedl, J. (2001), 'Item-Based Collaborative Filtering Recommendation Algorithms', WWW10, Hong Kong, pp. 1-4
- [10] Pazzani, M. Nguyen, L. Mantik, S. (1995), 'Learning from hotlists and coldlists: Towards a WWW information filtering and seeking agent', University of California, Irvine, pp 2-4
- [11] Pazzani, M. Billsus, D. (1999), 'Adaptive Web Site Agents', University of California, Irvine, pp
- [12] Pazzani, M. Muratmatsu, J. Billsus Daniel. (1996), 'Syskill & Webert: Identifying interesting web sites', University of California, Irvine, pp 3-4
- [13] Pazzani, M. (in Press), 'A Framework for Collaborative, Content-Based and Demographic Filtering', University of California, Irvine, pp 2-6
- [14] Pazzani, M. Billsus, D. (1999), 'Evaluating Adaptive Web Site Agents', University of California, Irvine, pp
- [15] Maes, P. Moukas, A. 'An Evolving Multi-Agent Information Filtering and Discovery System for the WWW', Software Agents Group, USA, pp 6
- [16] Maes, P. Shardanand, U. 'Social Information Filtering: Algorithms for Automating Word of Mouth', Software Agents Group, USA, pp 2-5
- [17] Pazzani, M. Billsus, D. (1999), 'A Personal News Agent that Talks, Learns and Explains', University of California, Irvine, pp 4
- [18] Perkowitz, M. Etzioni, O. (1998), 'Adaptive Websites: Automatically Synthesizing Web Pages', AAAI, WA, pp 2-3
- [19] Shepperd, M. 'Case-based Reasoning and Software Engineering', Bournemouth University, UK
- [20] Addyman.A. (2000), CS Research Centre Page, <http://www.cs.salford.ac.uk/csrg/csrc.nsf/index.htm?OpenPage>, [20 October 2002]
- [21] ISI Research Soft. (2001), EndNote Product specifications page, <http://www.endnote.com/en6info.asp>, [22] October 2002]
- [22] Shapland. M. (1999), Reference Management Software Evaluation Page, <http://www.cse.bris.ac.uk/~ccmjs/rmeval99.htm>, [20 October 2002]
- [23] Jacobsen. D. (1996), BibTex Explanation Page, <http://www.ecst.csuchico.edu/~jacobsd/bib/formats/bibtex.html>, [20 October 2002]
- [24] UKOLN Metadata Group. (1998), Further BibTex Explanation Page, http://www.ukoln.ac.uk/metadata/desire/overview/rev_02.htm, [20 October 2002]
- [25] ISI Web of Science. (2002), On-line database of references, <http://www.mimas.ac.uk/wos/>, [20 October 2002]
- [26] University of Minnesota (2002), GroupLens Home Page, <http://www.cs.umn.edu/Research/GroupLens/index.html>, [20 October 2002]

- [28] Netperceptions (2002), Net Perceptions Home Page, <http://www.netperceptions.com/>, [20 October 2002]
- [29] Object Management Group (2002), Corba FAQ Page, <http://www.omg.org/gettingstarted/corbafaq.htm>, [20 October 2002]
- [30] Corba (2002), Corba Success Stories Page, <http://www.omg.org/gettingstarted/corbafaq.htm>, [20 October 2002]
- [31] Jossi. F (2002), 'Net Perceptions Hits Big Time' Article Page, http://www.interactive-pioneers.org/net_perceptions.html, [20 October 2002]
- [32] Bembenick. L (2002), 'A Study of Advertising in Customer Relationship Management From a Personalization Perspective' Article Page, <http://www.andrew.cmu.edu/user/romoff/baseline1.htm>, [20 October 2002]
- [33] InternetNews (1998), 'Net Perception News' Article page, <http://www.internetnews.com/IAR/print.php/12381>, [20 October 2002]
- [34] AhonenMyka, H (2002), 'Processing of large document collections', <http://www.cs.helsinki.fi/u/hahonen/lad02/material/10>, [16 January 2002]
- [35] Shapiro, J (2002), 'The ID3 Algorithm', <http://www.cs.man.ac.uk/~jls/CS2411/id3.pdf>, [16 January 2002]
- [36] Zaniolo, C (2002), 'Classification and Data Mining', <http://www.cs.ucla.edu/classes/spring02/cs240b/presentations/Classifications.ppt>, [17 January 2002]
- [37] Xin, C. (2002), 'Induction of Decision Trees', http://xin.cz3.nus.edu.sg/group/personal/cx/modules/DM/decision_tree.ppt, [17 January 2002]
- [38] Babson College (2002), 'Neural Networks', <http://faculty.babson.edu/reck/mis7570f00/2>, [18 January 2002]
- [39] Queens University(2001), 'What are Bayesian Classifiers ?', http://www.cs.queensu.ca/home/thomas/notes/basic_bayesian_classifiers.html, [18 January 2002]
- [40] Hederman(1999), 'Machine Learning', <http://www.cs.tcd.ie/Lucy.Hederman/DipHI/ML.ppt>, [18 January 2002]
- [41] FreeVBCode(2000), 'ASPUload', <http://www.freevbcode.com/ShowCode.ASP?ID=975>, [15 April 2003]
- [42] The JavaScript Source(2003), 'Cookie Form Saver', <http://JavaScript.internet.com/forms/cookie-form-saver>, [15 April 2003]
- [43] Kaan, A.(1998), 'Using ODBC-JDBC with Java.SQL', <http://home.tiscali.nl/~bmc88/Java/sbook/015.html>, [15 April 2003]