

User Evaluation of the NASA Technical Report Server Recommendation Service

Michael L. Nelson, Johan Bollen
Old Dominion University
Department of Computer Science
Norfolk VA 23529
+1 757 683 6393
{mln,jbollen}@cs.odu.edu

JoAnne R. Calhoun, Calvin E. Mackey
NASA Langley Research Center
MS 157A
Hampton VA 23681
+1 757 864 9475
{joanne.r.calhoun,calvin.e.mackey}@nasa.gov

ABSTRACT

We present the user evaluation of two recommendation server methodologies implemented for the NASA Technical Report Server (NTRS). One methodology for generating recommendations uses log analysis to identify co-retrieval events on full-text documents. For comparison, we used the Vector Space Model (VSM) as the second methodology. We calculated cosine similarities and used the top 10 most similar documents (based on metadata) as “recommendations”. We then ran an experiment with NASA Langley Research Center (LaRC) staff members to gather their feedback on which method produced the most “quality” recommendations. We found that in most cases VSM outperformed log analysis of co-retrievals. However, analyzing the data revealed the evaluations might have been structurally biased in favor of the VSM generated recommendations. We explore some possible methods for combining log analysis and VSM generated recommendations and suggest areas of future work.

Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital Libraries.

General Terms

Measurement, Design, Experimentation, Human Factors.

Keywords

Digital libraries, recommendation servers, user evaluation.

1. Introduction

NASA’s public, web-based digital libraries (DLs) date back to 1993, when a WWW interface was provided for the Langley Technical Report Server (LTRS) [1]. However, it was not until 1995 that the NASA Technical Report Server (NTRS; <http://ntrs.nasa.gov/>) was established to provide integrated searching between the various NASA web-based DLs [2]. It offered distributed searching, mostly through the WAIS protocol [3], of up to 20 different NASA centers, institutes and projects. While NTRS was very successful for both NASA and the public, the distributed searching approach proved fragile. In

developed. The design and development of the OAI-PMH NTRS is detailed in [5]. A new feature of NTRS is a recommendation service. Based on user anecdotes, we knew the recommendation service was well received, but we desired a more quantitative evaluation of its performance.

1.1 NASA Technical Report Server Architecture

The new NTRS offers many advantages that the earlier, distributed searching NTRS did not. NTRS now provides both a simple interface and an advanced search interface that allows more targeted searching, including limiting the number of repositories to search. Syntactic differences between the 20 nodes of the previous version of NTRS made it infeasible to offer anything beyond just a simple search interface. The new NTRS also includes repositories not in the nasa.gov domain: the Physics eprint Server (arXiv), Biomedcentral, Aeronautical Research Council (the UK-equivalent of NASA) and the Department of Energy. The simple search interface searches only the NASA repositories by default. The advanced search interface (which features fielded searching) offers the possibility of including non-NASA repositories. Several other interfaces are provided as well, including: browsing, weekly updates, and administration. NTRS holds over 600,000 metadata records that point to over 300,000 eprints. NTRS averages nearly 30,000 monthly full-text downloads.

NTRS is implemented as a specialized bucket [6], and uses a variety of technologies: the Virginia Tech OAI-PMH harvester, an OAI-PMH repository (thus making NTRS an OAI-PMH aggregator [7]), a MySQL database, the awstats http log analysis facility, and a variety of scripts to integrate the various aspects. Both the user interface and baseURL for harvesters is <http://ntrs.nasa.gov/>. The bucket architecture includes advanced facilities for capturing and sharing logs – a necessary precondition for our recommendation service.

1.2 NTRS Recommendation Service

Taking advantage of the newer, more stable architecture, we added a recommendation server to NTRS in September 2003 (Figure 1). The recommendation server is based on two earlier developed techniques for the implementation of multimedia recommender systems. First we will briefly discuss the algorithm to generate document similarity matrices from user download sequences as reconstructed from NTRS logs. Then we will discuss how such document similarity matrices can be applied to the construction of spreading activation recommender systems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WIDM’04, November 12–13, 2004, Washington, DC, USA.
Copyright 2004 ACM 1-58113-978-0/04/0011...\$5.00.

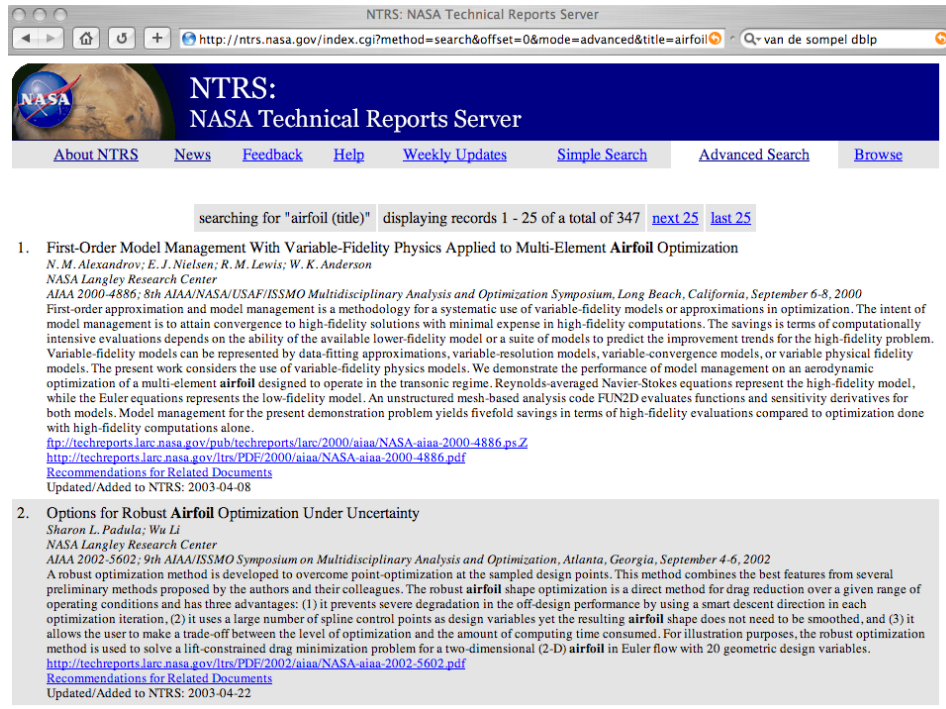


Figure 1. Recommendations Linked From a Search Results Page.

1.2.1 Similarity Matrices

When users download a set of documents, it does not necessarily indicate a stable, continued interest that can be used to build a reliable user profile. However, we may assume that the documents downloaded within a given session, and particularly within a window of three or four document downloads, are more often semantically related than not. A set of documents downloaded by the same user in close temporal proximity therefore does not necessarily indicate the user is permanently and stably interested in these and similar documents, but may indicate the downloaded documents correspond to a common information need and may thus be related or conceptually similar.

We developed a methodology that exploits user download behavior by generating document networks based on the concept of document co-retrieval events. A co-retrieval event is defined as a pair of documents retrieved by the same user in close temporal proximity. Each observed co-retrieval event provides a certain degree of additional support for the belief that the two documents involved may be semantically related. Given that we can reconstruct a set of co-retrieval events from a web log, we can gradually adapt the relationship weights between any pair of documents according to the frequency of their co-retrieval events, or the degree to which users have downloaded the pair of documents in temporal proximity. Such a collection of co-retrieval events reconstructed from a web log may be used to construct a network of weighted document relationships, regardless of the collection's text content or format. This methodology has been tested on hypertext collections and DL journal linking [8] and is similar to that proposed by [9].

The network of document relationships captures the semantic relationships expressed by users in the collective patterns of their document downloads as their download sequences overlap and

gradually update document relationship weights. The production of such networks is computationally highly efficient. Since the generated matrices are commonly highly sparse, sparse matrix formats can be efficiently employed for their storage.

1.2.2 Spreading activation recommender systems

When a network of weighted document relationships has been generated, it can be employed for an information retrieval technique known as spreading activation. Although spreading activation was originally formulated as a model of associative retrieval from human memory [10], it has found applications in IR systems [11]. We have successfully constructed spreading activation recommender systems on the basis of document and journal networks generated from web logs [12].

The process of spreading activation starts from an initial query set, i.e. a set of activated documents that jointly represent the user information need, i.e. a query-by-example principle. Activation is transferred from the query set to all connected documents modulated by the weights of the links connecting them, thereby expanding the initial query set. The total activation imparted on a particular document is defined as the weighted sum of the activations of all documents connecting to it.

This process of activation propagation is repeated k iterations for all documents after which the final activation state of the network is observed. The documents that have received the highest final activation levels are considered to be the most relevant retrieval results. In effect, spreading activation uses the document network and its weighted connections to determine which set of documents best correspond to the user information need by scanning the network for direct and indirect document relationships starting from the initial query set. It can be compared to asking the clerk at your local music store for the CD of a band "that sounds like 'Aphex Twin' meets 'Orbital' meets

‘Squarepusher’.” He or she will look for direct and indirect relationships starting from the mentioned bands to find those that are best connected to all (the band ‘µ-ziq’ is a good recommendation for the above example).

Spreading activation can be simulated by a repeated matrix-vector multiplication. The matrix in question, labeled M , represents the normalized adjacency matrix for the document network. Each entry m_{ij} corresponds to the weight value of the link between documents d_i and d_j . The vector representing the activation state of the network at each time t is labeled a_t . The final activation state of the network can then be calculated as follows. We determine the initial activation state of the network according to which document(s) have been activated by the user, i.e. the user query, resulting in the initial state vector a_0 . The activation state of the network at time $t=1$ is then defined as: $a_1 = f((M + \lambda I) \bullet a_0)$ where λ represents an attenuation value (or function) and I the identity matrix. The procedure can be repeated k times so that a_k represents the final activation state of the network. The set of documents can then be ranked according to the values of a_k to produce a set of recommendations. In this form, the spreading activation procedure indeed defines the activation state of each document as: $d_i = \sum a(d_j) \cdot w_{ij}$, i.e. the weighted sum of the activation values of all documents links to d_i .

The process of spreading activation is attractive for IR applications since it establishes the relevance of a document for a given query according to the overall structure of document relationships that can be defined independent of document content. Spreading activation is thus fit for large-scale DLs with heterogeneous content ranging from text files to multimedia content such as music and movies. Since activation spreads in parallel through the network, it can find pathways between related documents that could not have been identified by term matching or other procedures. Due to its parallel nature it is furthermore resistant to minor errors in network structure that may result from inadequate or missing data. However, it does require the generation of extensive document networks [13] and this has proven to be a considerable hurdle to its general application.

1.3 Related Work

Digital library evaluation is an area of growing interest. Presumably, this stems from the proliferation of DLs and their creators, managers and funding parties wanting to know “what is happening with our DL?” Jones et al. describe transaction analysis of the New Zealand Digital Library and derive interface suggestions from looking at user sessions and common mistakes [14]. Sfakakis and Kapidakis [15] studied the log files of the Hellenic Documentation Centre Digital Library and found users favor simple queries. Steinerova [16] describes DL usage patterns of Slovakian DL users as reported through questionnaires. Assadi et al. [17] describe DL usage patterns in France through surveys, interviews and instrumented DLs. Saracevic [18] offers a broad conceptual framework for DL evaluation.

There is an equal amount of work on the evaluation of recommendation systems. A number of studies have focused on improving existing recommendation systems, such as Efron & Geisler [19] and Hofmann [20] using singular value decomposition to improve recommendations. Kautz, Selman and Shaw [21] describe exploiting existing social networks to increase the quality of recommendations. Sinha and Swearingen [22] compared automated and human generated recommendations and

although human recommendations were preferred, the automated recommendations often provided interesting and novel recommendations. Herlocker et al. [23] provide a comprehensive framework for evaluating recommendation systems.

2. Methodology

We designed an experiment to evaluate the NTRS recommendation server using the domain knowledge of researchers at NASA Langley Research Center (LaRC). Using the terminology from Herlocker et al.’s framework, we designed a “find good items” user task. Twenty documents from the LaRC portion of NTRS were chosen at random. We verified that recommendations were available for these documents through the log analysis-spreading activation (labeled “log analysis”) method. For each of the LaRC documents in NTRS (approximately 4100), we calculated the top 10 most similar documents (according to the Vector Space Model) from the entire NTRS corpus. A program was written on a separate web site to allow the users to self-identify, and then step the user through each of the 20 documents. For each test document, an abstract was shown (with a link to the full-text document) and links were provided for the top 10 recommendations as computed by “Method A” (log analysis) and “Method B” (VSM). Figure 2 shows a screen shot of the data collection web page and the 2 pop-up windows with recommendations from methods A & B. Method A is the recommendation system in use in the production version of NTRS, but it was not described as such to avoid biasing the results.

The data collected for each document included the user’s level of expertise (1..5) with the document’s subject area, and the perceived number (0..10) of “quality” recommendations as generated by each method. A text box was included for any free-form comments the users wished to contribute. Before each session, the users were given a short preparatory presentation that included the history of NTRS, the purpose of the evaluation, and several ways to judge the quality of recommendations. We acknowledged to the participants that “quality” is largely subjective, but nonetheless suggested to consider such factors as:

- *similarity*: documents are obviously textually related
- *serendipity*: documents are related in a way that you did not anticipate
- *contrast*: documents show competing / alternate approaches, methodology, etc.
- *relation*: documents by the same author, from the same conference series, etc.

A call for participation was emailed to targeted organizations and posted on the LaRC intranet to solicit volunteers for one of the four 90 minute sessions, and a total of 13 volunteers were recruited. The volunteer profile is summarized in Table 1, and their organizational (and thus subject area expertise) is summarized in Table 2. The sessions were held at LaRC in a separate computer based training facility away from the participants’ normal offices. No monetary compensation was given to the participants, but refreshments were served. The actual test session was divided into two parts: everyone evaluated the same set of twenty documents, and then after completing that part, they could search NTRS for their own (or “favorite”) document and then rate the recommendations by both methods for those documents. During the first session it became apparent that

the users would not be able to do 20 evaluations in the allotted time, so the test set was truncated from 20 to 10 test documents. Evaluations for 29 documents were generated from the list of 10

standard document and the documents the users found themselves. The subject areas of the 29 documents are summarized in Table 3.

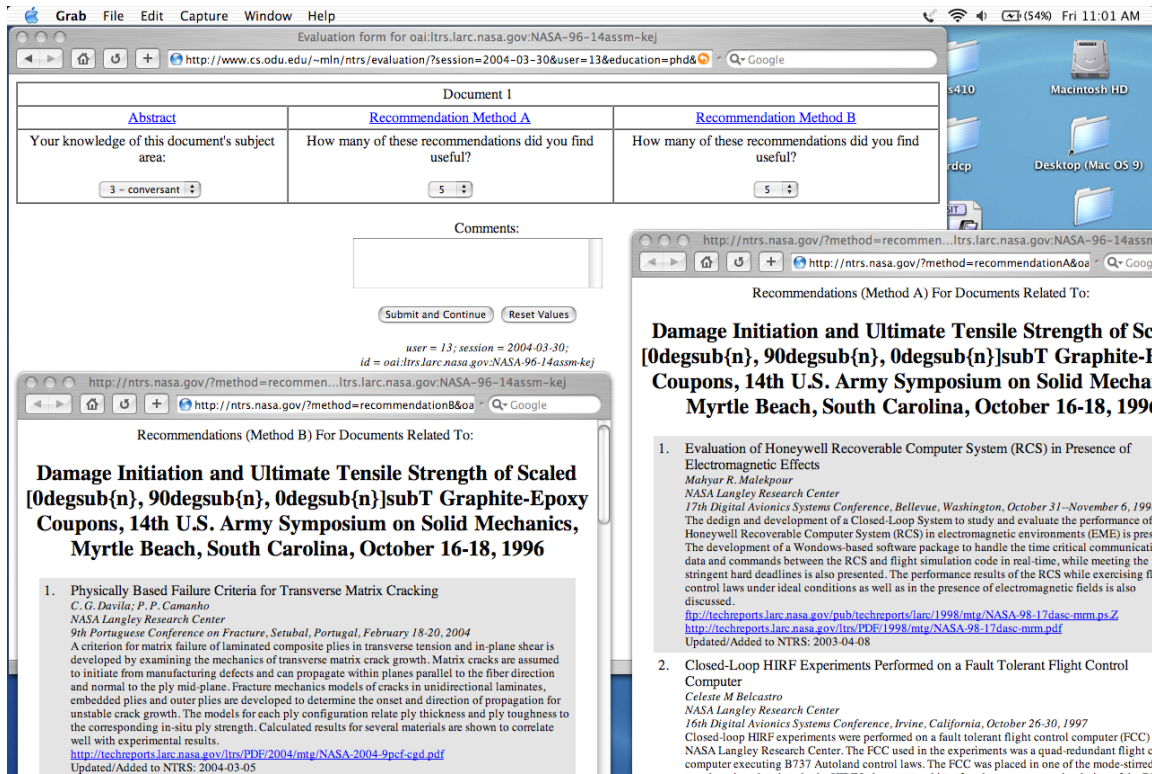


Figure 2. Data Collection and Recommendation Pages.

Table 1. Profiles of the 13 Volunteers.

Highest level of education	11 MS, 1 some graduate, 1 BS
Years of professional experience	Average=16, high=42, low=7
Average number of papers published over the last 5 years	Average=1, high=3, low=0
Experience with NTRS (1=low, 5=high)	Average=3.0
Experience with WWW for research (1=low, 5=high)	Average=3.84

Table 2. Organizational Affiliations of The Volunteers.

#	Competency / Branch
1	Atmospheric Sciences Competency/Radiation and Aerosols Branch
2	Aerospace Systems Concepts and Analysis Competency/Vehicle Analysis Branch
1	Aerospace Systems Concepts and Analysis Competency/Multidisciplinary Optimization Branch
1	Structures and Materials Competency/Nondestructive Evaluation Sciences Branch
1	Office of Chief Information Officer/Library and Media Service Branch
2	Systems Engineering Competency/Data Analysis and Imaging Branch
4	Systems Engineering Competency/Flight Software Systems Branch
1	Systems Engineering Competency/Test and Development Branch

Table 3. Subject Area of the 29 Documents

Subject code	#
Aeronautics	1
Aerodynamics	2
Air Transportation and Safety	1
Avionics and Aircraft Instrumentation	1
Aircraft Propulsion and Power	2
Launch Vehicles and Launch Operations	3
Space Transportation and Safety	1
Space Communications, Spacecraft Communications	1
Spacecraft Design, Testing and Performance	3
Metals and Metallic Materials	1
Fluid Mechanics and Thermodynamics	1
Instrumentation and Photography	2
Structural Mechanics	1
Earth Resources and Remote Sensing	1
Meteorology and Climatology	1
Mathematical and Computer Sciences	1
Computer Programming and Software	3
Solid-State Physics	1
Documentation and Information Science	2

3. Results

The evaluation sessions resulted a total of 129 observations, i.e. individual comparisons of the quality of recommendations issued by method A and method B for a specific document. In total, 129 comparisons pertained to a set of 29 documents. The results were tabulated so that each row of the resulting data set contained the document identifier, the rater identifier, results for

method A, method B and the self-reported level of rater expertise. All ratings were reported on a 10 point scale, 0 corresponding to no qualitatively adequate recommendations, 10 indicating all recommendations generated by a particular method are of high quality. Descriptive statistics were generated over all ratings for method A and B, and are listed in Table 4.

Table 4. Results for Methods A & B.

	method A (log)	method B (VSM)
min	0	1
mean	2.28	6.90
max	10	10
std	2.28	2.35

Table 5. Mean ratings for method A (log analysis) and method B (VSM) for different levels of rater domain knowledge.

Knowledge Level	Method A mean	Method B mean	Number of Raters	A-B
k=1 (lowest)	2.8955	6.4776	67	3.5821
k=2	3.1304	8.0435	23	4.9131
k=3	3.5000	7.3750	16	3.8750
k=4	2.0833	5.9167	12	3.8334
k=5 (highest)	1.4167	7.5833	12	6.1666

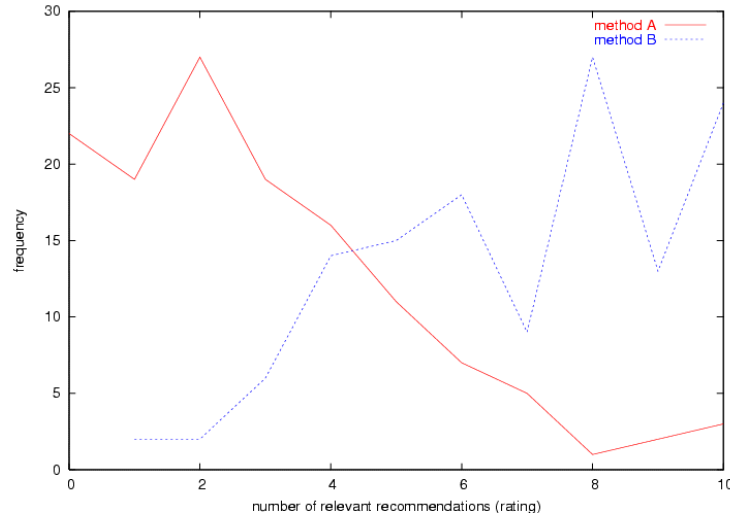


Figure 3. Frequency distribution of ratings for recommendation method A (log analysis) and method B (VSM).

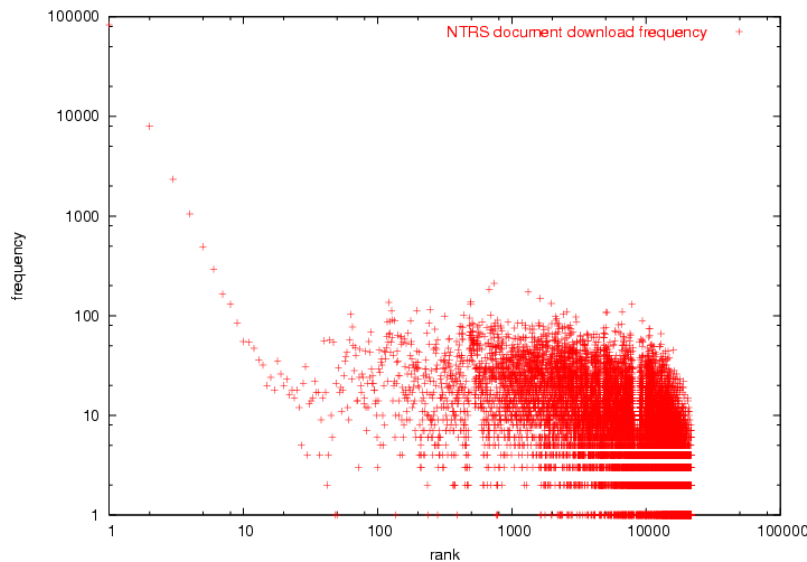


Figure 4. Only the top 20 NTRS downloads seem to follow a power law distribution.

The mean ratings for method A (log analysis) and method B (VSM) were 2.28 and 6.9, respectively, and thus diverged considerably. Method B outperformed method A, but the standard deviations of the distributions of ratings indicate both methods may outperform the other in particular instances.

An ANOVA (analysis of variance) was performed over the distribution of rating values to determine whether their means were significantly different. In this case, the null-hypothesis (the means of the two method ratings are not statistically different) was rejected at the $p < 0.1$ level, indicating the means are marginally different. This result is probably caused by the wide dispersion of the rating values. Figure 3 shows the frequency distribution of the observed ratings for method A and B which indicate that both methods perform at similar levels for a significant number of documents, but that the ratings definitely favor method B over method A.

Since our raters indicated their knowledge level on the document for which recommendations were issued, we

determined the degree of relationship between rater knowledge and recommendations ratings. It is conceivable that more knowledgeable raters prefer one method to the other. For each knowledge level ranging from 0 (layman) to 5 (expert) we determined the mean rating for recommendations issued by method A (log analysis) and method B (VSM). The results are listed in Table 5 and indicate that method B is preferred by raters at all knowledge levels, although roughly 2/3s of raters indicate their domain knowledge is on the layman level. Method A is rated best but still below method B for low knowledge levels. Although both methods are rated lower by expert raters, method B has a strong preference among that group as well.

Spearman correlation coefficients were calculated to determine the degree of relationship between rater domain knowledge and method ratings. The correlation between method A ratings and rater domain knowledge was $\rho = -0.156$ ($df = 129$, $\alpha < 0.1$) indicating there is a weak but marginally statistically

significant negative relationship between rater domain knowledge and method A ratings. For method B and rater domain knowledge we found a Spearman's $\rho=0.12931$ ($df=129$, $\alpha>0.1$) indicating the absence of a statistically significant relationship between rater domain knowledge and method B ratings. Method A and method B ratings over all rater domain knowledge levels was found to be $\rho=0.20100$ ($df=129$, $\alpha<0.05$) indicating a positive and statistically significant relationship between method A and method B ratings. The latter result indicates that some documents cause both methods to produce recommendations that are favorably rated.

An important matter of concern in the use of log analysis to produce document recommendations is the availability of sufficient usage data and its distribution. Document usage can be expected to follow an inverse power law where some documents are retrieved very often and many others only sporadically. However, figure 4 shows that after about the top 20 downloads, NTRS retrieval patterns do not follow a power law. This could be due to robot access to NTRS and only the most popular documents emerge from the cloud of evenly distributed web crawler accesses. The nature of method A's reliance on user retrieval patterns for the generation of document recommendations will naturally lead it to produce a higher number of more valid recommendations for frequently retrieved documents. Our sample of random documents may thus not accurately reflect the performance of method A under realistic circumstances where users would tend to frequently download a particular set of documents for which method A has the highest number of valid recommendations. It forces method A to compete with method B in the absence of retrieval data leading to an invalid comparison on grounds of missing data.

To test this hypothesis we retrieved the absolute number of recommendations available for each document under method. Due to our log analysis method this number corresponds to frequency of use. We then correlated the number of recommendation for each document with its ratings under method A and method B. Indeed, we found a statistically significant relationship between the ratings of method A and the number of available recommendations ($\rho=0.201$, $df=129$, $\alpha<0.05$). This result indicates that as more recommendations are available method A is rated higher. The rater preference of method B over method A can thus largely be caused by the absence of sufficient log data for the documents used in the evaluation. In addition, we found a negative but statistically significant relation between method B ratings and the number of recommendations available for recommendations A ($\rho=-0.32$, $df=129$, $\alpha<0.05$). Since the latter corresponds to document usage, we must conclude that method B produces less valid recommendations for more often retrieved documents. We have not yet formulated a hypothesis to explain this result, but speculate that often retrieved documents may be of a more general nature, and will therefore lead VSM recommenders to discover fewer valid document relationships due to the lack of precise term-relationships. On the other hand, method A relies on actual usage and will simply adopt whichever document relationships are favored by users in their actual retrievals.

To further explore this concept, we determined the ratio of the ratings of method A and method B, a metric indicating how strongly one method is preferred over the other, and the number of method A recommendations available. As expected, we found a strong and statistically significant relation ($\rho=0.384$, $df=129$,

$\alpha<0.01$) indicating that where sufficient log data is available, method A will increasingly improve its recommendations relative to those of B.

4. Future Work

We would like to repeat this evaluation, but with a larger user group. Unfortunately, there are tradeoffs to overcome in increasing the user size. It is hard to attract the people that are most qualified to rate the evaluations for the NTRS content. Since the staff members are U. S. Government employees, monetary compensation for participation is not bureaucratically feasible. We could do away with in person evaluation sessions and automate the process with features attached to the web page, but then we would risk encumbering the entire NTRS and turn away potential users. The intrusive nature of evaluation is well described by Bishop in the evaluation of the DLI system at the University of Illinois [24]. We are considering contacting aerospace undergraduate and graduate classes and incorporating NTRS awareness with a recommendation evaluation. This would result in more, albeit less experienced, subjects for evaluation.

One of the problems with the spreading activation approach to generating recommendations is the latency between the item entering the system and gathering enough downloads in order to increase the quality of the recommendations. This problem is two fold: first, logs are collected from NTRS and processed on a monthly basis, causing at least a 1 month delay before an item can be eligible for recommendations; second, items with "low popularity" (i.e., few downloads) can be in the system for many months before the quality of their recommendations starts to stabilize. We are experimenting with approaches to seed the recommendation process with VSM results, and then let spreading activation guide the recommendation process afterwards.

We are experimenting with approaches to minimize the impact of robots on the system. Large-scale robot downloads of the eprints in NTRS can create and reinforce links that are artifacts of accession order and do not represent semantic relationships. Only one robot, a LaRC robot affiliated with another project, was excluded from generating the current recommendations. While some robots identify themselves through the HTTP_REFERER field, many do not in order to avoid dynamic anti-robot tactics (i.e., "spider traps", HTTP status-code 503 (Service Unavailable), etc.). We will include in our log processing facility the ability to identify and discard repeated, rapid downloads that do not represent interactive user sessions. Another approach to lessen the impact of automated access would be to re-run the test using only the most popular documents (e.g., the first 20 downloads shown in figure 4).

5. Conclusions

We have described NTRS and the architecture of the NTRS recommendation server. We performed a quantitative analysis of the recommendation server and compared against baseline recommendations generated by VSM. The log analysis recommendation method did not perform as well as the VSM recommendation method, but the log analysis method was handicapped by a number of factors. The test documents were chosen randomly from the LaRC portion of NTRS, and not all documents had received enough downloads to have mature recommendations. Secondly, it was difficult to get a good match between document subject area and user expertise. The expertise of the evaluation subjects was largely outside the core focus area

of NTRS. For future evaluations, we will seek out more participants who are better situated to review the subject material, even if they are less experienced. We are investigating methods to lessen the impact of robots and to seed the log analysis recommendations with VSM results.

6. Acknowledgements

This work was supported in part by NASA Langley Research Center and the Los Alamos National Laboratory Research Library. We thank Aravind Elango for his help in preparing the VSM recommendations. We especially thank the volunteers who gave their time for this study.

7. References

- [1] Michael L. Nelson, Gretchen L. Gottlich and David J. Bianco, World Wide Web Implementation of the Langley Technical Report, NASA TM-109162, September 1994.
- [2] Michael L. Nelson, Gretchen L. Gottlich, David J. Bianco, Robert L. Binkley, Yvonne D. Kellogg, Sharon S. Paulson, Chris J. Beaumont, Robert B. Schmunk, Michael J. Kurtz and Alberto Accomazzi, The Widest Practicable Dissemination: The NASA Technical Report Server, Proceedings of Computers in Aerospace 10, AIAA-95-0964, San Antonio TX, March 28-30, 1995.
- [3] Brewster Kahle, Harry Morris, Franklin Davis, Kevin Tiene, Clare Hart, and Robert Palmer, Wide Area Information Servers: An Executive Information System for Unstructured Files. Electronic Networking Research, Applications and Policy, 2(1), 1992, pp. 59-68.
- [4] Herbert Van de Sompel and Carl Lagoze, Notes from the Interoperability Front: A Progress Report from the Open Archives Initiative, Proceedings of ECDL 2002, Rome Italy, pp. 144-157.
- [5] Michael L. Nelson, JoAnne Rocker, and Terry L. Harrison, OAI and NASA Scientific and Technical Information, Library Hi-Tech, 21(2), 2003, pp. 140-150.
- [6] Michael L. Nelson and Kurt Maly, Buckets: Smart Objects for Digital Libraries, Communications of the ACM, 44(5), 2001, 60-62.
- [7] Herbert Van de Sompel, Carl Lagoze, Michael L. Nelson, Simeon Warner, Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting, 2002. Available at: <http://www.openarchives.org/OAI/2.0/guidelines.htm>.
- [8] J. Bollen and R. Luce, Evaluation of digital library impact and user communities by analysis of usage patterns, D-Lib Magazine, 8(6), 2002.
- [9] P. K. Chan. Constructing web user profiles: a non-invasive learning approach. In B. Masand and M. Spiliopoulou, editors, Web Usage Analysis and User Profiling - LNAI 1836, San Diego, CA, August 1999.
- [10] J. R. Anderson, A spreading activation theory of memory, Journal of Verbal Learning and Verbal Behavior, 22, 1983, pp. 261-295.
- [11] F. Crestani and P. L. Lee, Searching the web by constrained spreading activation, Information Processing and Management, 36(4), 2000, pp. 585-605.
- [12] J. Bollen, Group user models for personalized hyperlink recommendation. In LNCS 1892 - International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2000), pages 39-50, Trento, August 2000.
- [13] A. Woodruff, R. Gossweiler, J. Pitkow, E. H. Chi, and S. K. Card, Enhancing a digital book with a reading recommender, Proceedings of the CHI 2000 Conference on Human Factors in Computing Systems, pp. 153-160, The Hague, The Netherlands, April 2000.
- [14] Steve Jones, Sally Jo Cunningham, Rodger J. McNab, and Stefan J. Boddie, A transaction log analysis of a digital library, International Journal on Digital Libraries, 3(2), pp. 152-169.
- [15] Michalis Sfakakis and Sarantos Kapidakis, User Behavior Tendencies on Data Collections in a Digital Library. Proceedings of ECDL 2002, Rome Italy, pp. 550-559.
- [16] Jela Steinerova, In Search of Patterns of User Interaction for Digital Libraries, Proceedings of ECDL 2003, Trondheim Norway, pp. 13-23.
- [17] Houssein Assadi, Thomas Beauvisage, Catherine Lupovici, Thierry Cloarec, Users and Uses of Online Digital Libraries in France, Proceedings of ECDL 2003, Trondheim Norway, pp 1-12.
- [18] Tefko Saracevic, Digital Library Evaluation: Toward an Evolution of Concepts, Library Trends 49(3), 2000, pp. 350-369.
- [19] Miles Efron and Gary Geisler, Using Dimensionality Reduction to Improve Similarity Judgments for Recommendation, Proceedings of the Second DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin City University, Ireland, June 18-20 2001.
- [20] Thomas Hofmann, Latent semantic models for collaborative filtering, ACM Transactions on Information Systems, 22 (1), 2004, pp. 89-115.
- [21] Henry A. Kautz, Bart Selman and Mehul A. Shah, Referral Web: Combining Social Networks and Collaborative Filtering, Communications of the ACM 40(3), 1997, pp. 63-65.
- [22] Rashmi R. Sinha and Kirsten Swearingen, Comparing Recommendations Made by Online Systems and Friends, Proceedings of the Second DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries, Dublin City University, Ireland, June 18-20 2001.
- [23] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John Riedl, Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 22 (1), 2004, pp. 5-53.
- [24] Ann Peterson Bishop, Logins and Bailouts: Measuring Access, Use, and Success in Digital Libraries, Journal of Electronic Publishing, 4(2), 1998.