

## Browsing and Keyword-Based Profiles: A Cautionary Tale

Michael Shepherd, Carolyn Watters, Jack Duffy and Raj Kaushik

*Faculty of Computer Science*

*Dalhousie University*

*Halifax, Nova Scotia, Canada B3H 1W5*

*{shepherd | watters | kaushik} @cs.dal.ca*

*jack.duffy@dal.ca*

### Abstract

*In this research, adaptive user profiles are used to rate Web pages with respect to possible user interest. The user profiles consist of weighted keywords and the adaptation is based on the Hebbian Learning Model with direct user feedback. A user study was conducted to determine if the system would “learn” over a series of five sessions when there was no explicit task other than to browse. The results are analogous to reading the news, i.e., it is not possible to predict what pages a user will read if there is no explicit task. We suggest that a shift away from content to document style (genre) and other user characteristics may be more effective.*

### 1. Introduction

In the context of finding information on the Web, users typically follow a strategy of browsing or searching, alone or in combination. Browsing has been defined as “an exploratory, information seeking strategy that depends on serendipity” [13] and can be characterized by an absence of prior strategic planning. Browsing is often used in new or unfamiliar information spaces, as is typically the case on the Web. Web users rely on mental browsing models based on structures commonly available there: hypertext links and hierarchical directories [17].

Searching and browsing strategies on the Web fit two behavioural theories that have previously been applied to news reading activities: utilitarian or *uses and gratification* and play or *ludenic*<sup>1</sup> [22]. The uses and gratification theoretical perspective is based on the assumption that the reader has some underlying goal, outside the Web activity, that the information sought will satisfy [25], i.e., “serves some ulterior purpose external to

the communication behaviour itself” [6]. Browsing, on the other hand, is a good example of ludenic or play behaviour as introduced by Stephenson [22]. This behavioural theory asserts that, “The process of news reading is intrinsically pleasurable, and that the intrinsic pleasure is at the root of a mature, orderly, and highly ritualized form of news reading as well as a *more casual, spontaneous, and unstructured* form of news reading.”

Two characteristics of ludenic reading behaviour are particularly relevant to Web browsing: individual path selection (convergent selectivity) and “knowing when they see it” (apperception). Convergent selectivity refers to the idiosyncrasy of path selection, so that even among the same document set users choose different articles in different orders, based on individual preference. Apperception is a characteristic of human learning wherein people perceive only those aspects of a complex situation that fit within their current interest and/or their current understanding.

Much of the personal use of the Web is based on browsing or “surfing” activities as a strategy to find information “of interest”. In this research we ask whether the findings on user behaviour reading the news is generalizable to behaviour on the Web. In particular, whether the contents of pages previously visited and rated by the user can be used as a predictor of future choices by that same user. That is, if we analyze the contents of the pages the user picked in one browsing session, can we use this information to predict which pages the user is likely to choose in the next session. The ludenic theory predicts that this will not be effective. In this study, we use a Hebbian learning model [3] for continuous revisions of the user profile. In our case, input is based on feedback from the user rating of the retrieved pages. The occurrence of keywords in those documents is used to update a profile of user interest that can be used by the system to rate new pages in terms of probable interest to the user.

---

<sup>1</sup> Although *ludic* is probably a more correct term, the term *ludenic* was used in the original reference.

## 2. Related Work on Adaptive User Profiles

There are two main variations on building a profile of user interest to assist in the rating or filtering of search results on the Web: indirect (i.e., watch the behaviour of the user) and direct (i.e., ask the user for feedback).

For example, Letizia [11] is a system that tracks the user's browsing behaviour, as the user follows links, makes search queries, or asks for help, and then uses this data to try to predict which pages would be of more interest to the user on the next click. Letizia automates a strategy of recommendation based on a best-first search augmented by inferred user interests based on previous browsing behaviour. The user can follow or ignore the recommendations. This system is not using the content of the pages to build the strategy but the behaviour of the user. No user study has been done to show whether this strategy is effective for the user.

A fair amount of work has been done with systems that recommend a set of Web pages as a guided tour or trail. Systems such as WebWatcher [2] watch the users actions and gain expertise on that part of the Web already visited by the user. The system then recommends sequences or paths of sites to the user based on knowledge of the users previous interests and knowledge of that part of the Web.

The results of Web searches based on known topics of interest can also be ranked according to closeness to a user's preferences. Systems like Syskill & Webert [17] keep a separate profile for each topic for each user. The profile is used to rate pages returned from a Web search. In the Syskill & Webert system, two users rated the returned pages on a three-point scale and this feedback was used to adapt the profiles. Various learning algorithms were evaluated and preliminary results show some improvement as the profiles adapt.

Non-Web-based research based on known topics has been submitted to the TREC [24] series of experiments in the filtering/routing category. For example, Stricker et al. [23] built a specific filter for each topic. Each filter is a neural net classifier trained on sample documents and then used to classifier subsequent documents. This system achieved only modest results. Ng et al. [16] achieved somewhat better results with a perceptron learning algorithm after terms were selected based on Rocchio's [20] formula.

The focused crawling [5] research of the CLEVER Searching project [9] is also based on known topics but is Web-based. This system starts with a canonical topic taxonomy with example Web pages. The user selects and/or refines specific topic nodes in the taxonomy and may provide additional example URLs which serve as starting points for the Web crawl. The user may inspect the system regularly to provide direct feedback by marking pages as useful or not useful. This system has achieved good results based on known topics, taxonomies

of topics, and the incorporation of the link structures into the process.

A different approach has been to discover enough about a user's interests that this can be used to discover information over time for the user. For example, Amalthea [15] works on the results from several search engines and extracts a digest of relevant information for the user. The Informant system [8] works like an SDI (selective dissemination of information) system in that users register queries with the system and these queries are run periodically. Only new results in the top 10 URLs are reported to the user, by email in this case.

A more intrusive use of profiles for users is found in systems like Avanti [7] where the profile is used to change the content and appearance of Web pages for the user using a set of adaptation rules. This has been used for elderly or handicapped users.

Community profiles are also used to provide a shared community level of feedback that can then be used by members of that community. Footprints [26] is an example of this type of system. Visitors can see common paths through a Website as an aid to navigation at that site. Bollen and Heylighen [4] have implemented a collaborative system that develops links structures based on user browse patterns and Hebbian learning. Users were asked to perform "associate" browsing, i.e., they were not given topics, but were given English nouns and asked to browse for associated terms. The browsing patterns were used to build and rank associative links among pages. This work may have a place in adaptive structuring of the Web, but has to be tested on large scale sites.

## 3. WebManthan<sup>2</sup> System

The WebManthan system is designed to be a personalized, on-demand information delivery assistant for users who are browsing. It first registers the user, accepts a root URL as input, then downloads that page and all links from that page recursively for a given number of levels, rates these pages against the user profile, and presents the results in a hierarchical directory presentation for the user (Figure 1). The colour codes indicate the relative ranking assigned by WebManthan. In this study, the system retrieved pages down two levels of the hierarchy and the colour code information was not given to the user.

The system is designed to assist people who are browsing rather than searching for specific information. This means that the person has not articulated a strategy or even a query but will recognize when enough information has been found or their interest wanes. Users usually

---

<sup>2</sup> According to an Indian legend, gods and demons together churned the ocean extracting the elixir of immortality.

combine strategies of short keyword searches, menu selections, and clicking on hypertext links found in documents to find information of interest. Previous studies [10, 19] have shown that users prefer browsing strategies when they are looking for “information about” or general information and that users prefer a keyword search strategy when they are looking for specific information.

WebManthan uses adaptive user profiles, with feedback, to provide information to inform the browsing

process for the user. In the system we are describing, the users interest profile is based on keywords extracted from Web documents deemed interesting by the user. This is based on the observation that although most users have a hard time describing exactly their search interests, they, nonetheless, recognize interesting Web pages when they see them. The user provides feedback on pages in a direct manner. We then use a learning algorithm to combine and aggregate user feedback based on the work of Baclace [3].

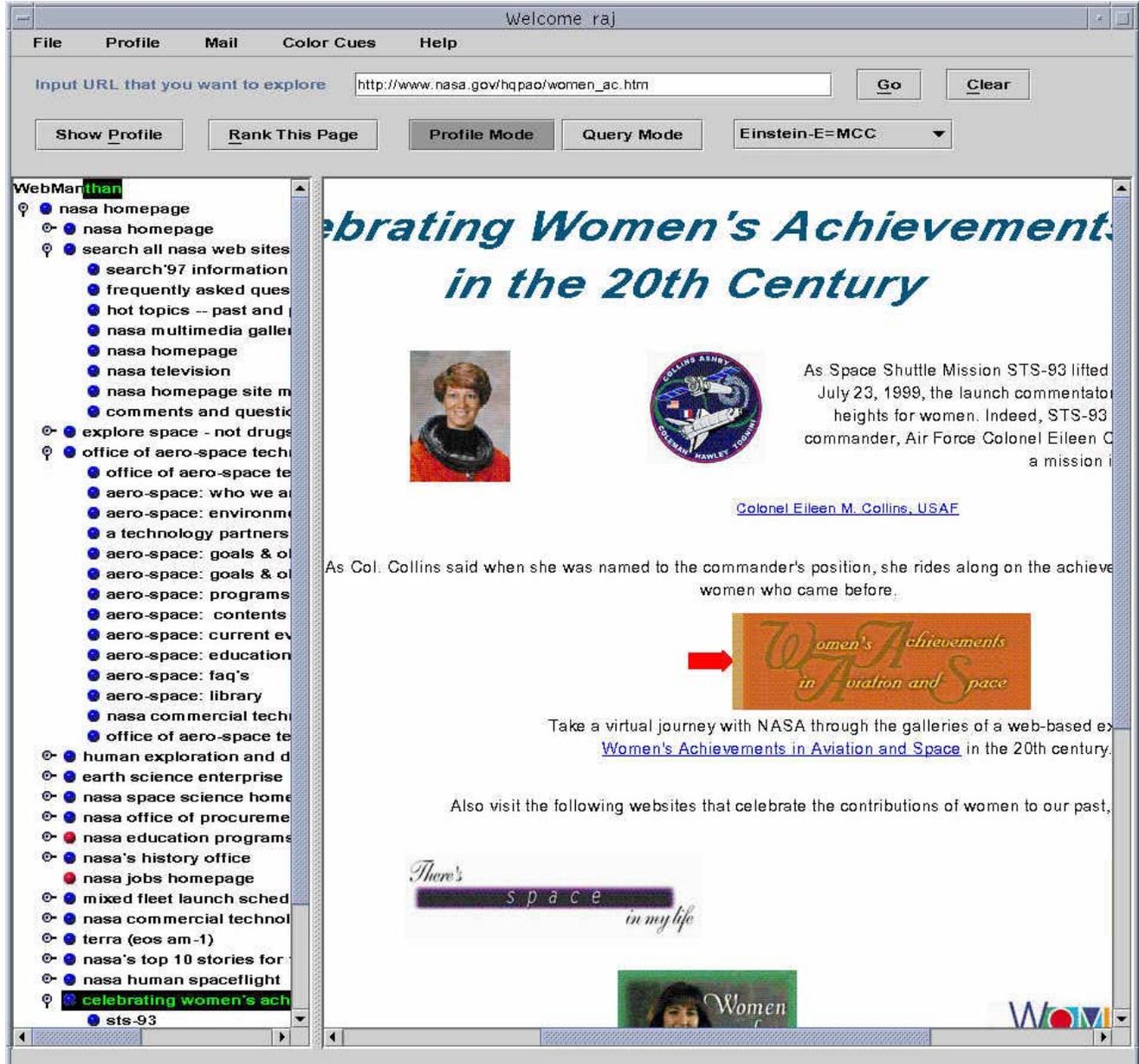


Figure 1. Web page with hierarchical directory.

### 3.1. User Profiles in WebManthan System

Systems that depend on direct feedback from the user to generate/update user profiles of weighted keywords typically generate the keyword set either automatically or directly from the user. If generated automatically, the user rates a document and the keywords are extracted from that document and weights assigned. If generated directly from the user, the user rates the document and submits a set of weighted keywords that represent that document.

Studies [12, 21] have shown, however, that users are not very good at articulating keywords that express their interests and typically enter as few as possible. We suspect this characteristic to be important when the task itself is vague. We have not found, on the other hand, that users have any difficulty in expressing whether a page was interesting or not. Consequently, for the browsing task targeted in this study, the keywords used for the profile are extracted from pages reviewed by the user with weightings based on the perceived interest of the whole page to the user. Ratings collected in this way are minimally articulated, which may be an advantage when individual keyword preferences are difficult to state.

Each profile in WebManthan is represented by a set of term-weight pairs, similar to Baclace's agent [3]. A term-weight is an attribute-value pair, where the attribute is the keyword and the value is its interest weight [0,1], where 0 is of no interest to the user and 1 is of significant interest to the user. The user profile begins with an empty set of term-weights and adapts automatically through user feedback.

The user provides feedback on a Web page using a rank slider, which is a scale  $[-2, 2]$  where  $-2$  is *dislike very much* and  $+2$  is *like very much*. As soon as a user offers feedback on a Web page, the keywords from that page are extracted and a stop list is used to remove <HTML> tags and common words. The resulting keywords are stemmed using Porter's stemming algorithm [18]. Each time the user provides feedback, the profile is updated. If a term extracted from the page is already present in the profile, a new weight for that term is calculated based on the Hebbian Learning Model. If this is a new term, it is defined with an initial weight based on the rating of that page and added to the profile.

A term-weight pair is triggered whenever a term in the user profile also appears in a document evaluated by the user. The term-weight pairs both compete and cooperate in making accurate predictions of the user's actual interest in a given document. The set of term-weights is private to the given user's profile. Each time a user ranks a Web page, a network of keywords from the document activates

the corresponding term-weights in the user profile for revision.

Iterative adjustments of the weight values of the term-weights provides for continual changes in the system performance. The learning rate of the term-weights can be adjusted, from 0 (not learning) to 1 (very rapid learning). In the browsing model we want the term-weights to learn continuously and quickly but with some memory and so the learning rate was set at 0.5, a moderate learning rate. The Hebbian Learning rule causes the term-weight for any keyword occurring in both the document and the profile to be strengthened when both are active at the same time.

### 3.2. Page Rating

Using a URL as the root page, the system follows the links outward for  $n$  levels, rating each retrieved page against the user profile.

WebManthan calculates an overall rating value, called the relevancy index value, [0,1], for each document. The relevancy index value, RI, is computed as the sum of the weights of all of the keywords co-occurring in the profile and the document, normalized by the number of keywords in the document. The page rating algorithm compares keywords in the document with the keywords in the user profile and so building the profile is a critical component. When there are more keywords in the user profile, there is a higher chance that a keyword will be common to both a given document and the user profile.

The value of this relevancy information is used to determine the colour coding associate with that document within an hierarchical directory view of all of the pages below the root URL (Figure 1). This information was not given to the user.

## 4. Evaluation Study

An evaluation study was conducted to see if the approach using profiles based on term-weights derived from the content of Web pages provided useful navigational support for browsing on the Web. Users were asked to use the system five times and complete a questionnaire at the end of each session. Three metrics are considered: how *effective* was the system in recommending relevant (interesting) pages, what was the level of *distrust* the user had in the system ratings, and what was the *correlation* of the system and user rating of pages.

The measure of *effectiveness* chosen was a relatively simple one measuring the agreement of ratings on those pages selected by the users. Effectiveness is the

proportion of the pages rated by the user for which the user and system ratings concur, i.e., the number of pages rated high by both the user and system plus the number of pages rated low by both the user and the system, divided by the total number of pages rated by the user. Note that the system rated all pages while the user was asked to rate at most 20 pages.

Our hypothesis is that effectiveness should increase as the profile is refined. We were also interested in knowing how long a learning period would be required for profile refinement to increase effectiveness.

The measure of *distrust* is an indication of how often the system mis-rated a page (by assigning a low rating when the user rated it high or vice versa). Distrust is the proportion of the pages rated by the user for which the user and system ratings did not concur, i.e., the number of pages rated high by the user but low by the system plus the number of pages rated low by the user but high by the system, divided by the total number of pages rated by the user.

The measure of *correlation* is a summary of the effectiveness and distrust indices combined. It is an indication of how closely the system mimicked the performance of the user in individual ratings for pages rated by both.

#### 4.1. Methodology

The participants were given the general task to “browse a Web site for pages of interest to you”. A session begins with a user entering a root URL of their choice. The system then downloads that page and all links from that page recursively for a given number of levels, rates these pages against the user profile, and presents the results in a hierarchical directory presentation for the user. The colour codes indicate the relative ranking assigned by WebManthan. In this study, the system retrieved pages down two levels of the hierarchy and the colour code information was not given to the user. The user then retrieves and provides feedback on any twenty pages in the Web site displayed by WebManthan. The profile is updated based on the keyword content of those pages and the ratings assigned by the user.

In the study, the participants were asked to rate the retrieved pages after each of five sessions. There were then two sets of rating data for each session: the data generated by the system, and the user’s input on relevance of the pages. We then compared the system and user ratings of the pages. After the fifth session each participant filled out an additional questionnaire.

The proposal for the user evaluation was approved by the Dalhousie University Human Research Ethics Board in accordance with the Dalhousie and TriCouncil *Policy on Ethical Conduct in Research Using Human Subjects*. Participants were recruited from graduate students in

several disciplines across campus. All students were volunteers. The participants engaged in the study at their convenience and could complete the five sessions. All potential participants were given a clear outline of the study and asked to sign an Informed Consent form. All information was kept confidential and participants could withdraw from the study at any time without penalty.

Twenty four graduate students took part in the study from computer science, biology, chemical engineering and psychology. All of the participants were between 21 and 39 years of age. Eighteen were male and six female. Nineteen of the participants completed all five sessions, three completed three or four sessions and one participant dropped out after the first session. All of the nineteen participants who completed all five sessions use the Web on a daily basis, mostly for assignments and entertainment.

#### 4.2. Data Collection

Initially, the WebManthan system has no knowledge about the user. As the user interacts with the system the term-weight population of the user interest profile “learns” and the profile is extended and revised. WebManthan rates Web pages based on its current knowledge of the user through his or her profile. At the end of each session, the user feedback and new keywords are added to the profile and so the profile is populated very quickly. The evaluation sessions were conducted when the profile reached approximately 2000, 3000, 4000, 5000, and 6000 keywords respectively. In each session the user was asked to rank the relevance of the pages presented.

#### 4.3. Results

**User preferences.** The participants stated that in general they do not like getting more than 100 hits (mean 3.68 on a 5 point scale with SD<sup>3</sup> 0.31), reported that they feel frustrated when results include many non-relevant documents (mean 4.47 on a 5 point scale with SD 0.21), and like to know what is ahead (mean 4.47 on a 5 point scale with SD 0.19).

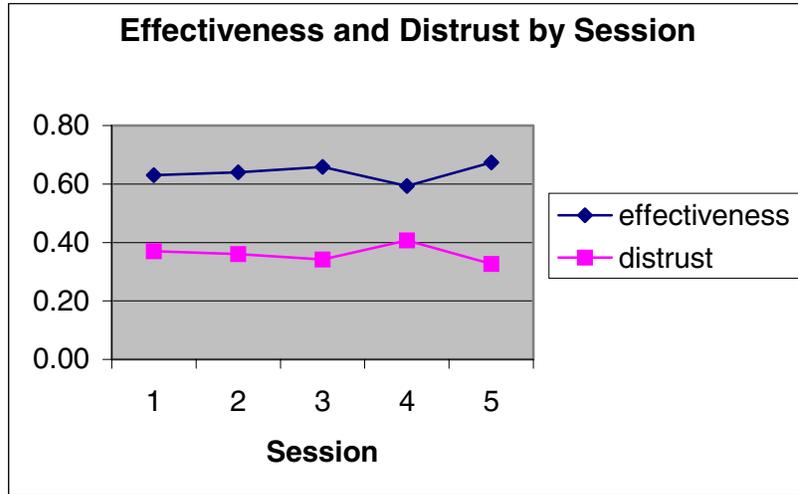
The responses from the users in the study indicated that, overall, they liked the concept of building their personal profile and that the presentation of the structure of the site as a hierarchy was an important factor. Most participants (mean 3.53 on a 5 point scale with SD 0.24) found that understanding the hierarchical relationship of the pages within the site provided a look-ahead function and that this was important (mean 4.47 on a 5 point scale with SD 0.19) in planning a navigation strategy.

---

<sup>3</sup> SD is the standard deviation of the obtained scores per user.

**Effectiveness of ratings.** The effectiveness measure of the WebManthan system is a measure of the ability of the system to predict ratings consistent with the user ratings (high or low). On average, effectiveness was 64% (Figure 2). Using a chi-squared goodness of fit test, the effectiveness did not differ significantly from session to

session at the 0.05 level of significance, ( $\chi^2=2.26$ , table value=9.49, 4 degrees of freedom) . From this we cannot reject the null hypothesis that the adjustments in the profile have no effect on the effectiveness of the system when used for browsing.



**Figure 2. Average effectiveness and distrust measures by session.**

**Distrust measure.** On average, 36% of the ratings of the system did not match the ratings of the users (high or low). This mis-rating of pages, we called *distrust*. The level of distrust was not affected by the number of sessions and remained fairly consistent, at 36%, throughout the evaluation (Figure 2). Using a chi-squared goodness of fit test, the distrust did not differ significantly from session to session at the 0.05 level of significance, ( $\chi^2=3.70$ , table value=9.49, 4 degrees of freedom) .

**Correlation system rating to participant rating.** In this study, the WebManthan system uses explicit links down two levels from the root site to build a tree of page links. The number of pages in the resulting trees varied widely and could be quite high. In the study, two methods were used to evaluate the “goodness” of the rating with respect to the individual user. The participants were asked to provide feedback on any twenty of the pages in the tree for each of the five sessions. At the same time, the system ranked all of the pages using the personal profile. We then determined the correlation of the rating for the twenty

pages chosen by the participant during the evaluation session. In an ideal situation, the correlation would be 1, i.e., the ratings would be the same.

The correlation coefficients for system and user preferences were calculated for each participant over the five sessions (Table 1). The correlation between the system and participant ratings for combined data of the 19 participants who completed all five sessions, shown below, was 0.29 for all sessions. A chi square test indicated that the average relationship (correlation = 0.29) between user and system ratings was significant at 0.05 level ( $\chi^2=140.2$ , table value=16.9, 9 degrees of freedom). In addition, the correlation for each session was also significant. However, a one-way analysis of variance test using all five correlations indicated that there is no significant difference observed among the five sessions. This means that the system and the user have significant agreement on their ratings in each and every session. But the level of agreement does not change from session to session.

**Table 1. Correlation after system and user preferences for each session.**

Session	1	2	3	4	5
Correlation Coefficient	0.30	0.28	0.35	0.26	0.26

One of our hypotheses was that the correlation would improve the longer the user used the system and the profile increased in size and weights adjusted. In fact, no specific pattern was found in the behaviour of the correlation coefficient when plotted against the number of sessions. Rather, the *behaviour was found to be idiosyncratic and personal.*

**Improvements.** One objective of this work was to establish a benchmark for system-based rating of Web pages. During the evaluation study it became clear that several extensions of this approach are in demand by the users: integrating bookmarks (79%), displaying history paths (45%), and restricting paths to explicit links found in the documents rather than including analysis of text of the documents for implicit links.

In addition, there are some obvious improvements or variations of the system that may or may not be worth pursuing to improve effectiveness. In particular, effectiveness may be improved by eliminating pages of no value to the user from the system evaluation, such as, page not found, graphics, incorrect tags in the document, pages requiring login, or pages with very few keywords (e.g., weather page). Variations in the value of the learning rate may also effect outcome.

## 5. Results

The results of this user study indicate that there were no improvements found in the performance of the system in predicting which pages a user would find interesting while browsing based on the content of the documents that the users rated in previous sessions. The keyword content of documents preferred by a user in one browsing session were not shown to be good indicators of documents that will be preferred in a future browsing session. Each session was an independent event as far as the system rating and user rating correlation was concerned. The Hebbian learning model, at least with learning rate set at 0.5, was not effective in providing changes in the profile that improved performance.

The effectiveness of the system, at 64%, was nearly twice as good as the distrust measure, at 36%, during the evaluation sessions. From this we conclude that for the purposes of browsing, the system did mimic the user's behaviour to some degree, but it did not improve as the

user used the system. It is interesting to note that at 64%, the system is in the low end of effectiveness of systems designed to use interest profiles based on page content where the task is more narrowly defined by a topic of interest [17].

Many factors obviously contributed to this result: learning model, learning rate, selection of keywords, feedback method, etc. Nonetheless, the results do indicate that another approach to adaptive user profiles, where the task is browsing, may be more profitable than refinements to this approach.

## 6. We have heard this before

Previous research into the predictive value of content analysis for browsing related to news reading has indicated that we can not predict, on past behaviour, which news items a person is likely to choose. Allen [1] showed that even a person who is very familiar with another's reading habit, is not a reliable judge of which articles that person is likely to read. A study done for news reading using a learning algorithm based on feedback [14] showed initial improvement followed by a period of near total failure. This was understandable in terms of the reader avidly following a "hot" story in the news and followed quickly by disinterest. A user study at Dalhousie [21] used keyword profiles for both the individual and community and found that a significant majority of readers preferred the community profile over their own. A recent study using a learning neural net algorithm based on user feedback to news items again showed no improvement from the user perspective [12].

The study, reported in this paper, on Web browsing indicates that the serendipitous factor that drives ludenic behaviour in news reading may also be dominant in Web browsing behaviour. Just as with the selection of items for news reading, the selection of Web pages based on a user profile of extracted keywords, weighted by the user, from previously viewed documents is not helpful to the user.

## 7. Browsing strategies

If we still want to help users find relevant information when browsing, then we need to examine what data are

used to build the profile. This study and earlier studies have used data from the content of the documents previously chosen by the user to build the profile of interest. This method is an extension of the selective dissemination of information (SDI) systems used for library services. SDI systems are effective when the interest area is focussed and relatively stable. This does not apply to users browsing the Web and content based profiles do not seem to be the appropriate tool for predicting or recommending pages for Web users in browsing mode.

Rather than abandon profiles for use in browsing, we suggest that a shift away from content to document style and user characteristics may be more effective. That is, the profile should be built on user *use characteristics* rather than document content characteristics. Characteristics such as categorization of content, genre of documents preferred, document visit patterns, link topologies, and community use patterns, may be used to profile use characteristics for use in recommending pages in subsequent browsing sessions by the user.

## 8. References

- [1] R. Allen, 1990. "User Models: Theory, Method, and Practise", *Intern. Journal of Man-Machine Studies*. 1990, 3: 511-543.
- [2] F. Armstrong, F. Freitag, T. Joachims, and T. Mitchell. "WebWatcher: A Learning apprentice for the World Wide Web", 1995, [Available at: <http://www.cs.cmu.edu/project/theo-6/Web-agent/www/project-home.html>].
- [3] P. Baclac. 1991. "Personal Information Intake Filtering", *Belcore Information Filtering Workshop*. 1991. [Available: [www.baclac.net/Resources/ifilter1.html](http://www.baclac.net/Resources/ifilter1.html)]
- [4] J. Bollen and F. Heylighen. "Dynamic and Adaptive Structuring of the World Wide Web Based on User Navigation Patterns." [Available at: [www.c3.lanl.gov/~jbollen/pubs/Bollen97.htm](http://www.c3.lanl.gov/~jbollen/pubs/Bollen97.htm)]
- [5] S. Chakrabarti, M. Van den Berg and B. Dom. "Focused crawling: a new approach to topic specific resource discovery", *Proc. of the Eighth International World Wide Web Conference*, 11-14 May 1999, Toronto, Canada. p. 545-562.
- [6] D. Dozier and R. Rice. "Rival Theories of Electronic Newsreading". In *The New Media*, (ed) R. Rice. Sage Publications: London, 1984, p.103-128.
- [7] J. Fink,, A. Kobsa and A. Nill. "User-oriented adaptivity and adaptability in the AVANTI project", *Designing for the Web: Empirical Studies*. Microsoft Usability Group, Richmond. 1996. [Available at: [fit.gmd.de/hci/projects/avanti](http://fit.gmd.de/hci/projects/avanti)]
- [8] B. Gray. *Architecture of the Informant*. 1996. [Available at: <http://actcomm.Dartmouth.edu/~rgray>]
- [9] IBM. <http://www.almaden.ibm.com/cs/k53/clever.html>
- [10] A. Large, J. Beheshti, A. Breuleux and A. Renaud. "A comparison of information retrieval from print and CD-ROM versions of an encyclopaedia by elementary school students", *Inf. Processing and Management*, 1994, (30)4:499-513.
- [11] H. Lieberman and A. Letizia. "An agent that assists Web browsing", *Proc. Of International Joint Conf. On A.I.* August, 1995, Montreal, Canada.
- [12] A. Marath. *Neural Net for Predicting News Items*. M.Sc. Thesis, Dalhousie University, Halifax, Canada,. 2000.
- [13] G. Marchionini and B. Shneiderman. "Finding Facts vs. Browsing Knowledge in Hypertext Systems", *IEEE Computer*, 1988, 21(1): 70-80.
- [14] K. McGillivray. *Personalizing Electronic News through User Modelling*. Honours Thesis. Dept of Mathematics, Statistics, and Computer Science, Dalhousie University, Halifax, Canada, 1995.
- [15] A. Moukas and G. Zacharia. "Evolving a Multi-Agent Information Filtering Solution in Amalthea", *Proc. Of Autonomous Agents 97*, 1997, Marina Del Rey, California, p. 394-403.
- [16] H.T. Ng, H.T. Ang and W.M. Soon. "DSO at TREC-8: A Hybrid Algorithm for the Routing Task". [Available: <http://trec.nist.gov/pubs/trec8/papers/DSO.pdf>]
- [17] M. Pazzani,, J. Muramatsu and D. Billsus. "Syskill & Webert: Identifying interesting Web sites", *Proc. Of the National Conference on Artificial Intelligence*. 1996, Portland, Oregon.
- [18] M. Porter. "An Algorithm for Suffix Stripping", *Program*, 1980, 14(3): 130-137.
- [19] L. Qiu. "Analytical Searching Verses Browsing in Hypertext Information Retrieval Systems", *Canadian Journal of Information Science*, 1993, 18(4): 1-13.
- [20] J.J. Rocchio. "Relevance Feedback in Information Retrieval". In *The SMART Retrieval Systems – Experiments in Automatic Document Processing*, (ed) G. Salton. Prentice Hall, Inc: Englewood Cliffs, NJ, 1971, p. 313-323.
- [21] M. Shepherd, J. Duffy, C. Watters and N. Gugle. "Evaluating the role of user profiles for news filtering", *JASIS*, Accepted May 25, 2000.
- [22] W. Stephenson. *The Play Theory of Mass Communication*. University of Chicago Press: Chicago. 1967.
- [23] M. Stricker, F. Vichot, G. Dreyfus and F. Wolinski. "Two-Step Feature Selection and Neural Network Classification for the TREC-8 Routing". [Available: <http://trec.nist.gov/pubs/trec8/papers/s2n2.pdf>]

[24] TREC. <http://trec.nist.gov/>

[25] C. Watters and M. Shepherd. 1997. "The Digital Broadsheet: An evolving genre", *Proc. Of HICSS'31 Conference*, Hawaii, Jan7-10, 1997. Vol. 7, 22-29.

[26] A. Wixelbalt and P. Maes. "Footprints: visualizing histories for the Web". 1997. [Available at: <http://www.media.mit.edu/people/wex/Footprints/footprints1.html>]