Inferring User Interest

Mark Claypool, David Brown, Phong Le, Makoto Waseda
{claypool,dcb}@cs.wpi.edu

Computer Science Department
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609, USA
Phone: 508-831-5357, Fax: 508-831-5776

August 31, 2001

Abstract

Recommender systems provide personalized suggestions about items that users will find interesting. Typically, recommender systems require a user interface that can determine the interest of a user and use this information to make suggestions. The common solution, *explicit ratings*, where users tell the system what they think about a piece of information, is fairly well-understood and precise. However, having to stop to enter explicit ratings can alter normal patterns of browsing and reading. A less intrusive method is to use *implicit ratings*, where a rating is obtained by a method other than obtaining it directly from the user. This research studies the relationship between various implicit ratings and the explicit rating for a single Web page, and the impact of implicit interest indicators on user privacy. We developed a Web browser that records a user's actions (implicit ratings) and the explicit rating for each page visited. The browser was used by over 70 people that browsed more than 2000 Web pages. We find that the time spent on a page, the amount of scrolling on a page and the combination of time and scrolling has a strong positive relationship with explicit interest, while individual scrolling methods and mouse-clicks are ineffective in predicting explicit interest.

1 Introduction

That we are in the "age of information" is clearly evident from the ever increasing amount of Usenet News, email and Web traffic. It is impossible to access even a small portion of the information generated in a day. We need automated information filters to prioritize information so that we only access information of interest. Since people have different opinions about the importance or relevance of information, personalized filters are needed.

Adaptive filtering and recommendation systems are becoming a significant factor in the economy due to increased use of electronic commerce. It is likely that these techniques will be extended to assist with a wide variety of information using and seeking activities, such as use of electronic libraries, for example. Any improvements in the assistance that such techniques provide to users will have a large impact, that can only grow given increasing computer and Web use.

Filtering/Recommendation systems need to know each user's level of interest in the material currently being examined (the current Web page) so that an accurate profile of the user or the Web page can be built, and so that those profiles can be used for recommendation or filtering. The most common and obvious solution is for the interface to use *explicit ratings*, where users tell the system what they think about some object (e.g., a music CD) or piece of information (e.g., a Newspaper article). Explicit ratings are fairly well-understood and precise, and appear in everyday life due to movie reviews, restaurant "stars", *etcetera*.

However:

- Having to stop to enter explicit ratings can alter normal patterns of browsing and reading;
- Unless users perceive that there is a benefit from providing ratings, they may stop providing them [4]. Hence, users may continue to read, resulting in system use, but no ratings at all [1]; and
- Research on the GroupLens system [12] found that with explicit ratings, users were reading a lot more articles than they were rating.

Hence, explicit ratings, while common and trusted, may not be as reliable as is often presumed. The solution? Use *implicit ratings*. An implicit rating is a rating that is obtained by a method other than obtaining it directly from the user. For example, if the user bookmarks a page, or spends a long time looking at the page, one might infer that the user is interested in the page. Other behavioral signs from the user may be more subtle, such as if the user scrolls down the page it may indicate they are reading out of interest, or they may be skimming and not be interested. Many user actions may be quite unreliable indicators of interest while others may only be reliably interpreted as interest when seen in combination with other indicators.

Advantages of implicit ratings are:

- they remove the cost of the user examining and rating items;
- potentially, every user interaction with the system (and, sometimes, the absence thereof) can contribute to an implicit rating.

Although each implicit rating is likely to be less accurate than an explicit rating, they:

- can be gathered for "free";

- can be combined with other implicit ratings or with explicit ratings for a more accurate rating.

We believe that the capture and use of implicit ratings is a significant problem that has yet to be thoroughly investigated. Combining implicit ratings offers the potential for determining the user's interest in some item or piece of information in situations where the intrusion needed to obtain an explicit rating is either not possible or is not desirable.

In our research, we concentrate on indicators of interest for a single, current page. These indicators might be from a single behavioral sign or from a pattern of behavior. The main objective of the research is to collect, measure and evaluate the predictive power of *implicit interest indicators*: i.e., interest indicators obtained via implicit rating. To accurately gather implicit interest indicators, we developed a Web browser, called *The Curious Browser*, that allows us to capture user actions as they browse the Web. We deployed the browser in a user study with over 70 people browsing over 2000 Web pages.

We analyzed the individual implicit ratings and some combinations of implicit ratings and compared them with the explicit ratings. While there are many implicit interest indicators that warrant study, our goals in this work were to concentrate on several promising implicit interest indicators that could be obtained with our limited resources. We found that the time spent on a page and the amount of scrolling on a page had a strong positive relationship with explicit interest, while individual scrolling methods and mouse-clicks were ineffective in predicting explicit interest. Moreover, implicit interest indicators may be as effective as explicit interest indicators in terms of accurate coverage while having none of the user-costs from explicitly requesting user interest.

The rest of this paper is as follows: Section 2 describes related work in gathering implicit interest indicators; Section 3 describes a general categorization of interest indicators; Section 4 details our approach towards gathering implicit interest indicators; Section 5 describes our user study experiments and results; Section 6 analyzes the results from the experiments; Section 7 discusses issues of personal privacy that come with implicit interest indicators; Section 8 mentions some possible future work; and Section 9 presents our conclusions.

2 Related Research

Nichols [10] discusses the costs and benefits of using implicit ratings for information filtering applications. He categorizes implicit ratings by the actions a user may perform, such as "Examine" for reading a whole item, or "Save" for saving, bookmarking or printing an item. He observes that the limited evidence suggests that implicit ratings may have great potential, but that there has been little experimental work evaluating their effectiveness. Our research provides experimental evaluation of the effectiveness of implicit ratings.

Chan [2] proposes measuring a user's interest in a Web page based on the number of visits to that page, whether or not the page is bookmarked by the user, the time reading the page normalized by the page length, how recently the page was visited and the percentage of links in the page that are followed. While they show some preliminary evaluation of the use of their techniques for recommendations, they do not directly analyze the correlation between their implicit measure of user interest and the users explicit interest.

Morita and Shinoda [9] study the amount of time spent reading a Usenet News article. They examined users in a carefully controlled experimental environment in which users were not allowed to interrupt their reading and only read a carefully chosen news domain. They find that the time people spend reading Usenet News articles is the primary indication of them having interest in it. We extend the study of implicit ratings into a less well-controlled environment, with more types of implicit ratings, to see if their statistically significant results still hold.

Konstan et al [7] describe how the GroupLens system for filtering Usenet News allowed study of the correlation between time spent reading an article and the explicit ratings. They could obtain substantially more ratings by using implicit ratings, and predictions based on time spent reading are nearly as accurate as predictions based on explicit ratings. Our work seeks to extend their experiments into alternative domains, as well as to greatly expand the number of implicit ratings examined.

Hill et al [5] monitor "read" and "edit" actions on a document. The amount of time spent reading or editing an item is termed the "wear" on the item, and is implicitly assumed to indicate interest. However, these implicit ratings were not analyzed to determine how accurately they correlated with interest, but were merely displayed in a scrollbar so that users can infer interest themselves by the "wear" provided by other users. Our work provides a methodology for an analysis of the correlation between such marking actions to user interest.

Lieberman [8] uses different levels of marking to imply different amounts of interest. Lieberman's system Letizia, which works in a Web-based environment, infers that saving a reference to an item implies a strong amount of interest, following a link implies a tentative amount of interest, repeated visits indicate an increasing amount of interest, and passing over a link indicates no interest unless the item is selected later. Our work explicitly measures the level of interest for similar interest indicators.

Kim and Oard [6] provide a framework for considering alternate sources of implicit feedback, and specifically conduct experiments that examine whether reading time is useful for predicting interest of journal articles. In addition, they note printing as an additional implicit interest indicator that enhances predictions of interest made by using reading time alone. While we do not examine printing as an implicit interest indicator, our work measures implicit interest in Web page browsing in general and explores additional implicit interest indicators beyond reading time.

Wittenburg et al [13] focus on tools that monitor users' Web bookmarks to help create an enhanced browsing structure. They present a tool called WebWatch that monitors Web pages of interest and alerts users when significant updates appear. While we do not explore

Figure 1: **Explicit/Implicit Dimension of User Input.**

bookmarks specifically, we present experiments that relate implicit interest indicators to explicit interest.

3 Categories of Interest Indication

Implicit interest indicators can be categorized in a variety of ways. The most basic is to consider them on an Implicit/Explicit dimension, as depicted in Figure 1.

This dimension is based on the time at which the user provided input (i.e., an action), and on whether, and how much, inference is needed. The time might be "now", at the time of viewing the page (e.g., explicit rating) or earlier (e.g., user provided keywords). By "user input" we mean an action that is intended to indicate interest. An example of Explicit input is "providing a rating", of Mixed input is a match of keywords provided apriori by the user with keywords in a document, and of Implicit input is "time spent reading". While this dimension clearly needs some additional study and refinement (e.g., as it mixes action, intent and inference), another beneficial view is to consider *what* the user's input is.

Figure 2 extends Figure 1 by depicting a two-dimensional categorization of interest indicators. The horizontal axis represents how explicit or implicit the interest indicator is. The vertical axis represents whether the interest indication comes from a characterization of the whole item, such as the entire Web page, or from a characterization of the structure or content of the item, such as the Web page layout or font color. Explicit interest ratings are at the bottom left of the Figure. The implicit interest indicators we propose to measure are in the bottom middle to bottom right of the Figure.

It is worth noting that some indicators may be context sensitive, depending on the user's task/goal (e.g., browsing versus searching), or the "category" of the page: i.e., whether it is a page of links in a menu-style, or just plain text with embedded links. This might effect the importance of links *not* taken. In general, layout has an effect on page function, which affects the user's behavior.

In addition, different combinations of indicators might mean different things. For example, if a user does not read a document for very long, but they do bookmark it, the short time indicator might suggest that they do not like the page, while the bookmark indicator might suggest that they do. In this case, they probably bookmarked it for later reading and we do

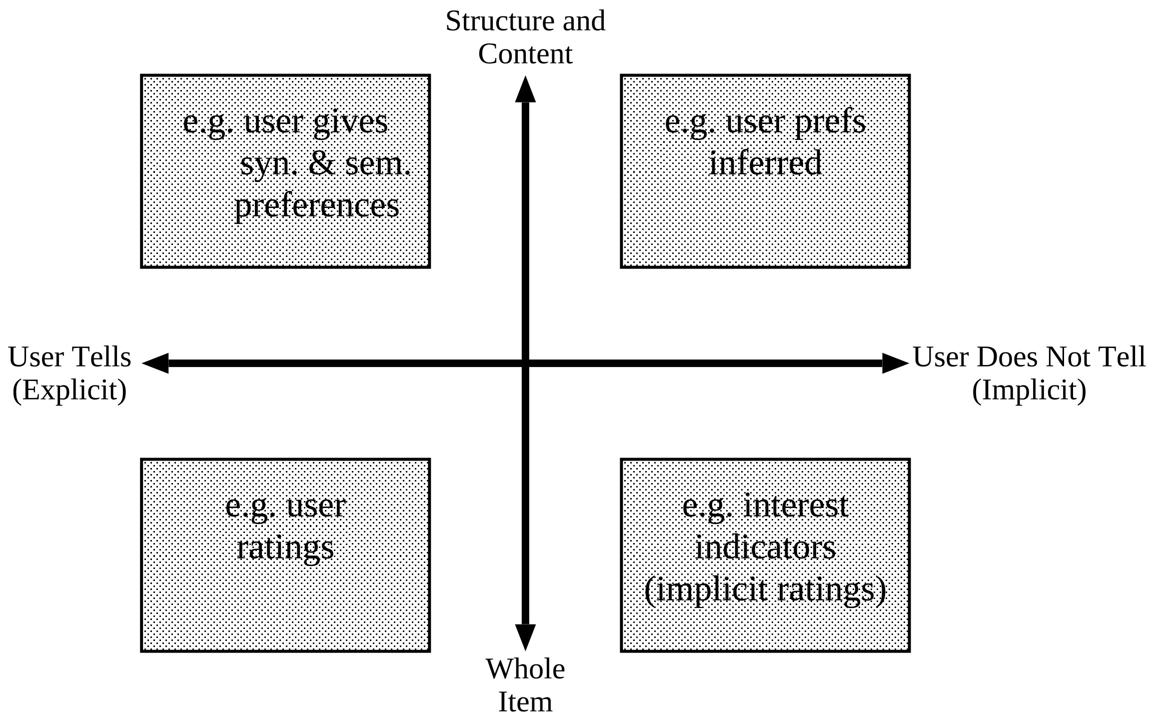

Figure 2: **Categories of Interest Indicators.**

not yet know if they like it or not.

4 Approach

Our approach is to experimentally measure and analyze several promising indicators in order to ascertain their effectiveness in predicting explicit interest. We used the following methodology:

- Implement a browser to gather data about several implicit interest indicators.
- Conduct a user study with many participants browsing the Web using our custom browser.
- Analyze relationship between implicit interest indicators gathered and explicit interest.

This section details the Web browser we implemented, called *The Curious Browser*, to capture some implicit interest indicators from user actions as they browsed the Web. The Curious Browser provides a Graphical User Interface (GUI) that also captures mouse and the keyboard actions as the user browses the Web. The first time each Web page is visited, the Curious Browser stores the user name, the URL, the time and date, the explicit rating and all implicit interest indicators. Subsequent returns to the same page are not recorded.

Figure 3: **The Curious Browser.** This is a screen shot of the main interface, showing WPI's home page.

4.1 Graphical User Interface

The graphical user interface is written with Microsoft's Internet Explorer (version 5.0) in mind, with additional buttons for evaluation, user study instructions, and exiting. Figure 3 shows the main interface of the Curious Browser.

As in normal Web browsing, clicking on a link will load the appropriate Web page. However, before the current Web page is closed, the user is presented with an evaluation window that prompts the user for their explicit rating on the page just visited (see Section 4.5). Figure 4 shows a screen-capture of the evaluation window. The explicit rating is indicated by checking one of five unlabeled radio buttons presented with a scale labeled from "least" to "most" interest. There is a sixth button labeled "no comment" that is the default button selected.

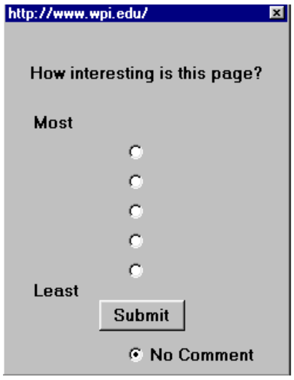

Figure 4: **Explicit Interest Indication Window.** This is a screen capture of the window that pops up for users to give their explicit rating of the current Web page.

4.2 Mouse Activities

The Curious Browser captures two mouse activities: the number of mouse clicks and the time spent moving the mouse, in milliseconds. Mouse activities are only captured when the mouse is inside the browser window and the browser is in focus. The mouse is out of the browser window when the mouse cursor is out of the main HTML page, the vertical scroll bar, and the horizontal scroll bar. The browser window is not focused when a user activates another application. The mouse activities are accumulated for each user while on the page.

4.3 Scrollbar Activities

The Curious Browser captures two kinds of scrollbar activities: the number of mouse events (clicks) on the horizontal and vertical scroll bars and time spent scrolling. Similar to the mouse activities, scrolling activities are only captured when the mouse is inside the browser window and the browser is in focus.

4.4 Keyboard Activities

As some people prefer using a keyboard to scroll instead of the mouse, the Curious Browser captures action on 4 keys: Page Up, Page Down, Up Arrow and Down Arrow. There are two different keyboard activities: the number of times that a user holds down these keys; and the other is the amount of time, in milliseconds, that these keys were held down. The Curious Browser stores the data separately for each key.

4.5 Explicit Ratings

The Curious Browser explicitly asks for ratings (using the window shown in Figure 4) whenever the user changes from one page to another. This is typically done by following a link, but there are also several other ways to change to another page: push the Back button, push the Forward button, or type a URL address directly into the Address Bar and hit the Enter key. The Curious Browser only records one rating per URL, so subsequent visits to the same page do not elicit a request for a new rating. However, the user can also select the Evaluation button at any time to enter an explicit rating, overwriting any old rating that may be present.

5 Experiments

We installed the Curious Browser on about 40 PC's running Microsoft Windows 98 in a computer lab open to all WPI students and in a private computer lab open only to computer science students enrolled in our *Webware* (cs4241) course.

Students from a *Human-Computer Interaction* course (cs3041) as well as students from *Webware* were encouraged to participate in the user study experiments. Students were instructed to open up the Curious Browser and browse the Web for 20-30 minutes, but were not told the purpose of the experiments.

The Curious Browser was available from March 20, 2000 to March 31, 2000. During this time, 75 students visited a total of 2267 Web pages. 72 of the students visited all their Web pages in 1 session, since their 20-30 minutes could typically be done in one sitting, while 3 students had 2 sessions each. The students provided explicit ratings on 80% (1823) of the Web pages (the others were rated "no comment"). Figure 5 depicts a histogram of the explicit rating breakdown. The mean explicit rating was 3.3.

Of the Web pages with explicit ratings, 75% (1366) were from from the `.com` or `.net` domain, 22% (406) were from the `.edu` domain and rest (3%) were from other domains. Web pages local to WPI accounted for 17% (313) of the pages.

6 Analysis

The implicit interest indicators we analyze in this section are:

1. The time spent on a page (Section 6.1).
2. The time spent moving the mouse (Section 6.2).
3. The number of mouse clicks (Section 6.3).
4. The time spent scrolling (Section 6.4).

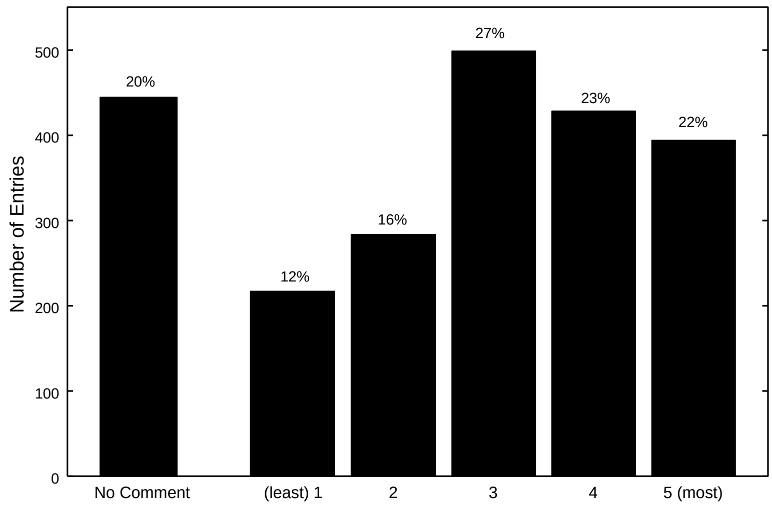

Figure 5: **Explicit Rating Histogram.** This figure shows the number of occurrences of each explicit rating (1 is the least interesting and 5 is the most interesting), along with its percentage of all ratings.

Initially, we analyzed the mean of each implicit interest indicator versus the explicit rating. However, because most of the indicators have a heavy-tailed distribution, the mean of the implicit indicator proved to be a poor indicator of explicit interest. Thus, we focus on the median and distribution of each indicator using a Kruskal-Wallis test[1] (based on .05 level of significance) to examine the degree of independence of the medians among each explicit rating groups for each implicit interest indicator [2]

We present the results with a box-plot, where the box represents the range of values from the bottom quartile (25%) to the top quartile (75%) and the median is depicted by a line in the middle. Although typical box-plots are extended on the top and bottom by two "whiskers" that extend to the full range of values, most of the whiskers are cropped in the figures below.

6.1 Time on Page versus Explicit Rating

The time spent on a page is captured immediately after loading the page until right before the page is exited. It includes all the actions and the actual reading time for the page, but does not include the time that the Curious Browser is not in focus. Thus, factors that influence its accuracy include rendering time (which, in turn, depends upon speed of connection, CPU speed and the amount of Internet traffic) and how much of the active window time the user actually spends looking at the Web page (as opposed to going out for coffee). Before running the test, we filtered out 91 outliers: 4 data points that have more than 1,200,000 milliseconds

[1]Details on the Kruskal-Wallis test can be found in most introductory statistics books.

[2]Details on the test results can on the Web at: `http://www.cs.wpi.edu/~claypool/mqp/iii/`, but are only summarized here due to lack of space.

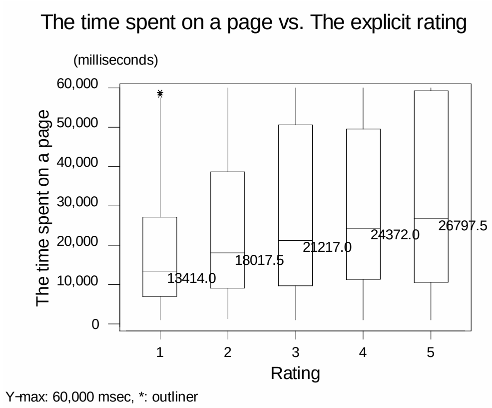

Figure 6: **Time versus Explicit Rating.**

(about 20 minutes) spent on a page as the users had likely stopped reading the page, and 87 data points that had less than 1000 milliseconds (1 second) spent on a page as we believe users cannot accurately assess interest in a page in less than 1 second.

Figure 6 depicts a box-plot of the time spent on a page versus the explicit rating. The Kruskal-Wallis test rejected the null hypothesis (that the median values are the same), meaning that the median values for each explicit rating group differed. Our conclusion is that the total time spent on a Web page is a good indicator of interest. This is a more general result than found in [9] and [12] which showed the correlation between time spent reading News articles and explicit interest.

6.2 Time Moving Mouse versus Explicit Rating

The time spent moving the mouse is measured as the total time the mouse position is changing inside the active browser window. Some users move the mouse while reading the window text or looking at interesting objects on the page, while others move the mouse only to click on interesting links. Either way, we hypothesized that the more mouse movement, the more interesting a user would find the page.

Figure 7 depicts a box-plot of the time spent on a page versus the explicit rating. The results from the Kruskal-Wallis test rejected the null hypothesis, meaning that the median values for each explicit rating group differed. The median for a rating of 1 is significantly less than the median for the other explicit rating groups. The other explicit rating groups (2-5) have only small differences in the median and distribution. Thus, we can observe that the time spent moving the mouse is proportional to the explicit rating. However, they are not linearly proportional to the explicit rating.

Our conclusion is that there is a relationship between the time spent moving the mouse and the explicit rating, but mouse movements alone appear only useful for determining which

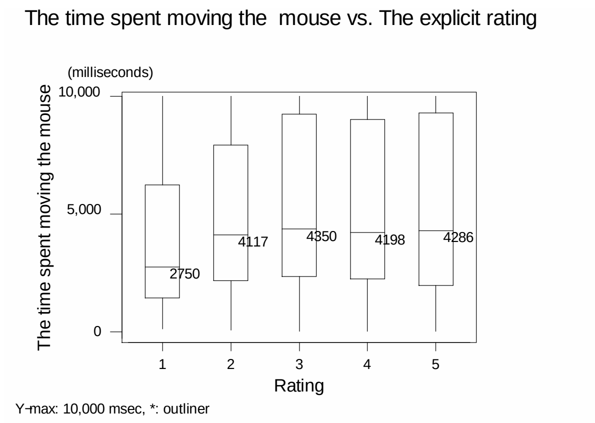

Figure 7: **Time Moving Mouse versus Explicit Rating.**

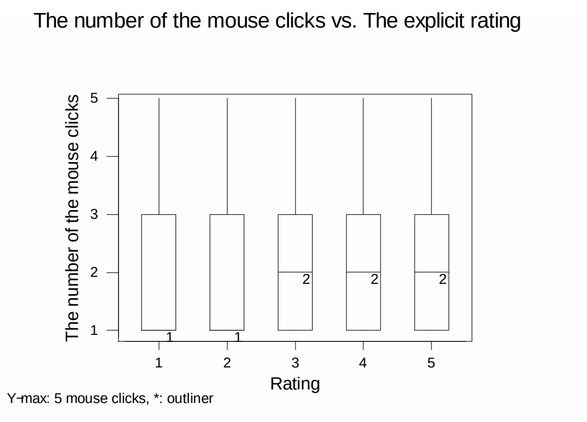

Figure 8: **Number of Mouse Clicks versus Explicit Rating.**

pages have the least amount of interest but are not accurate for distinguishing amongst higher levels of interest.

6.3 Number of Mouse Clicks versus Explicit Rating

Mouse clicking may be a useful interest indicator, too, as users click on links they find interesting (suggesting the current page is a good gateway to interesting sites) and may click on items on the page that look appealing.

Figure 8 depicts a box-plot of the number of mouse clicks versus the explicit rating. The Kruskal-Wallis test failed to reject the null hypothesis, meaning that the median values for each explicit rating group may be the same. Our conclusion is that for this experiment the number of mouse clicks is not a good indicator of interest.

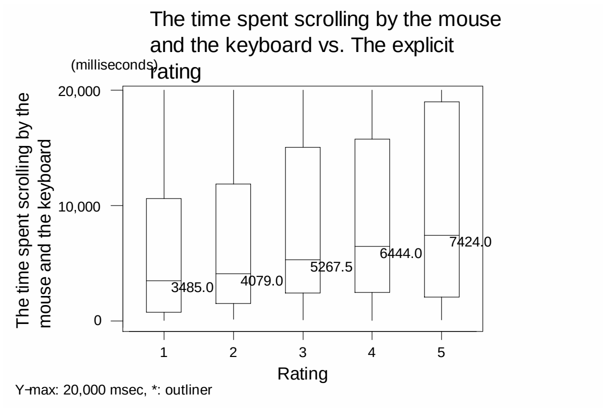

Figure 9: **Combined Scrolling versus Explicit Rating.**

6.4 Scrolling versus Explicit Rating

We hypothesized that users scroll down a page that they find interesting, most likely as they read the material or occasionally as they search the page for interesting links to follow. Users may scroll in a variety of ways: clicking on the scroll bar, clicking and dragging the scrollbar, hitting page up/down keys or hitting up/down arrow keys. Early analysis of each scrolling method by itself revealed them to be poor indicators of interest, probably because most users have one preferred means of scrolling. We then combined some of scrolling methods by adding the time spent on each in an attempt to capture a the "total" scrolling amount.

Figure 9 depicts a box-plot of the time spent scrolling by the mouse and the keyboard versus the explicit rating. The Kruskal-Wallis test rejected the null hypothesis, meaning that the median values for each explicit rating group are different. We conclude that the total time spent scrolling by the mouse and the keyboard is a good indicator of interest.

7 Privacy

As indicated in Section 3, many user actions (and inactions) can provide information about a user's level of interest in a Web page. To a user, this means not only may typical explicit actions be gathered such as purchase history, keyword searches, or Web pages visited, but also many other actions may be gathered such as time spent on a page, amount of scrolling and pages bookmarked. In fact, it is a goal of this research to try to deduce user interest in a Web page implicitly from user actions that are not typically used as interest indicators. Our work has only begun to quantify level of interest from user actions, but if successful, it will identify many user actions that can be gathered non-intrusively to infer interest in online material. This increased understanding of user interest has the potential to greatly improve information filtering systems by supplying them with more accurate user profiles.

Unfortunately, the same implicit interest indicators also lend themselves to increased opportunity for abuse of privacy. The pattern of accesses made by an individual can reveal how they intend to use the information. For example, if a CEO from a company scans the financial statements of a smaller company it could indicate a possible takeover. Or, if an employee spends time reading the job classified section of an online newspaper it could indicate a search for new employment. The input to searches can also be particularly revealing since that provides another source of explicit information. If such inferences are strengthened by similar research to ours, the knowledge gained about user actions, and hence the potential for abuse of privacy, becomes that much greater.

In this work, we developed a customized browser to capture implicit interest indicators. However, instead of requiring users to have a custom browser, embedded Java script could monitor actions and send those in cookies to the Web server. Alternatively, embedded Java script or an active X control could be downloaded and started on the browser's computer and provide the implicit interest indicators studied here, and probably more. A browser's history, hotlists, and cache could be mined for implicit interest indicators.

Unfortunately, even disabling cookies or active X controls is not enough to ensure user actions are not measured. Implicit interest can also be captured at a Web server by analyzing log files. Most Web servers record information to a log file for every access. The log usually includes the IP address or the host name where the browser is running, the time and date, the user's name (if known, possibly from user authentication or obtained by the `identd` protocol), the Web page requested (including any values present from a form), and the size of the data transmitted. Some Web browsers also provide the browser version, the Web page address that the client came from, and the user's e-mail address.

Increasingly, Web browsers are being run from a single-user machine, thus a download can usually be attributed to an individual. Consequently, many implicit interest indicators can be attributed directly to an individual simply by mining a Web log. For example, the time spent reading a particular page can be inferred directly from the log. Thus, most of the extra data gathered by the Curious Browser does not enable any additional abuses of privacy above and beyond the potential abuses from data already gathered by Web servers and proxies.

The above risks to personal privacy are not necessarily different than from information filtering systems that use only explicit interest indicators. Personalized information filtering requires construction of a user profile that accurately represents a user's interest. Much of the research in recommender systems has sought to more accurately capture user interests in order to make better recommendations. Implicit interest indicators are another means to help build a user's profile, providing the potential to capture interest more quickly than explicit interest indicators alone. This, in turn, should yield more accurate recommender systems with less time and effort from the user. Protecting the privacy of a user's profile is crucial, both when using implicit or explicit interest indicators. However, a profile that is enhanced by the use of implicit interest indicators will have more, and more accurate, information about a users interests, making the protection of that profile even more important. In addition, with explicit ratings, a user can choose to give or not to give a rating, while implicit ratings, by definition, are collected without user choice.

Many government sites, such as the U.S. Federal agencies, are not allowed to publish or even collect many types of data about their clients. In most U.S. states, libraries and video stores cannot legally sell or otherwise distribute the checkout history of their customers. While the courts have not yet applied the same legal standard to Internet information, many users have the same expectation of privacy on the Web. Users should expect protection of their privacy in their personal profile, whether that profile is built from implicit or explicit interest indicators.

8 Future Work

In this work, we have considered only implicit interest indicators alone. There are many more implicit interest indicators present in other literature [10, 11], such as bookmarking or printing, that need to be empirically evaluated. Combinations of interest detectors, such as time spent on a Web page *and* the amount of scrolling, may prove to be more accurate than any indicator alone. Implicit interest indication may be combined with more explicit indicators, such as ratings or even purchase history, to provide even more effective interest indication.

Additional analysis tools, such as a rule learner, decision tree or Bayesian network, should indicate stronger positive relationships between implicit interest indicators and explicit interest than those presented in this work.

This suggests searching for a prediction function that accurately predicts explicit interest for a large percentage of users on a large percentage of pages tested. Similarly, there may be a personalized prediction function that can be tailored to an individual user, resulting in a more accurate means of predicting explicit interest.

In this study, we relied upon a large number of users and Web pages to randomize the samples. Clearly, there is a need to study the relationship between implicit interest indication and particular tasks. While our intent here was to establish the relationship between implicit interest indicators and any kind of Web browsing, it may be possible to come up with more accuracy if the test domain is limited to specific types of pages or a specific task. For instance, the correlation between time spent reading a page and a user's interest may be stronger if we know the user's task.

9 Conclusions

Personalized information filtering systems require indication of interest from users. Explicit methods of interest indication, such as asking users to rate the documents they read or evaluate books they have read, intrude upon the normal browsing process and often are ignored by users. Implicit methods, such as the amount of time spent reading a Web page or the purchase of a book, promise to provide more interest indicators without the "cost" to users.

In this research we have categorized and experimentally evaluated the effectiveness of several implicit interest indicators in determining the explicit interest in a Web page. Based on over 40 hours of Web browsing by over 70 students, we find that time is good implicit indicator of interest, while mouse movement and mouse clicks by themselves are ineffective implicit interest indicators. However, using mouse clicks and keyboard actions to infer the level of scrolling, we determined that the "amount" of scrolling does provide an effective indicator of interest.

In this work we consider the information being viewed or read by users to be in the form of Web pages presented by a browser, but in principle the techniques and theories could apply to any computer-based information delivery system. Our results promise to strengthen information filtering systems by clearly directing efforts to include effective indicators and divert effort from indicators that are not effective. The relationships specifically demonstrated may even be directly combined with an explicit measure of interest for future Web-based systems. Other types of personalization, such as adaptive Web sites, might also utilize the results of this research to build more effective systems.

We believe that our work is highly significant because any improvement to information filtering will have significant impact, and because over-reliance on explicit indicators is an important problem that has hardly been addressed by current research. Our approach provides a fresh view of these issues.

10 Notes and Acknowledgments

The *Curious Browser* and the data gathered from our experiments can be downloaded from `http://perform.wpi.edu/downloads/index.html#iii`.

We would like to thank the anonymous reviewers for their extensive comments on an earlier version of this document.

References

[1] C. Avery and R. Zeckhauser. Recommender Systems for Evaluating Computer Messages. *Communications of the ACM*, 40:88 – 89, Mar. 1997.

[2] P. Chan. A Non-Invasive Learning Approach to Building Web User Profiles. In *Workshop on Web Usage Analysis and User Profiling*, pages 7 – 12, 1999.

[3] M. Claypool, P. Le, M. Waseda, and D. Brown. Implicit Interest Indicators. In *Proceedings of ACM Intelligent User Interfaces (IUI)*, pages 33 – 40, Jan. 2001.

[4] J. Grudin. Groupware and Social Dynamics: Eight Challenges for Developers. *Communications of the ACM*, 35:92 – 105, 1994.

[5] W. C. Hill, J. D. Hollan, D. Wroblewski, and T. McCandless. Edit Wear and Read Wear. In *Procedings of ACM CHI Conference on Human Factors in Computing Systems*, pages 3 – 9, 1992.

[6] J. Kim, D. Oard, and K. Romanik. User Modeling for Information Filtering Based on Implicit Feedback. In *Proceedings of ISKO-France*, July5-6 2001.

[7] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3):77 – 87, 1997.

[8] H. Lieberman. Autonomous Interface Agents. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, 1997.

[9] M. Morita and Y. Shinoda. Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In *Proceedings of SIGIR Conference on Research and Development*, pages 272 – 281, 1994.

[10] D. M. Nichols. Implicit Rating and Filtering. In *Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering*, Nov. 1997.

[11] D. Oard and J. Kim. Implicit Feedback for Recommender Systems. In *Proceedings of the AAAI Workshop on Recommender Systems*, July 1998.

[12] B. Sarwar, J. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl. Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, 1998.

[13] K. Wittenburg, D. Das, W. Hill, and L. Stead. Group Asynchronous Browsing on the World Wide Web. In *Proceedings of the World Wide Web Conference*, pages 51 – 62, 1995.