Passive Feedback Collection – An Attempt to Debunk The Myth of Clickthroughs

Christopher C. Vogt
Computer Science, Math and Physics Department
Chapman University, One University Drive
Orange, CA 92866
vogt@chapman.edu
www.chapman.edu/~vogt

Abstract We report the results of a pilot study designed to investigate the feasibility of collecting information about user actions over the Web. By logging simple events (queries, document views, redisplay of query results) and noting their relative timing, we hoped to be able to predict relevance of viewed documents. Although design problems cast doubt on the accuracy of our results, analysis of the cleanest data reveals that clickthroughs are not very predictive of relevance, but that viewing times, when normalized by document length, are somewhat predictive.

1 INTRODUCTION

Collecting feedback regarding the relevancy of documents from users is an expensive process. The amount of time required can be prohibitive for large collections, even if only the top-ranked documents are scored by humans. Furthermore, in operational settings, users rarely want to be bothered by having to explicitly mark documents as relevant or nonrelevant.

The idea that users select documents for viewing which they think are relevant is an attractive one – it makes determining relevant documents a simple matter of noting "clickthroughs." This idea has gained acceptance in the context of the Web and has also been used in any context where determining relevant documents is difficult [Dreilinger and Howe, 1997].

Our hypothesis is that clickthroughs are actually a poor indicator of relevance, and much too coarse to be of real use. Instead, we suggest that a finer view of the user's actions needs to be used, one which takes into account, amongst other things, the timing of the user's actions. For example, a document which is viewed and then immediately discarded seems much more likely to be *not* relevant. Our experiments in the TREC9 interactive track were designed to make a first step towards investigating the idea that a rich transcript of the actions taken by a user might be used to more accurately predict the relevance of the documents they have viewed.

The most relevant experiment relating to our hypothesis comes from [Morita and Shinoda, 1994]. In this study, users of a newsgroup reader were monitored, and their reading time for each news article recorded. Morita and Shinoda found that the length of time spent reading articles was related to how interesting they were, but not related to their length, their "density," or the amount of backlog (unread articles). Their user task, however, was significantly different from the one used for the interactive track at TREC9. Our subjects were not reading the articles for pleasure, but were actively searching for answers to specific questions. Thus, relevance should play a key role in length of time reading. Furthermore, the reason length of an article played such a small role in the Morita and Shinoda study was that most articles in their study were "uninteresting," so the user spent a very small amount of time on them, regardless of length. The articles that users in our study examined all had a reasonably good chance of being relevant (since the user chooses them based on a headline), so the secondary factor of length could come into play.

2 METHOD

The details of the Interactive Track's experimental protocol are described elsewhere in these proceedings. We describe here only those portions of our experimental setup which were left to individual experimenters to interpret.

2.1 System Designs

As required, we had two different search systems. There is only one difference between our two systems (hereafter cleverly referred to as System 1 and System 2), so we will first describe the common aspects of both systems. We proceeded by indexing each of the 6 information sources (AP, FBIS, FT, LA, SJM, and WSJ) using SMART's [Salton, 1971] "ltc" weighting scheme, no stemming, and the standard stop list. Each of these sources was indexed individually. The results list shown to a user was the result of combining the top 30 documents from each of the 6 sources when the user's query was run against those indexes. System 1 merely sorted these 180 (or fewer) documents according to the RSV (retrieval status value) reported by SMART. System 2 first multiplied the RSV by a weight. This weight was the same for all documents from a given source, but varied by query. The weights were calculated based on half of the subjects' experiences using System 1 in the following manner.

During the first week of the experiment, approximately half of the subjects answered their first four questions using System 1. Using these subjects' answers as a guide, the experimenters assigned all documents that had been viewed to a relevance category (a relevant document being one which directly supported the correct answer to a question). Weights were then assigned to each source on a per-query basis based on the number of relevant (R_s) and nonrelevant (N_s) documents viewed from that source, normalized by the weights assigned to other sources. The initial weight (w_s) given to a source was:

$$w_s = 5R_s - N_s$$

Which was then normalized based on other sources to the final weight, W_s:

$$W_s = 100\left(\frac{w_s - \min}{\max - \min} + 1\right)$$

Where max and min are over w_s from all sources. Table 1 shows the final weights, which were used as multipliers on all documents from a given source on a given question.

2.2 Subjects

The experiment was run entirely over the web. Subjects logged on to a password controlled web site using whatever configuration they had at their disposal. As such, the testing conditions for each subject were not uniform – each used a different computer in a different setting with a different type of internet connection. Users were all volunteers – about half from the greater Chapman community, and half being personal friends of the experimenter. Education level varied from high school diploma through Ph.D. Although over 32 people volunteered, only 30 actually logged any time, and only 25 attempted all 8 questions (with one more attempting 7).

Users were divided into two groups: Group 1 and Group 2. The timeline of the experiment was as follows. During the first week, users in Group 1 completed four questions using System 1. Their results were then used to create System 2, which users in Group 2 used during the second week to answer 4 questions. During the third week, all users answered their remaining 4 questions using the system they had not used for the first four questions. Not all users in all groups adhered to this schedule strictly. Users were allowed to log their sessions at times convenient for them.

2.3 Logging User Actions

One of the main goals of this experiment was to investigate the feasibility of monitoring user actions from afar. To achieve this goal, the system was designed as a collection of CGI scripts (CSH and Perl). User actions would thus be automatically recorded via the HTTP server used to activate these scripts. The the user interface was designed so as to try to maximize the number and kind of events that would be recorded, without interfering too much with usability. It was hoped that the relevance of a document could be inferred from the pattern of browsing events.

2.4 User Interface

An example of what the user interface looked like can be seen in the Appendices. The system divided the Web browser's working area into two frames. The upper, smaller frame contained a textbox for the user's query and a search button. It also contained a textbox for the user's answer, a drop-down list for answer certainty indication, and a Submit Answer button. The question currently under consideration was also displayed in this frame. This upper frame was always visible during the search process. The lower, larger frame served a dual purpose:

1. It displayed the list of results of a search, which consisted of a sorted list of document IDs (in red) and headlines (hyperlinks), along with how many times each word in the query appeared in that document. This word count was included to make scanning though the list easier (and to let the user do a sort of visual version of AND and OR, which were not part of the query language).

2. When the user requested the full text of a document (by clicking on its headline), this frame would then display the full contents. In this configuration, the lower frame was divided vertically into a left and right subframes. The left subframe was narrow, and contained only a "Back" hyperlink for returning to the results list. The right subframe displayed the contents of the document (again, with document ID highlighted in red to facilitate cutting and pasting it into the answer window).

The full text of the instructions given to the user is also included in an Appendix.

In order to facilitate user-action collection, the browser's navigation buttons were disabled, but only for Netscape users (due only to lack of knowledge on the experimenter's part of how to do so for Internet Explorer). The five minute time limit per question was not strictly

Question → Source	1	2	3	4	5	6	7	8
AP	**200**	**200**	183	112	**200**	**200**	**200**	135
FBIS	100	154	100	100	133	128	100	111
FT	100	145	116	162	133	171	109	100
LA	125	100	183	137	100	100	128	**200**
SJM	175	154	**200**	112	133	128	109	141
WSJ	100	163	**200**	**200**	133	128	152	135

Table 1: Multiplicative Weights for System 2, by Question and Source (largest weights per question in bold)

enforced. Rather, at the five minute mark, a dialog box popped up asking the user to summarize their answer and submit it. For technical reasons, users could actually continue to search after dismissing the dialog box, but in the over 200 searches recorded, this only happened at most 3 times.

3 RESULTS and DISCUSSION

The results and discussion for the experiment is broken down into two parts: first, we will discuss the relative efficacy of the two systems, then we will address the central concern in this study – tracking user actions in an attempt to implicitly identify relevance.

3.1 System 1 vs. System 2

Because System 2 was designed to uprank relevant documents directly (essentially placing documents with the correct answers near the top of the results list for the query formulations used during the first week), our hypothesis is that it will outperform System 1 both in terms of actual answers found by users and also in terms of user satisfaction.

User Performance

We turn our attention first to actual user performance. For comparison purposes, we will use two measures:

1. A binary measure which is 1 for each search session where a user correctly and fully answers the question and provides complete supporting evidence in the form of document IDs, and 0 otherwise.
2. An adjusted measure, which assigns a score as follows:
   - 4 if the user correctly identifies all answers and fully supports them
   - 3 if the user correctly identifies all answers and partially supports them, OR if the user correctly identifies some answers and fully supports them
   - 2 if the user correctly identifies some answers and partially supports them

	Binary		Adjusted	
Question	Sys 1	Sys 2	Sys 1	Sys 2
1	0	.250	.018	.250
2	.071	0	.071	0
3	0	.200	.383	.425
4	0	0	.150	.375
5	.571	.417	.571	.417
6	0	0	0	0
7	.571	.583	.571	.583
8	.133	0	.133	0
Overall	.167	.193	.239	.261

Table 2: Binary and Adjusted Measures per Query and Overall

Table 2 displays these measures on a per-query basis, and also when calculated across all queries. On the binary measure, System 1 does better than System 2 a total of 3 times, and the reverse is also true. Overall, however, System 2 does slightly better. On the adjusted measure, System 2 beats out System 1 on a total of 4 question, whereas the opposite is only true for three questions. Once again, the overall measure favors System 2. So, on both measures, System 2 eeks out a slight lead. These differences still need to be statistically verified.

User Preference

As part of the experiment, users filled out many feedback surveys - one for each question, each system, and an overall survey at the end. In response to the question "Which of the two systems did you like the best overall?" 4 out of 26 subjects liked System 2 better, 1 liked System 1 better, and the rest indicated no preference. When asked "Which of the two systems did you find easier to use?" 3 indicated System 2, 1 indicated System 1, and the rest said no difference. When asked "How different did you find the systems from one another?" 17 found no difference, 2 found little difference, and 6 indicated somewhat of a difference. These results corroborate the performance results from the previous section, indicating that System 2 is only slightly better than System 1.

3.2 Tracking User Actions

On both systems, user's actions were tracked via entries in the HTTP daemon log on the server that implemented the systems. There were three types of salient actions that were recorded:

1. QUERY - when the user pressed the Submit button to issue a query. The user was then shown the results list of document headlines, document IDs, and term counts. The beginning of this list was displayed immediately, and the list continued to grow until all documents were retrieved. This allowed the user to browse the top ranked documents immediately, to facilitate quick searches. This action could be initiated at any time.

2. VIEW - when the user clicked on a document headline in the results list. The frame the results list had been in was replaced with the contents of that document. The user could only perform this action when a results list was displayed.

3. RETURN - when the user clicked the specially generated Back button while viewing a document, thus replacing the document text with the results list from the previous query. The user could only perform this action when a full document was displayed.

Of course, these are only a small subset of all of the user actions that could conceivably be recorded (see for example, [Oard and Kim, 1998]), but represent what was available using the given technology.

Problems

Unfortunately, three design flaws subjected the event logs to a lot of noise when it came time to interpret them. In increasing order of seriousness, they are:

1. Back Button. Although Netscape browsers had the browser back button disabled, Internet Explorer users were free to use it (although they were told not to in the instructions, habits die hard). Thus, the log file could be missing RETURN events. Although it would normally be possible to infer these events, the timing would be unknown, and as it turns out, inferring them isn't possible due to the problems described next.

2. Multiple Processes. Because the user was able to request to VIEW a document before their QUERY had finished processing, it was possible to still have the computer process which was handling the QUERY running at the same time as the process handling the VIEW, which slowed things down a little. This scenario wasn't really a problem, but when the user swapped back and forth between viewing and querying (every RETURN event also initiated a new process to re-generate the results list from scratch), the system could get very bogged down. In and of itself, this would not be a problem for the logging of events, but due to the next problem, it greatly complicated things.

3. End time stamp. Contrary to what one might assume (and what the experimenter did assume), the HTTP daemon does not log the start time of requests, but the finish time. This would not be a problem in a serial, one-process-at-a-time computer. However, when coupled with the previous problem, it nearly rendered the event logs useless.

We have attempted to compensate for the above three flaws in two ways. First, event start times were approximated based on first calculating run-times of processes in standalone mode, and then using a simple algorithm which works backwards from the end of the log, estimating start times based on number of processes running and the standalone runtimes. Second, we combed through the logs and tossed out any of those belonging to any users who exhibited behaviors which resulted in logs that were inconsistent with the constraints above (e.g., ones with two VIEW events in a row or two RETURN events in a row, etc.), on 4 or more of the questions. This left us with the results from 13 users.

3.3 The Myth of Clickthroughs

One of our goals in this experiment was to find a way of inferring which documents were relevant to a query based on the user's actions, without explicitly asking the user for feedback. One common simple assumption is that when a user selects a document for display (an event sometimes called a "clickthrough" in the context of the Web), that document is more likely to be relevant (or, an even simpler version of this heuristic is that the document *is* relevant). In this section, we attempt to debunk that idea. We propose a hypothesis that the amount of time spent reading the document (when adjusted for the document's length) is a more closely correlated with relevance than just whether the document was clicked-through.

In this section, we only analyze the data provided by the 13 "well-behaved" subjects described previously. We are interested in how much time the users spent examining each document that they saw and how that corresponds to relevance. In this section, a document is defined as relevant to a question if the NIST assessors indicated that it supported (at least in part) a correct answer to the question, otherwise it is nonrelevant. As

	Mean	Std Dev
Relevant	0.01808	0.0467
Nonrel	0.00977	0.0159

Table 3: Mean and Standard Deviation of Normalized View Time

a first step in debunking the clickthrough assumption, we note that across all 13 users, a total of 181 VIEW events occurred. Of these, only 66 (or about 36%) were actually for relevant documents. This low number is partially a result of the lack of information the users had to go on: only about 2/3rds of the document's headline was shown. Nevertheless, this argues that there is a very good chance that clicked-through documents are *not* relevant. Perhaps timing information will prove more informative.

Analyzing Normalized View Times

We hypothesize that a user's behavior when reading documents for this task is to scan the document quickly to find an answer or partial answer to the question at hand. They will encounter two kinds of nonrelevant documents, ones which are obviously not relevant and ones which look promising. Relevant documents will also look promising. The amount of time spent on obviously nonrelevant documents will be very small on average. The amount of time spent on promising but nonrelevant documents will likely be a little bit longer than the time spent on relevant documents, because the user will want to scan the entire document to verify the absence of an answer. On the other hand, they will stop scanning a relevant document as soon as they find an answer. Of course, these times will also depend a lot on the length of the document, so we propose normalizing view times by document length.

Thus, we hypothesize that the distribution of normalized view times for nonrelevant documents will be bimodal, with one peak for the obviously nonrelevant and one for the promising but nonrelevant. On the other hand, the distribution of normalized view times for relevant documents will be flatter, and likely have a higher overall mean.

We have calculated the normalized view times by dividing the view time (in seconds) by the number of characters in indexed sections of each document. Table 3 shows the mean and standard deviation for these two distributions. As hypothesized, the average normalized view time for relevant documents is higher than that for nonrelevant. Also, since the distribution of relevant times has a higher variance, it is more spread out, or flatter. The question still remains as to whether the nonrelevant distribution is bimodal or not.

Figure 1 shows two histograms of the nonrelevant normalized view times, which are further normalized by the total number of data points (115) to facilitate comparison to the histograms for relevant document view times. Figure 2 mirrors this for relevant view times. Contrary to our hypothesis, the nonrelevant distribution only looks vaguely bimodal, but in keeping with our hypothesis, the relevant distribution shows signs of normality. Currently, the bins in these histograms have very few (on the order of 10) data points. It is possible that a larger number of data points would better show the true nature of these distributions.

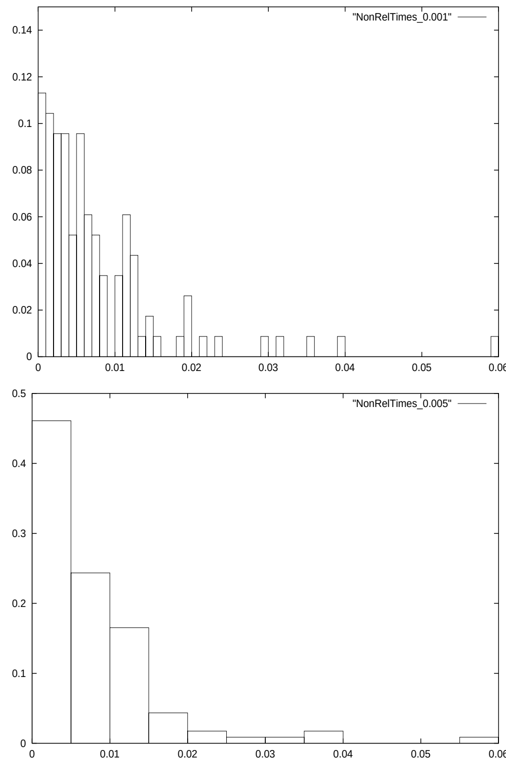

Figure 1: Distribution of Nonrelevant Normalized View Times (bin sizes 0.001 and 0.005)

In the end, what matters is whether or not one can distinguish relevant documents from nonrelevant using normalized view time. Figure 3 shows the receiver operator characteristic curve for using normalized view time as a predictor of relevance. Although the plot shows that it does have some predictive value, the fact that it is not that far from the plot of a random predictor ($y = x$) shows that it is not a great predictor. Clearly, there are other factors that need to be taken into consideration.

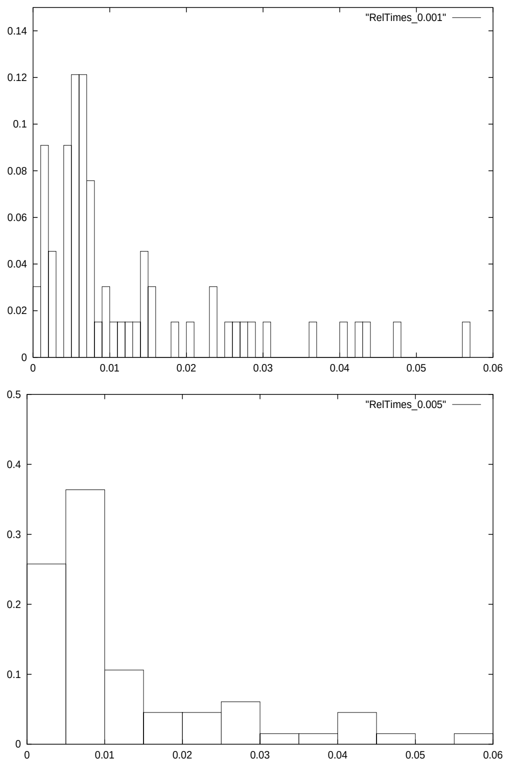

Figure 2: Distribution of Relevant Normalized View Times (bin sizes 0.001 and 0.005)

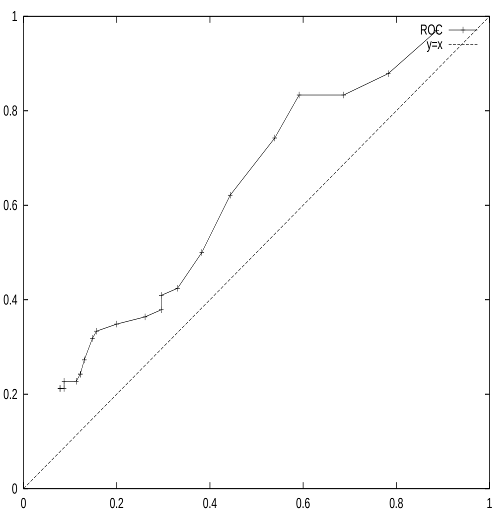

Figure 3: ROC Curve for Using Normalized View Time as Relevance Predictor (False Positive Rate vs. True Positive Rate for Various Theshhold Cutoffs)

4 CONCLUSIONS and FUTURE WORK

If we assume that the steps we took to correct for the errors in the design of our user action logging system were indeed effective, then we have shown that clickthroughs are not necessarily indicative of relevance. In fact, for this experiment, where the preview of a document was quite short (about 2/3rds of a headline), only about one third of the clickthroughs were actually to relevant documents. Clearly, this result depends heavily on our system, the corpus, and the task assigned to the users.

We have introduced the idea of length-normalized viewing time as a predictor of relevance. We have shown that the distribution of this measure differs between relevant documents and nonrelevant documents, although not to such a great extent that it could be used as the single predictor of relevance with real accuracy. Nevertheless, it is an important factor which can more accurately predict relevance than clickthroughs alone.

In the future, we hope to correct the deficiencies of this experiment by creating a better testbed which accurately records event timing, and expands on the types of events logged (e.g., scrolling, mouse-overs, highlighting, etc.), possibly using the Java programming language. It is our belief that a detailed log of user actions could very accurately predict relevance, thus alleviating the need for explicit feedback.

5 Acknowledgments

This research was supported by a Chapman University faculty research and development grant.

References

[Dreilinger and Howe, 1997] Dreilinger, D. and Howe, A. E. (1997). Experiences with selecting search engines using metasearch. *ACM Transactions on Information Systems*, 15(3):195–222.

[Morita and Shinoda, 1994] Morita, M. and Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In Croft, W. B. and van Rijsbergen, C., editors, *SIGIR 94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pages 272–281, Dublin. Springer-Verlag.

[Oard and Kim, 1998] Oard, D. W. and Kim, J. (1998). Implicit feedback for recommender systems. In *Proceedings of the AAAI Workshop on Recommender Systems*, Madison, WI.

[Salton, 1971] Salton, G., editor (1971). *The SMART Retrieval System – Experiments in Automatic Document Retrieval.* Prentice-Hall Inc., Englewood Cliffs, N.J.

6 Appendix

On the next few pages are the instructions given to subjects, including an ASCII rendition of the Web interface.

Instructions

Please read these instructions carefully. Please reread them multiple times until you feel comfortable with what they say. You will not see these instructions again during the course of the experiment. If you want, you can print this page out and/or copy-and-paste it to another window for reference when you're searching.

Overview of the Experiment

The goal of this experiment is to determine how well an information retrieval system can help you to answer questions you might ask when searching newswire or newspaper data. The questions are of two types:

- Find a given number of different answers. For example: Name 3 hydroelectric projects proposed or under construction in the People's Republic of China.
- Choose between two given answers. For example: Which institution granted more MBAs in 1989 - the Harvard Business School or MIT-Sloan?

You will be asked to search on four questions with one system and four questions with another. These two sessions might be separated by a period of up to a week. You will have **five minutes** to search on each question, so *plan your search wisely* (you will be shown the question before you have to start searching). You will be asked to answer the question and provide a measure of your certainty of your answer both before and after searching, and to indicate which documents were helpful for determining the answer.

You will also be asked to complete several additional questionnaires:

- Before the experiment - computer/searching experience and attitudes
- After each question
- After each four questions with the same system
- After the experiment - system comparison and experiment feedback

System Details

The searching system you will be using is similar to the typical Web search engine. You will type keywords into a text box, press a "Search" button, and then view a list of the titles of documents ranked according to how well the document matches your search query.

The screen will be divided into two portions. In the top portion, you will see the following elements: the question you are looking to answer, a search textbox and button, an area for typing in your answer to the question, and a drop-down list used to indicate how certain you are of your answer. There is also a countdown timer that indicates how long you have to complete your search (this is also displayed in the status bar at the bottom of your browser). The bottom half of the screen serves a dual purpose:

- It displays the list of results of a search, which consists of a sorted list of document IDs and titles, along with how many times each word in your query appears in that document.
- It also displays the full content of individual documents whenever you click on their title in the results list.

An example of what the screen would look like at some point in a search is shown on the last page of these instructions.

Every document in the database has a unique identifier, typically 2-4 letters which indicate the source (e.g., WSJ for Wall Street Journal), followed by a possibly hyphenated series of numbers. These will be displayed in a red font for easy identification. For example, WSJ911231-0122. It is **very important** that you include the document identifiers for those documents in which you found supporting evidence as part of your answer. Otherwise, we won't be able to verify your answer.

Verification of your answer will be done by human reviewers. Unfortunately, you will receive no immediate feedback as to whether your answer was correct, but we will let you know how you did after the manual scoring is done (some time this fall).

Hints and Special Requests

- **Please do not use the navigation buttons (Back, Forward, Reload, Refresh, etc.) on your browser.** Click only on buttons and links that are in the main portion of your browser screen.
- **Please observe the 5 minute time limit** - when the system tells you to submit your answer, do so without delay. This is a scientific experiment, so certain guidelines need to be followed by all subjects to ensure the results are accurate. The time you spend searching is logged, and if you spend significantly more than 5 minutes, we can't use your data.
- Some users of this system have noted the following ideas for fast and efficient use of the system:
  - Maximize the size of your browser window (you can do that right now if you want).
  - When you get through the initial questionnaire and are on the search screen, scroll the top frame of the browser window so you can see both the search box and the answer box.

- After issuing a query, the number of candidate documents might be quite large - you may have to do some scrolling to find what you're looking for. To help you, a portion of the document's headline will be displayed, along with how many times each word in your query appears in that document. This should help you quickly eliminate irrelevant documents.
- You don't have to wait for the whole results list to be displayed before choosing a document to examine - you can click on the document headline at any time.
- Use the browser's "Find" feature to quickly locate relevant parts of a document that is currently being displayed in the lower half of the screen (the keyboard shortcut for Find is usually Ctrl-F or Command-F)
- If one word is more important than others in your query, repeat it multiple times (e.g. "dog dog dog bones").
- It is often useful to modify your query by removing bad keywords (i.e., ones that cause the system to retrieve irrelevant documents) and adding new keywords. When you're reading a document, be on the lookout for specific terms that you may want to add to your query. If you're getting bad results, don't be afraid to modify and re-issue a query.
- This search engine does NOT support any of the fancy searching features you may be used to (e.g., AND, OR, +, enclosing phrases in quotes, etc.). It only allows you to type in keywords
- Use your browser's copy-and-paste facilities to copy document identifiers to the answer box (usually, Ctrl-C and Ctrl-V)

You should have been mailed a four-character username via email. Make sure you have about an hour's worth of time available and a consistent and good internet connection, and then enter your username to begin:

If you are having problems, please contact Chris Vogt

Example screen (rendered in ASCII) begins on next page:

```
=====================================================================================
3:14 Minutes Remaining
Please answer the following question by searching for relevant documents:
Which institution granted more MBAs in 1989 - the Harvard Business School or MIT-Sloan?
+-------------------------+
| harvard sloan mba       |  {Search}
+-------------------------+
Enter your answer in the space below along with one or more documents
that supports the answer.  (You must list at least one document that
answers the question.)
+-----------------------------------------------------------+
| MIT-Sloan FT944-15805 WSJ881104-0152                      |
|                                                           |
|                                                           |
+-----------------------------------------------------------+
Please choose how certain you are about your answer:
o Extremely Uncertain
o Somewhat Uncertain
o Neutral
o Somewhat Certain
o Extremely Certain
{Submit Answer}
=====================================================================================
DocID            mba  harvard  sloan  Title
FT944-15805      124  3        .      -Survey of Business Schools - An A-Z Guide (
AP881017-0012    .    32       .      -Presidential Campaign Elevates Elite Rival
FT924-176        22   1        .      -Management: Successful to a degree - Critics
FT932-12256      18   1        .      -Survey of Business Schools (2): Some employ
LA093090-0082    .    .        16     -ARTS FESTIVALS; IT'S SAN FRANCISCO'S TURN
LA113089-0061    .    14       6      -HARVARD-WESTLAKE MERGER IS FOCUS OF LAWSU
FT941-2127       12   6        .      -Survey of Management Education &amp; Trainin
WSJ881104-0152   .    1        14     -MIT's Sloan School Tries To Drop a Name
WSJ911007-0114   .    22       .      -Harvard Marks Down Speculative Fund by $
LA111989-0153    .    21       .      -COLLEGE FOOTBALL; THE SCENE AT ANOTHER RI
LA110289-0081    .    21       .      -PARENTS' REVOLT DELAYS MERGER OF 2 SCHOOL
LA113089-0094    .    12       6      -BATTLE GROWS OVER SCHOOLS' MERGER; EDUCAT
FT941-2136       16   .        .      -Survey of Management Education &amp; Trainin
FT932-12258      13   2        .      -Survey of Business Schools (1): Horses for
WSJ900827-0138   10   3        .      -Manager's Journal: How Much Is an MBA Re
FT944-14774      13   1        .      -Management: Degree of reluctance - George B
WSJ870615-0133   10   2        .      -Manager's Journal: The Value of Today's
FT922-13885      12   1        .      -Survey of Management Education and Training
FT944-9861       .    .        10     -Management: Pioneers and prophets - Alfred P
=====================================================================================
```