How People Recall Search Result Lists

Jaime Teevan
MIT, CSAIL
32 Vassar Street, G472
Cambridge, MA 02139 USA
teevan@csail.mit.edu

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
People commonly interact with lists of information – incoming emails are listed in a person's Inbox, search engines return lists of results, news stories appear as lists on newspaper Web sites, and people navigate file systems by listing directory contents. While the changes that occur to a list can be interesting – a new search result is interesting to the person searching for new information – changes are often secondary to the primary goal of using information. New search results are inconsequential when a person wants to summarize a set of results or return to a previously viewed Web page. This paper presents a study of which aspects of a list are memorable and thus are likely to be noticed when they change, and which are not memorable and thus are unlikely to be missed if changed. The study shows that what people remember about a list item is a function of their interaction with the item and the item's location in the list.

Keywords
Re-finding, search, personalization, implicit feedback, dynamic information, user profiling, lists, memory.

ACM Classification Keywords
H5.4. Information interfaces and presentation (e.g., HCI): Hypertext/Hypermedia – User issues. H3.3. Information storage and retrieval: Information Search and Retrieval – Search process, Relevance feedback.

Introduction

Information in the electronic world often changes. Search results change as new information is indexed and email Inboxes change as new emails arrive. The growing ease of electronic communication and collaboration, the rising availability of time dependent information, and even the introduction of automated agents, suggest information is becoming ever more dynamic. Even traditionally static information like a directory listing on a personal computer is beginning to become dynamic; Apple, for example, has introduced *smart folders*, folders based on queries that change their content as new information becomes available.

Changes are often useful – people want to see new email, better search results and new news stories. But it is important that changes occur in a way that does not disrupt the user's interaction with existing information. Time and time again changes that should help the user appear instead to get in the way. For example, dynamic menus were developed to support faster menu item access by bubbling common accessed items to the top. However, rather than decreasing access time, research revealed dynamic menus slow users down because the items no longer appear where expected [7, 9]. As another example, White et al. [12] tried to help people search better by giving them lists of relevant sentences that were dynamically re-ranked based on implicit feedback gathered during the search process. Despite the improvement in list quality, people did not enjoy the search experience as much or perform as well as when the sentence list was static.

It is important when information changes to understand what aspects of the information a person has already interacted with are memorable before updating the information. This understanding can be used to highlight important changes (by having changes occur to memorable aspects of the information) or to hide unimportant changes (by only allowing changes to occur to unmemorable aspects of the information).

This paper present a study of what 119 people found memorable about search result lists. In the study, participants were asked to interact with a result list and later asked to recall what they remembered about it. The paper begins with a brief overview of relevant cognitive psychology literature relating to list recall. It then discusses the study methodology and presents the results of the study. Although only search result lists are studied, the lessons learned likely apply to other types of lists, including lists of directory files, news stories, message board threads and navigational links.

Studies of List Memory

How lists of information items, and in particular lists of words, are recalled has been well studied by cognitive psychologists [1, 3, 4, 8, 11]. Several main effects have been observed, including the *primacy effect*, the *recency effect*, and the *von Restorff effect*. Because these effects are pertinent to the study presented here, they are described in greater detail in the side column.

Studies of list recall tend to take place in highly controlled environments and often do not involve information items that are of interest to the participants or actively interacted with in the way one might interact with a search result. Studies of more complex items than words, such as television commercials [11], and of more complex uses, such as the forming of impressions of personality [1], have found the effects mentioned above to be true with some variation.

J B T 2 A M H

Primacy effect
Items in a list that are first tend to be memorable. From the list of characters above, the *J* is particularly likely to be remembered.

Recency effect
Items in a list that are last tend to be more memorable than items in the middle. The *H* in the above list is more likely to be remembered than the *M*.

von Restorff effect
Distinctive items are memorable. The *2* in the above list is memorable even though it appears in the middle of the list because it is a number in the midst of letters.

Study Methodology

To discover what people found memorable about search result lists, participants were asked to interact naturally with a list of 8 to 12 results for a self-generated query. While typical studies of list memory require all items to be attended to, participants were not required to view every result. By allowing natural interaction, the study revealed which aspects of the result lists were attended to and remembered. Queries were issued to a search engine via a Web form accessed from the participant's own computer, and clicked results were logged. Within an hour, participants were emailed a survey asking them to recall the result list without referring back to it. Participants were asked to remember their query, the number of results, and basic information about each result, including its title, snippet, URL, whether it was clicked, and if so, what the Web page was like.

Table 1: Study demographics. Note that numbers do not add to 100% because not all participants responded to demographic questions.

Gender	
Male	52%
Female	45%
Age	
18-24	15%
25-39	64%
40+	18%
Affiliation	
MIT	27%
Computer use	
Daily	97%

Approximately half of the 245 people who issued an initial query completed the follow-up survey. Removing survey responses that were clearly erroneous (*e.g.*, the remembered information did not at all match the initial search, or the results were remembered exactly, which suggests the information was copied) yielded 119 responses for analysis. The study demographics are shown in Table 1. The follow-up survey was typically completed within several hours of the initial search. Sixty-nine percent of all responses were received within three hours, and all but five were received within a day.

The observable behavior captured was similar to the behavior commonly observed for Web search. The average query length was 3.2 words, comparable to what has been found by others with query log analysis [10, 13]. When interacting with the results participants on average followed 1.9 results, and this number is comparable to the 1.5 clicks per result page observed by others on considerably larger data sets [13].

Participants did not accurately remember how many results were returned in the initial list. The correlation coefficient between the actual number returned and the number recalled was 0.09, which is not significantly different from 0. Not surprisingly, a large majority of the participants assumed ten results were returned regardless of how many actually were. Although participants received anywhere from 8 to 12 results, the figures below present data for exactly ten results – the first five results from the beginning of list, and the last five results from the end of the list.

Because a result's recalled rank may not correspond to its true rank, it was necessary to match the description of each recalled results with an actual result. Two independent coders performed this matching with an 84% inter-rater reliability. The 189 results with descriptions that were rich enough for both coders to make the same match were considered to have been "memorable". These memorable results are analyzed to provide insight into how to predict which results will be remembered, and how to understand the relative likelihood that various different types of changes that can occur in a result list will be noticed.

What Makes a Result Memorable

Participants recalled little about the original result list. Although only a few hours elapsed between the first search and the follow-up survey, only 15% of all results displayed were memorable. Two main factors emerged from the data as affecting how like a result was to be remembered: where in the result list it was ranked and whether or not the result was clicked.

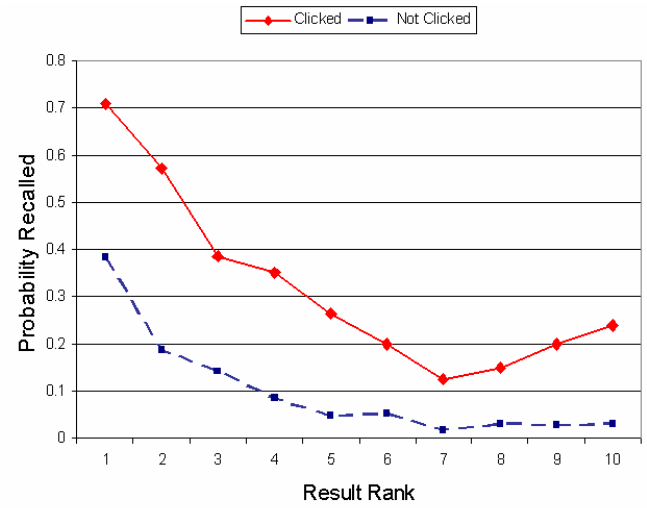

Figure 1: The probability of recalling a result of a given rank.

Rank Affects the Likelihood a Result is Remembered
As suggested by previous studies of list memory, the likelihood of a result being remembered was affected by where in the result list the result occurred. Figure 1 shows the probability a result was remembered given the result's rank for results that were clicked (solid line) and results that were not clicked (dashed line). The general shape of the curves is similar to what is seen in cognitive psychology literature. In particular, those results presented first appear memorable compared to later results (similar to the primacy effect) and those results presented last are somewhat more memorable than earlier results (similar to the recency effect).

Highly ranked results appear particularly memorable. This is probably because top results get more attention than lower ranked results that require scrolling to view. Results "below the fold" (typically result 7 and below) are often never seen at all [2]. Further, people tend to find highly ranked results more relevant, regardless of whether the results actually are more relevant [5]. Results thought to be highly relevant probably stand out because they are of direct interest and distinctive (similar to the von Restorff effect).

While the last results in the list are more memorable among clicked results, this is not the case for results that weren't clicked. This discrepancy may be because low ranked results that were not clicked were also often not read. Thus the last result seen for non-clicked results varied as a function of the individual (*e.g.*, the resolution of the participant's screens, how likely the participant was to review all of the results, etc.).

Clicked Results are More Memorable
Whether a result was clicked affected how likely it was to be remembered. The importance of click through data has been studied for its value as an implicit measure to determine result quality [5, 6]. In this analysis, click through is looked at as a way to determine how likely a result is to be remembered. On average only 8% percent of results that were not clicked were recalled, whereas significantly more results that were clicked were recalled (40%, $p<0.01$). The greater likelihood of a result that was clicked being recalled can be seen graphically in Figure 1.

The last result that was clicked was particularly likely to be remembered. A 12% increase in recall was observed if a result was the last clicked, compared to other clicked results. The last result clicked may be more memorable because it was also the last result seen (similar to the recency effect), and because it was

what the participant was looking for and thus distinctive (similar to the von Restorff effect).

Other Factors Affecting Recall

The number of times a result was visited also appeared to affect its likelihood of being recalled. A comment that often accompanied a result recalled with great detail was that it pointed to a Web page the participant visited often. This comment was common despite the fact that participants were not explicitly asked to share how often they visited a particular result.

In the above analysis, whether a result was memorable is considered independent of the other results in the list. However, there was some evidence that content of the result set affected the likelihood that an individual result would be remembered. In fact, some results were recalled that did not even occur in the original list. For example, a participant recalled a result with the URL "www.mlb.com" as being returned for the query "pittsburgh pirates", when in reality no such result was listed. Such phantom results probably represent pages that were found via other means but that conceptually belong in the search result set.

How Result Ordering Was Remembered

How result ordering was remembered was also analyzed to give insight into how changes to ordering might affect one's ability to interact with a search result list. Participants often were observed mistakenly recalled rank. The recalled rank differed from the actual rank 33% of the time. Mistakes tended to be less common for highly ranked results, and the first result's rank was remembered correctly 90% of the time. Accuracy dropped quickly as rank dropped, as can be seen graphically in Figure 2. This implies that

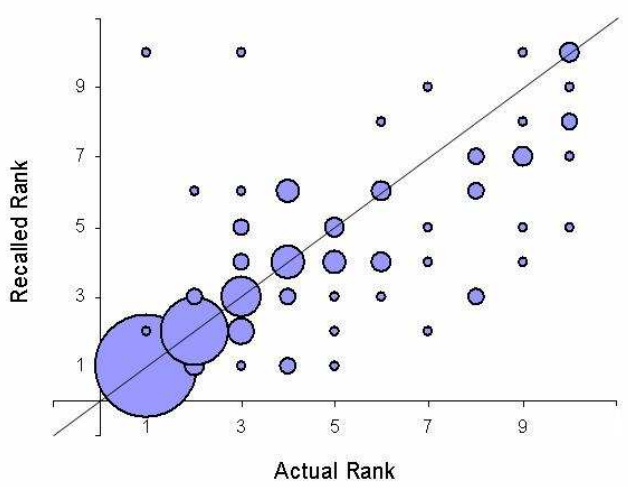

Figure 2: The result's location in the result list as the participant remembered it, compared with the result's actual location. The size of each point represents the number of people remembering that combination

moving a result from the number one position in a result list is more likely to be noticed than moving a lower ranked result.

Figure 2 also illustrates another interesting trend in the data. The greater weight of the data occurs to the right of the identity line. This means that remembered results were much more likely to be recalled as having been ranked higher than they actually were. Recalled results moved up in the result list 24% of the time, significantly more often than they moved down (10% of the time, $p<0.01$). The trend to remember results as highly ranked probably reflects the fact that remembered results were more likely to be relevant to the participant's information need and thus in the

participant's mind "should have been" ranked more highly than they actually were.

Although psychology literature suggests that relative ordering is important for recall [3], swaps in result ordering occurred in 10% of the responses where a swap was possible (*i.e.*, a participant's remembered results mapped to at least two real responses).

Design Implications

This paper has shown that people remember little about the search results they interact with, and presented several ways to identify particularly memorable results based on user interaction and result positioning. The findings of this study can be used to help people cope with changes to result lists. When it is important for a new result to be noticed quickly, it can be added in a way that disrupts the memorable aspects of the list. On the other hand, when valuable new information becomes available that is not of immediate interest, a consistent interaction can be provided by swapping in new results for unmemorable ones. I am currently in the process of developing and testing a system that invisibly incorporates new results into previously viewed result lists, presenting the best new information while maintaining the memorable aspects of the old.

Acknowledgements

This research has benefited from the involvement of David Karger, Susan Dumais, Rob Miller, Mark Ackerman, Molly Potter and the study participants.

References

[1] Asch, S.E. Forming Impressions of Personality. *Journal of Abnormal and Social Psychology*, 41 (1946), 1230–1240.

[2] Granka, L. A., Joachims, T. and Gay, G. Eye-Tracking Analysis of User Behavior in WWW Search. In *Proceedings of SIGIR '04*, 2004, 478–479.

[3] Henson, R. Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, 36 (1998), 73–137.

[4] Hunt, R.R. The Subtlety of Distinctiveness: What von Restorff Really Did. *Psychonomic Bulletin & Review*, 2 (1995), 105–112.

[5] Joachims, T., Granka, L, Pan, B. Hembrooke, H. and Gay, G. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *Proceedings of SIGIR '05*, 2005, 154–161.

[6] Kelly, D. and Teevan, J. Implicit Feedback for Inferring User Preference: A Bibliography. *SIGIR Forum*, 37, 2 (2003), 18–28.

[7] Mitchell, J. and Shneiderman, B. Dynamic Versus Static Menus: An Exploratory Comparison. *ACM SIGCHI Bulletin*, 20, 4 (1989): 33–37.

[8] Murdock, B.B. The Serial Position Effect of Free Recall. *Journal of Experimental Psychology*, 64 (1962), 482–488.

[9] Somberg, B.L. A Comparison of Rule-Based and Positionally Constant Arrangements of Computer Menu Items. In *Proceedings of CHI/GI '86*, 1986, 255–260.

[10] Spink, A., Wolfram, D., Jansen, B.J. and Saracevic, T. Searching the Web: The Public and Their Queries. *Journal of the American Society for Information Science*, 52, 3 (2001), 226–234.

[11] Terry, W.S. Serial Position Effects in Recall of Television Commercials. *The Journal of General Psychology*, 132, 2 (2005), 151–163.

[12] White, R., Ruthven, I. and Jose, J.M. Finding Relevant Documents Using Top Ranking Sentences: An Evaluation of Two Alternative Schemes. In *Proceedings of SIGIR '02*, 2002, 57–64.

[13] Xue, G.-R., Zeng, H.-J., Chen, Z., Yu, Y., Ma, W.-Y, Zi, W. and Fan, W. Optimizing Web Search Using Web Click-Through Data. In *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining*, 2004, 118–126.