Remembrance of Things Tagged: How Tagging Affects Human Information Processing

Raluca Budiu (raluca@alumni.cmu.edu)
Peter Pirolli (pirolli@parc.com)
Lichan Hong (hong@parc.com)
Palo Alto Research Center
3333 Coyote Hill Rd
Palo Alto, CA 94304, USA

Abstract

We describe an experiment investigating how tagging affects human memory. We compare a low-cost tagging by mouse-clicking interface (Click2Tag) to traditional tagging by typing and to a baseline, no tagging condition. Our results suggest that Click2Tag is comparable to the no-tagging condition in terms of memory performance, and it leads to better recognition of facts from read text than tagging by typing. These results support the hypothesis that tagging by clicking strengthens the memory traces by repeated readings of relevant words in the text and, thus, improves recognition.

Keywords: Tagging, memory, interaction cost, social bookmarking.

Introduction

In recent years, there has been an explosion of social bookmarking systems (e.g., del.icio.us, diigo, ma.gnolia, CiteUlike). These systems allow users to generate labels or keywords (*tags*) to content encountered on the Web (for example, a user may tag the URL *http://www.nytimes.com* with the tags *news* and *nytimes*). The tags and the associated URLs are then stored by the bookmarking system; they can be accessed later by their owner, but also by other uses. Thus, the social aspects of these systems emerge from implicit or explicit sharing of tags among users. As evidence of the growth of social bookmarking system, the Pew Internet Survey (Rainie, 2007) reports that 28% of U.S. Internet users have tagged or categorized content such as videos, photos, news stories, or blogs. Theoretical explanations (Benkler, 2002; Pirolli, 2007), discussed below, attribute the growth of participation in socially mediated systems partly to the reduction in the costs of producing sharable content (e.g., bookmarks and tags): tagging tools that have become part of the web browser allow people to easily tag content as they read. Consequently, in the development of our own social tagging system SparTag.us (Hong, Chi, Budiu, Pirolli & Nelson, 2007), we have been motivated to develop techniques that lower the costs of producing tags and other annotations. However, theories of memory (e.g., Anderson, 2000) also suggest that changes in the kind and amount of tagging effort involved may affect how well people remember the original content or tags. We want to avoid tagging techniques that provide low interaction costs, but make it harder for people to remember the content that they have tagged.

In this paper, we unravel how different techniques for producing tags to Web content affect individual memory. We are also generally interested in whether the act of tagging increases memory for content. In particular, we performed an experimental contrast of a lower interaction cost technique (Click2Tag), developed for our system SparTag.us, against a standard (higher interaction cost) type-to-tag technique, similar to ones used in popular tagging systems such as del.icio.us. Both of these were contrasted with a baseline condition of no tagging. Click2Tag allows users to simply mouse-click words in a text to have the words become tags for the content. Type-to-tag allows users to type their own tags for the content. We examined the effect of these techniques on both recognition and recall tests for the original material. As we discuss below, previous memory research (Anderson, 2000; Bradshaw & Anderson, 1982; Pirolli & Anderson, 1985) leads us to hypothesize that Click2Tag would produce better recognition of the original material, but type-to-tag would produce superior recall. These results are predicted by differences in the way the techniques either strengthen or elaborate the memory traces for the original content. We also discuss the implications of this trade-off for the functionality and design of social tagging systems.

Tagging Costs

Tagging Cost and Participation

Social information foraging theories (Pirolli, 2007), as well as microeconomics theories of networked information economies (Benkler, 2002) predict that as the costs of production of shareable knowledge (e.g., tags) are driven down, more individuals will participate and reap greater net benefits. Thus, reductions in the cost of tagging will improve the value of the system to the individual user. More tags, and presumably more useful tags for the individual, will be generated as more people join a social tagging site.

Click2Tag: Lowering the Cost of Tagging

To lower the cost of tag production, SparTag.us makes each content word on a Web page clickable (see Figure 1). As users click words, the words are added to a tag list. The assumption is that this technique (hereafter called Click2Tag) provides a low-cost method for tagging in comparison to the more standard type-to-tag technique.

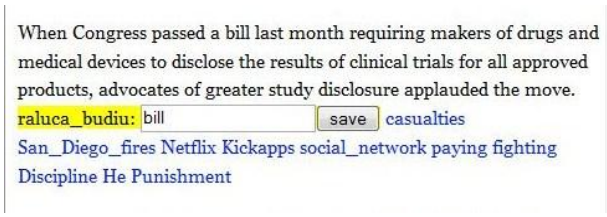

Figure 1. Using Click2Tag to tag a paragraph in SparTag.us.

In type-to-tag, users can generate any string as a tag and enter it into a tag list. The difference is that this involves the additional costs of mentally generating the tag plus the cost of typing.

One worry about Click2Tag might be that it trades off costs in tagging time for cognitive costs to subsequent memory. Human memory research suggests that Click2Tag and type-to-tag may have different effects on subsequent memory.

The *strength* of a memory trace captures the relationship of practice to memory performance. It has been shown that repeated practice increases strength, and strength decays as a function of time since last practice (see Anderson, 2000 for a review). Reaction times and accuracy on recognition tests both improve with strength (e.g., Pirolli & Anderson, 1985). Click2Tag appears to encourage users to re-study the original content and, thus, increase the strength of memory traces for that content. Indeed, as shown in Figure 2a, eye tracking data from pilot studies in our lab suggest that, in the Click2Tag condition, participants often read a passage and then re-scan the passage to seek out words to click. In the no-tags condition (see Figure 2c), we see participants fixating more on relevant words when they read text, although perhaps less so than in the Click2Tag condition. In comparison, when participants have to type in tags (see Figure 2b), they tend to fixate less on any specific words. Thus, we expect Click2Tag to produce better recognition performance than type-to-tag.

Elaborateness of encoding of material also generally improves aspects of memory. Memory research (e.g., Bobrow & Bower, 1969; Bradshaw & Anderson, 1982) has shown that, when people are asked (or provided) with additional information that is highly semantically related to the content they are studying, they typically show superior recall to content that has been processed less elaborately. One explanation (Anderson & Reder, 1979) suggests that the memory traces that elaborate the original content provide additional retrieval routes to recall the content. This is because self-generated elaborations have some high likelihood of being re-generated at recall time as a retrieval route to the content. Since type-to-tag requires users to generate tags to associate with the original content, we expect it to produce more elaborative encodings and to improve recall performance.

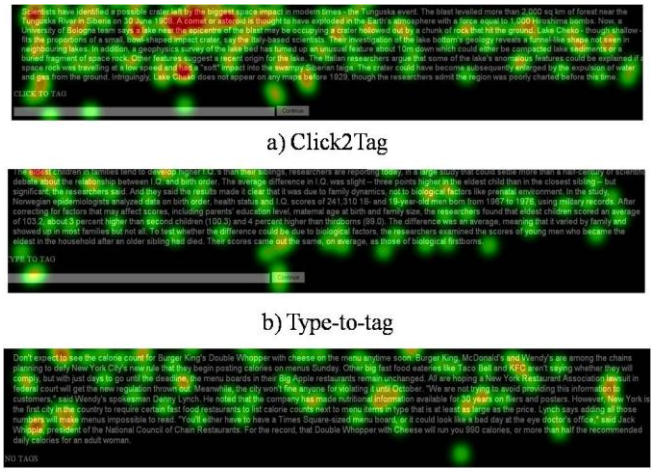

Figure 2. Eye tracking of reading when using two different tagging techniques versus no tagging. Red spots signify more time spent at those locations.

Experiment

Method

Participants. We recruited 27 participants for this study; they were each compensated with $20 in cash or Amazon gift certificates.

Materials. We selected 18 passages from news articles as well as from various web pages on the Internet. The passages reflected a variety of topics (medicine, education, general science, aviation, history, etc). On average, the passages were 267 words long (ranging from 253 to 279).

Procedure. Participants went through three study-recall blocks. A study-recall block had two parts: in the first part, participants performed 6 study trials, and in the second part, participants performed 6 memory trials. Participants were instructed to perform these trials as fast and accurate as possible.

Study trials. In each study trial, participants read a passage, selected randomly from the list of 18 passages. Participants were instructed to read at their own paces, but if they spent more than 2 minutes on a trial, they were automatically moved to the next trial. The trial could belong to one of three conditions as follows:

No-tags: In this condition, no tagging was performed.

Click2Tag: Participants had to tag the passage with relevant words by clicking on words from the passage. The tags were displayed in a box under the passage and could not be modified by the participants.

Type-to-tag: Participants had to tag the passage with any relevant tags that they could generate, and type those tags in a box under the passage. Subjects were not allowed to cut and paste in this condition.

Memory trials. After 6 study trials, the participants completed 6 memory trials. A memory trial had two components, presented in the following order:

Recall: Participants were given two cues about one of the 6 passages they had previously studied (e.g., "One of the passages you read was about Christmas and Santa Claus"). Then they had to remember and type as many facts from that passage as possible. There was a time limit of 1.5 minutes per passage for this phase.

Recognition: Participants had to answer 6 true/false sentences (3 true and 3 false) about the passage they had just recalled. They had 1 minute to answer all questions.

The order of presentation of questions was randomized, and the order of passages within a block was randomly generated for each participant, as was the assignment of a passage to a particular block or condition. The passages in the memory trials were presented in a random order, unrelated to the order in which the passages were studied. The only constraint was that there be 6 passages per block, and within each block, there be 2 passages per condition. To keep them engaged in the experiment, at the end of each block, subjects were given feedback about their overall performance so far in the recognition tests. We did not give feedback after individual recognition trials.

For each trial, we measured the study time, the number of tags, the number of facts recalled and the recall time, the recognition time per question and the recognition accuracy.

Results

We performed ANOVAs with subjects as the random factor using Block (0,1,2) and Condition (no-tags, Click2Tag, type-to-tag). When there was a correlation with study time, we performed ANCOVAs with study time as a covariate.

Encoding

Table 1 shows the average study time per condition and per block. There was a significant effect of block on the study time ($F(2,52)=9.94$, $p<0.001$), with participants taking longer to study the passages as they went through the experiment. There was also a significant effect of condition ($F(2,52)=52.72$, $p<0.001$) and a significant block-by-condition interaction ($F(4,104)=3.97$, $p<0.005$). Contrasts showed that participants spent less time in the no-tags condition than in the Click2Tag condition ($p<0.001$) and in the type-to-tag condition ($p<0.001$), and also participants in the Click2Tag condition were faster than in the type-to-tag condition ($p=0.05$). These results pointed to a time cost associated with the tagging conditions**,** and also confirmed that Click2Tag was a lower-cost interaction technique than type-to-tag.

	Block 0	Block 1	Block 2	Avg
No-tags	71.95	83.35	88.82	81.38
Click2Tag	93.78	98.10	96.13	96.01
Type-to-tag	96.06	103.45	102.92	102.92

Table 1. Study times (seconds).

People tended to attach more tags ($p<0.001$) in the Click2Tag condition (6.95 on average) than in the type-to-tag condition (3.98 on average), suggesting that they made use of the ease to tag in the Click2Tag condition to attach more tags faster than in the type-to-tag condition.

Recall

Time to recall. There was no significant correlation ($r=0.03$, $p=0.6$) between the time to recall and the study time. Our ANOVAs resulted in no significant effects or interactions – people took the same amount of time to recall the content of the passages in all conditions (61.60 seconds on average).

Number of recalled facts. Table 2 shows the average number of facts recalled per condition, and Figure 3 splits these data per condition and per block[1]. The number of recalled facts was marginally correlated with the study time ($r=-0.10$, $p<0.1$), so we performed an ANCOVA with study time as a covariate. We obtained a significant effect of block ($F(2,29)=4.91$, $p=0.01$), but no other effects or interactions were significant.

	Number of facts recalled	Recall efficiency (facts/sec)	Recognition accuracy
No-tags	4.03	0.061	0.84
Click2Tag	4.20	0.052	0.82
Type-to-tag	3.84	0.043	0.79

Table 2. Average recall and recognition measures for the three conditions.

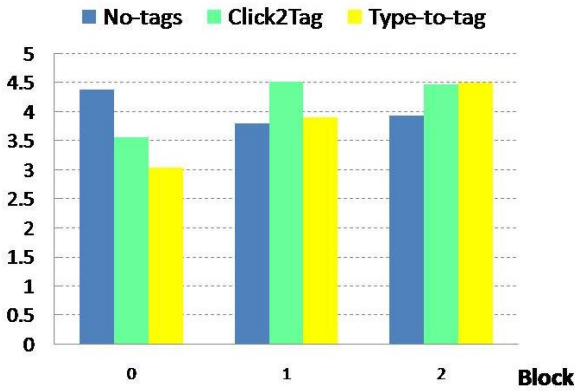

Figure 3. Number of recalled facts per block.

These results show that there was an overall learning effect through the experiment, with people getting better at recalling as the experiment went on. Contrary to our initial

[1] Due to an error in the experiment software, we did not have access to the recalled facts for 198 trials (out of 486 trials). However, the recorded trials were evenly distributed among conditions: 96, 97, and 95 trials in the three conditions respectively.

hypothesis, we did not find any advantage of the tagging conditions (and specifically of type-to-tag) for recall, although there is a non-significant trend in the data for the tagging conditions to surpass the no-tags condition over time.

Recall efficiency. Another measure that we looked at was the recall efficiency: How many facts did subjects recall per unit of study time (i.e., the ratio between number of facts recalled for a passage and study time for the same passage) – see Table 2. ANOVA showed a significant effect of condition ($F(2,31)=6.73$, $p<0.005$) and a marginal condition-by-block interaction ($F(4,60)=1.99$, $p=0.1$). Pairwise comparisons indicated higher efficiency for no-tags than for type-to-tag (0.06 vs. 0.04 facts/second). Contrasts also showed that the no-tags condition was better than the two tagging conditions combined ($t(285)=2.64$, $p<0.004$), although there was no significant difference between Click2Tag and no-tags.

These results are encouraging for Click2Tag. They show that, although for type-to-tag, the supplemental cost of tagging affects the recall, by using a low-cost technique the effects on recall become comparable to the no-tags condition.

Recognition

We ran ANCOVAs with the truth of the sentence included as an additional factor in the analyses.

Accuracy. The last column in Table 2 shows the average recognition accuracies in the three conditions, and Figure 4 splits these data by block and condition. The correlation between the study time and the accuracy at recognition is low ($r=-0.03$) and not significant, and therefore we only report the ANOVA. We found a significant effect of condition ($F(2,47)=6.38$, $p<0.005$), a significant effect of truth ($F(1,20)=4.86$, $p<0.05$), and a significant interaction between condition and block ($F(4,102) = 3.81$, $p<0.01$), as well as a significant interaction between truth and condition ($F(2,46)=3.27$, $p<0.05$). People responded correctly on 79% of the trials in the type-to-tag condition; this was significantly lower than the 84% accuracy in the no-tags condition ($p<0.005$) and the 82% accuracy in the Click2Tag condition ($p =0.05$). The difference between Click2Tag and no-tags was not significant.

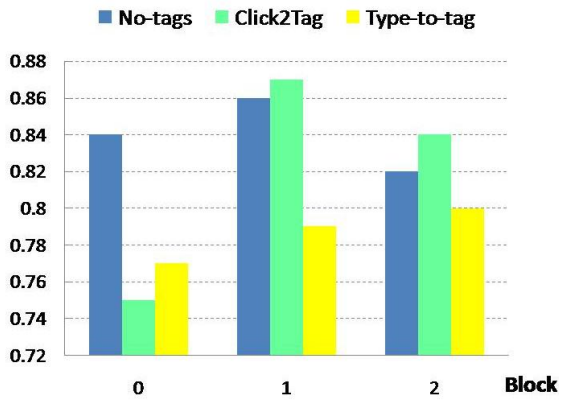

Figure 4. Recognition accuracy per block.

These results show that Click2Tag did not impair memory, in spite of reducing the tagging costs. People recognized facts equally well in the no-tags and Click2Tag condition. However, they were slightly less accurate when they typed. This supports our hypothesis that Click2Tag strengthens facts in memory, and thus has better effects on recognition than type-to-tag. As in the case of recall, the no-tags condition was better in the beginning than the tagging conditions, suggesting some learning of the experimental setup.

Recognition times. Table 3 shows the recognition times. There was a significant correlation ($r=0.28$, $p<0.001$) between study time and recognition time, so we report results from an ANCOVA with study time as a covariate. Condition was significant ($F(2,46)=3.24$, $p<0.05$), as were truth ($F(1,20) =11.60$, $p<0.005$) and condition by block interaction ($F(4,101)=5.89$, $p<0.001$). Contrasts showed that Click2Tag led to faster recognition times than both no-tags (marginally significant: $p=0.07$) and type-to-tag ($p <0.01$).

	Block 0	Block 1	Block 2	Avg
No-tags	4.83	5.02	4.83	4.90
Click2Tag	5.08	4.43	4.64	4.70
Type-to-tag	5.01	5.13	4.86	5.00

Table 3. Recognition times (seconds).

Since Click2Tag was faster than the other conditions and also more accurate than type-to-tag and as accurate as no-tags, there was no time-accuracy trade-off involved. The lower time of Click2Tag validates our assumption that Click2Tag is a low cost interaction that also has beneficial effects on recognition memory, compared with the more traditional type-to-tag technique.

Discussion

Our results indicate that participants took longer to read the passages in our experiment when they also had to tag them (Click2Tag or type-to-tag) compared to the no-tags condition. Moreover, people were faster to encode text in the Click2Tag condition than in the type-to-tag condition, and they also attached more tags, suggesting that, as initially hypothesized, Click2Tag is indeed a more efficient tagging technique in terms of interaction cost.

With respect to our question of how tagging affects memory, we did not find any clear benefits of the tagging conditions with respect to the no-tags condition. For recall, there was no difference among conditions in the number of facts recalled; this was contrary to our initial hypothesis that type-to-tag encourages elaboration and therefore may promote superior recall. For recognition, although Click2Tag and no-tags were comparable, people tended to recognize facts faster in the Click2Tag condition than in the no-tags condition. Moreover, the type-to-tag condition was worse in both accuracy and recognition time than the Click2Tag condition. This result is consistent with our

initial hypothesis (and with the observation from our pilot eye-tracking study – see Figure 2) that Click2Tag encourages people to revisit material more and thus strengthens the memory traces.

Our results confirm the intuition that Click2Tag decreases the cost in tagging time, and show that it succeeds to avoid cognitive costs to subsequent memory. Indeed, the study time in the Click2Tag condition is lower than in the type-to-tag condition, and the memory measures show comparable or better performance than in the other conditions.

Two questions still remain: (1) why we do not see any memory advantage for the tagging conditions in general, and (2) why the type-to-tag condition in particular does not perform better at recall. A possible explanation is that of a trade-off between interaction and elaboration time (spent in actually performing the tagging) on the one hand, and reviewing time (spent in reviewing and encoding the text), on the other hand. To explore in more detail this hypothesis, we built a simple model.

Let us assume that the encoding time is given by the equation:

$$T_{study} = T_{read} + T_{\text{interaction}} + T_{elaboration} \quad (1)$$

Thus, in the study phase, participants spent some time reading and encoding the material T_{read}, then they attach the tags (the interaction time $T_{\text{interaction}}$), and finally they may elaborate on the text they just read (the elaboration time $T_{elaboration}$). We assume that the elaboration time is zero for no-tags and Click2Tag, and that the interaction time is zero for no-tags. We also assume that people spend the same amount of time T_{read} encoding the material in the no-tags and Click2Tag conditions (and that is given by the overall study time for the no-tags condition – namely, 81s). Note that the interaction time contains not only the time to actually type or mouse click the tags (those would be very small compared to the overall reading time), but also the time needed to think about what would make a good tag.

We assume that the strength of the memory trace is proportional to the time spent reading (see Anderson, 2000):

$$S = A \times T_{read}^{B}, \text{ with A, B constants.} \quad (2)$$

Moreover, to recall a fact from memory, we assume that one could either directly retrieve it, or, if it could not be retrieved, one could attempt to infer it making a plausibility judgment based on an elaboration performed at encoding. Following Anderson and Lebiere (1998), we assume that the probability of retrieving a fact from memory is:

$$P = \frac{1}{1 + K \times e^{-S/\varsigma}}, \text{ with K,}\zeta\text{ constants.} \quad (3)$$

Thus, the probability of recalling a fact from memory is:

$$P + (1-P) \times P_{elaboration}$$

$$= \frac{1}{1 + K \times e^{-S/\varsigma}} + \frac{K \times e^{-S/\varsigma}}{1 + K \times e^{-S/\varsigma}} \times P_{elaboration} \quad (4)$$

The accuracy at recognition is simply *P*: the probability that a fact will be retrieved from memory.

Thus, we can assume that in the type-to-tag condition the time T_{read} is smaller than in the other two conditions, both because of the increased interaction time and because subjects also elaborate the content in order to find appropriate tags. If that is the case, we can expect that the strength of the memory traces for the facts in the passage be smaller than in the other conditions, and thus the probability to retrieve those facts may be smaller than in the no-tags or Click2Tag conditions. However, because of elaborations, in the type-to-tag condition there is actually another way to access those facts (via inferences) – and that is the second term in Equation 4. Thus, the overall probability of recall is the same for this condition as for the other two conditions.

Unlike for free recall, for recognition the elaboration does not play a big role (Anderson, 2000). Therefore, the probability to recognize a fact correctly is simply the probability to retrieve it from memory, as given by Equation 3. Consequently, we get that the recognition accuracy in the type-to-tag condition is lower, simply because the reading time T_{read} is lower than in the other conditions, and thus the strength and the probability of recognition are lower.

Table 4 shows the values with which we have instantiated these parameters. Note that A and ζ can be compressed into a single parameter corresponding to $\frac{A}{\varsigma}$, since if we replace *S* of Equation 3 with Equation 2, the two variables only occur in that ratio. To determine the components of the study time in Equation (1) corresponding to the three conditions, we used the study times that we observed in our experiment, assuming that the encoding time T_{read} was 81s (corresponding to the study time for the no-tags condition in Table 1) for both the no-tags condition and the Click2Tag condition. For the type-to-tag condition, T_{read} was a estimated as 53.77s.

A/ς	1.20
B	0.22
K	5.45
T_{read} for type-to-tag (s)	53.77
$P_{elaboration}$	0.1

Table 4. Model parameters.

With these values, the above model predicts a probability of recognition of 0.82 for no-tags or Click2Tag and 0.77 for type-to-tag. The predicted number of recalled facts is virtually the same for all the conditions (4.3 for the no-tags or Click2Tag and 4.1 for type-to-tag)[2].

[2] The number of recalled facts was obtained from the probability of recall assuming that the average number of facts per passage was 10.

Thus, using these equations, we can see that the time spent on elaborating is effective for recall, but on the other hand it impacts the overall time spent for revisiting the material. So, although elaboration per se helps recall, in the context of our experiment it interferes with the encoding, ultimately the elaboration benefits compensate the encoding deficits.

Our results are similar in spirit to note-taking research. One study (Bauer & Koedinger, 2007) compared note-taking using typing versus pasting and found that pasting led to better learning than typing. However, Bauer and Koedinger (2006) found that note-taking using pasting combined with typing made participants record more ideas, but also forget more. Other research (Bretzing & Kulhavy, 1979). has shown that verbatim notes (presumably similar to Click2Tag) led to poorer learning than paraphrased or summarized notes.

Conclusions

We started this paper with an argument over the importance of reducing the interaction cost for social tagging systems. However, in reducing the interaction costs, we must ensure that we do not increase cognitive costs. We showed that Click2Tag, a method that reduces the tagging cost by allowing users to click words in order to tag content, does not impair memory performance. In fact, compared to typing, it can be beneficial to recognition memory, due to strengthening of relevant words in text. In contrast, although typing tags may encourage text elaborations, the cost of typing is too high and impacts the recall efficiency, as well as the performance in recognizing seen content.

However, there may be other benefits of type-to-tag that remain to be explored. Type-to-tag is a top-down procedure, and it induces users to fit the content into their own "ontology" and retain only those facts matching their view of the world. On the other hand, Click2Tag is more bottom-up, content-driven. People tag with relevant words in the passage, paying less attention to their own ontology. One question that remains to be explored is whether the two techniques will lead to different tag "folksonomies". Another related question refers to the impact of the tagging technique on information retrieval. If Click2Tag leads to a proliferation of tags, would those tags contain less information and impact the ease of retrieval? How about the combination of Click2Tag and typing (which is indeed what has been implemented in SparTag.us) – how would it compare to typing or clicking alone?

Our research suggests that it is worth exploring lower cost techniques for tagging. Although the full implications of tagging by typing still remain to be explored, we have shown that a lower-cost interaction technique (Click2Tag) has beneficial effects on human information processing when compared to a higher cost technique (tagging by typing), possibly because it allows the users to tag without trading-off content reading time.

Acknowledgments

We thank Michelle Gumbrecht and Ed Chi for useful comments; Leila Takayama for revising the materials; and our participants.

References

Anderson, J. R. (2000). *Learning and Memory: An Integrated Approach.* John Wiley, New York, NY.

Anderson, J. R. & Lebiere, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum.

Anderson, J.R. & Reder, L. (1979). An elaborative processing explanation of depth of processing. In L.S. Cermak and F.I.M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Erlbaum.

Bauer, A. & Koedinger, K. (2006). Pasting and Encoding: Note-Taking in Online Courses. *Proc. ICALT'06,* 789-793.

Bauer, A. & Koedinger, K. (2007). Selection-Based Note-Taking Applications. *Proc. CHI'07,* 981-990.

Benkler, Y. (2002). Coase's Penguin, or Linux and the Nature of the Firm. *Yale Law Journal,* 112, 367-445.

Bobrow, D. & Bower, G. (1969). Comprehension and recall of sentences. *J. of Experimental Psychology, 80,* 455-461.

Bradshaw, G. & Anderson, J. (1982). Elaborative Encoding as an Explanation of Levels of Processing. *J. of Verbal Learning and Verbal Behavior,* 21, 165-174.

Bretzing, B. & Kulhavy, R. (1979). Note-Taking and Depth of Processing. *Contemporary Educational Psychology,* 4, 145-153.

Hong, L., Chi, E., Budiu, R., Pirolli, P., & Nelson, L (2007). SparTag.us: A Low Cost Tagging System for Foraging of Web Content. *Under review.*

Pirolli, P. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, New York, NY.

Pirolli, P. & Anderson, J. (1985). The Role of Practice in Fact Retrieval. *J. of Experimental Psychology: Learning, Memory, and Cognition,* 11, 136-153.

Rainie, L. (2007). 28% of Online Americans Have Used the Internet to Tag Content. *Pew Internet & American Life Project.* http://www.pewinternet.org/ pdfs/PIP_Tagging.pdf. Retrieved Jan, 2008.