

A comparison of LSA, WordNet and PMI-IR for predicting user click behavior

Ishwinder Kaur & Anthony J. Hornof
 Department of Computer & Information Science
 University of Oregon, Fall 2004
 {ishakaur, hornof}@cs.uoregon.edu

ABSTRACT

A predictive tool to simulate human visual search behavior would help interface designers to inform and validate their design. Such a tool would benefit from a semantic component that would help predict search behavior even in the absence of exact matches. This paper discusses a comparison of three semantic systems - LSA, WordNet and PMI-IR - to evaluate their performance in predicting the link that people would select given an information goal and a webpage. PMI-IR best predicts the human performance.

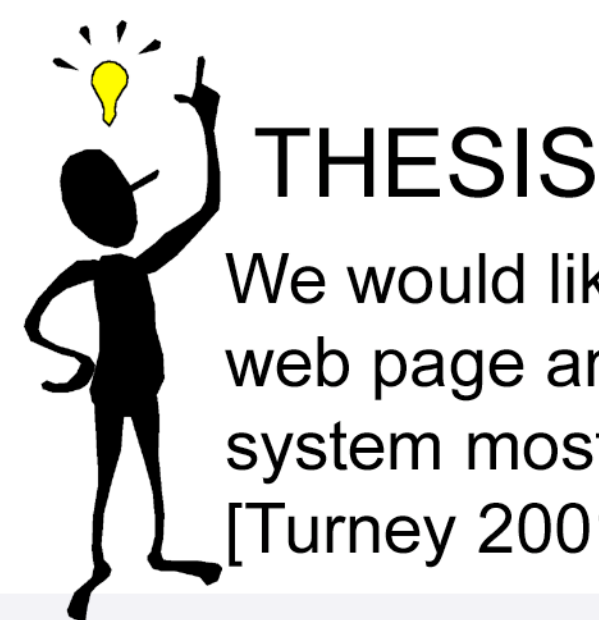
MOTIVATION

Visual interfaces are the predominant method of conveying information in human computer interaction. Predicting human behavior in tasks involving their use would be useful for the designers of the interfaces. A number of tools that evaluate some parameters of web usability, such as ease of navigation or lack of confusability, have tried to incorporate semantics in some form [Cognitive Walkthrough for the Web (CWW); Blackmon et al. 2002], [Bloodhound; Chi et al. 2003]. It would be useful in such cases to have empirical validation of the choice of a semantic system in the context of HCI tasks.

Semantics have a profound effect on the way visual interfaces are searched [Brumby et al. 2004, Pierce et al. 1993]. Exact matches between information goals and link labels are perhaps rare in real world tasks. The vocabulary problem, for example, points out that for a single concept, there can be no single word or description (however well chosen) that will be used by all people [Furnas et al. 1987]. This applies to the choice of a label for a concept by the designer of the interface, and to the goal in the mind of the users while searching that interface.

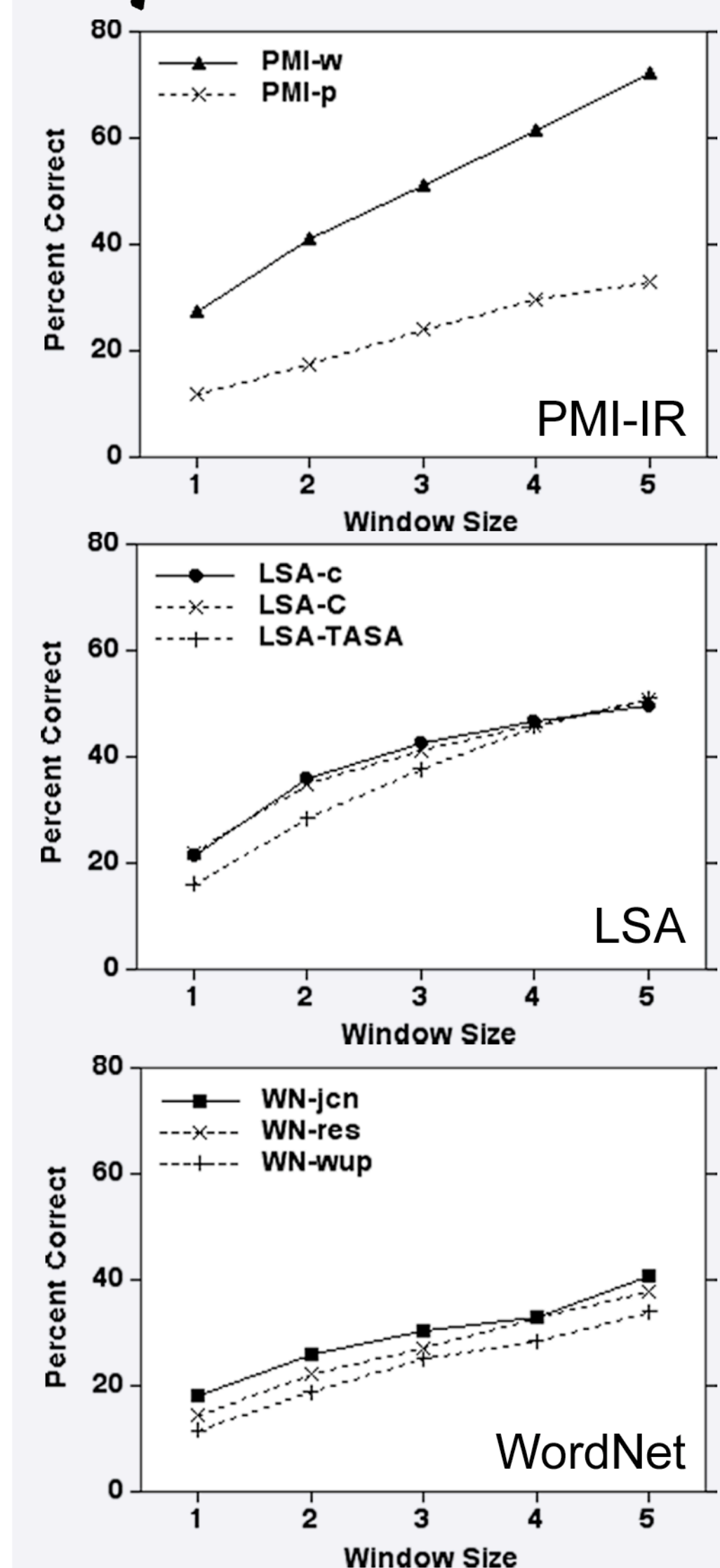


How do I speak the users' language??



THESIS

We would like to evaluate existing semantic systems that could be plugged in and be the semantic "brain" of the tools in question. Given a web page and an information goal, the ideal system would choose the same link label that a human would. We want to determine which system most accurately predicts the human link selection behavior. WordNet [Miller 1995], LSA [Landauer & Dumais 1997] and PMI [Turney 2001] were all applied to predict the behavior.



Percent correct as a function of window size for the three systems (a) WordNet, (b) LSA, & (c) PMI-IR.

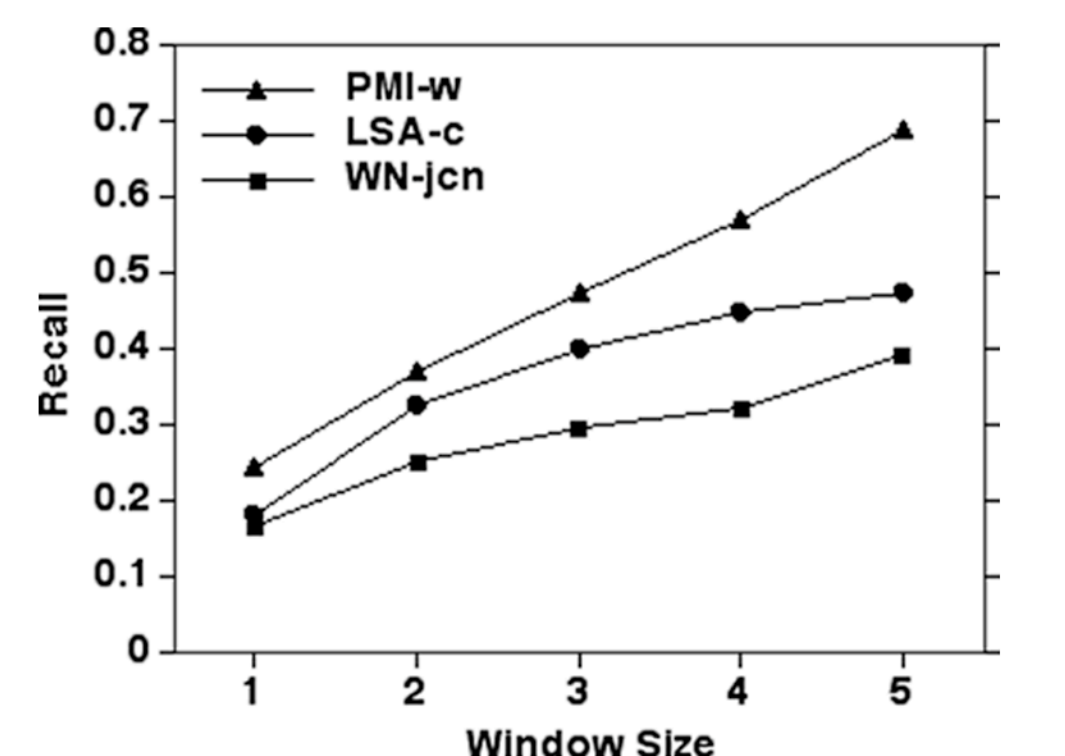
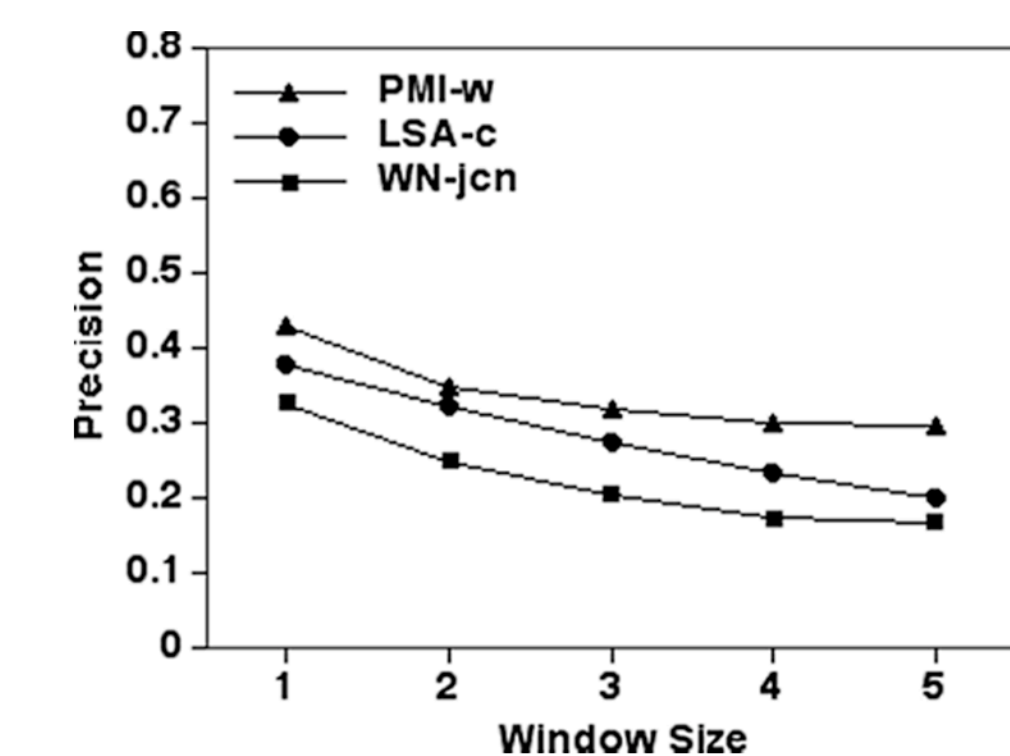
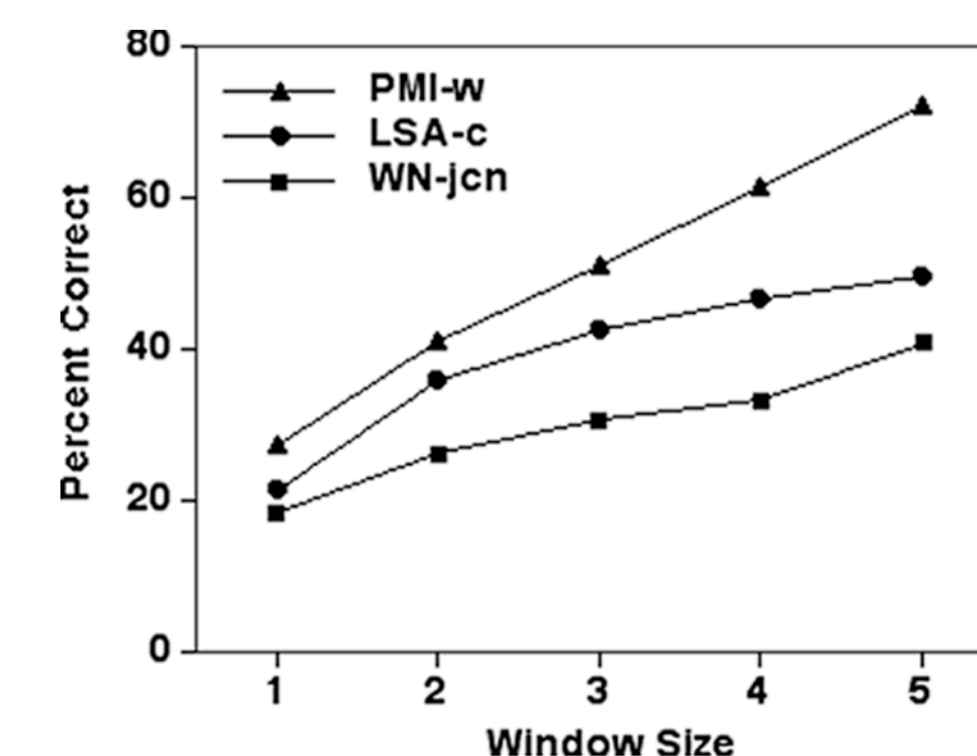
SYSTEMS' PREDICTIONS

This section compares the predictions of the various systems. For each webpage-goal pair (WP-G), participants in the survey responded with a set of links they would click on. These human preferences are taken to be **target responses**. The target responses are weighted by the number of votes received by each of them in the survey. For the same WP-G pair, each semantic system was used to rank the link labels from most similar to least similar. The top N labels in this ranking are taken as the **system responses**. N is the window size. The overlap between the target and the system responses are termed the **correct responses**. We used the three dependent variables, which were computed as a function of the semantic measure used and the **window size N**, to compare the performance of WordNet, LSA & PMI-IR:

Percent Correct: votes received by correct responses as a percentage of the total number of votes for the target responses (that is, achieved votes vs. maximum possible votes).

Precision: number of correct responses divided by the number of system responses (that is, accuracy of prediction).

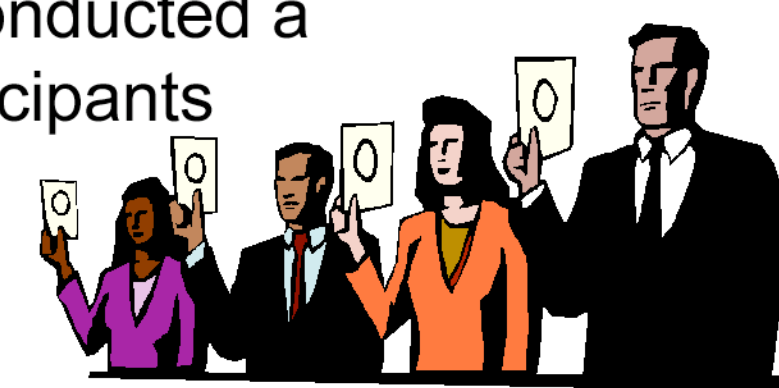
Recall: number of correct responses divided by the number of target responses (that is, potency of prediction).



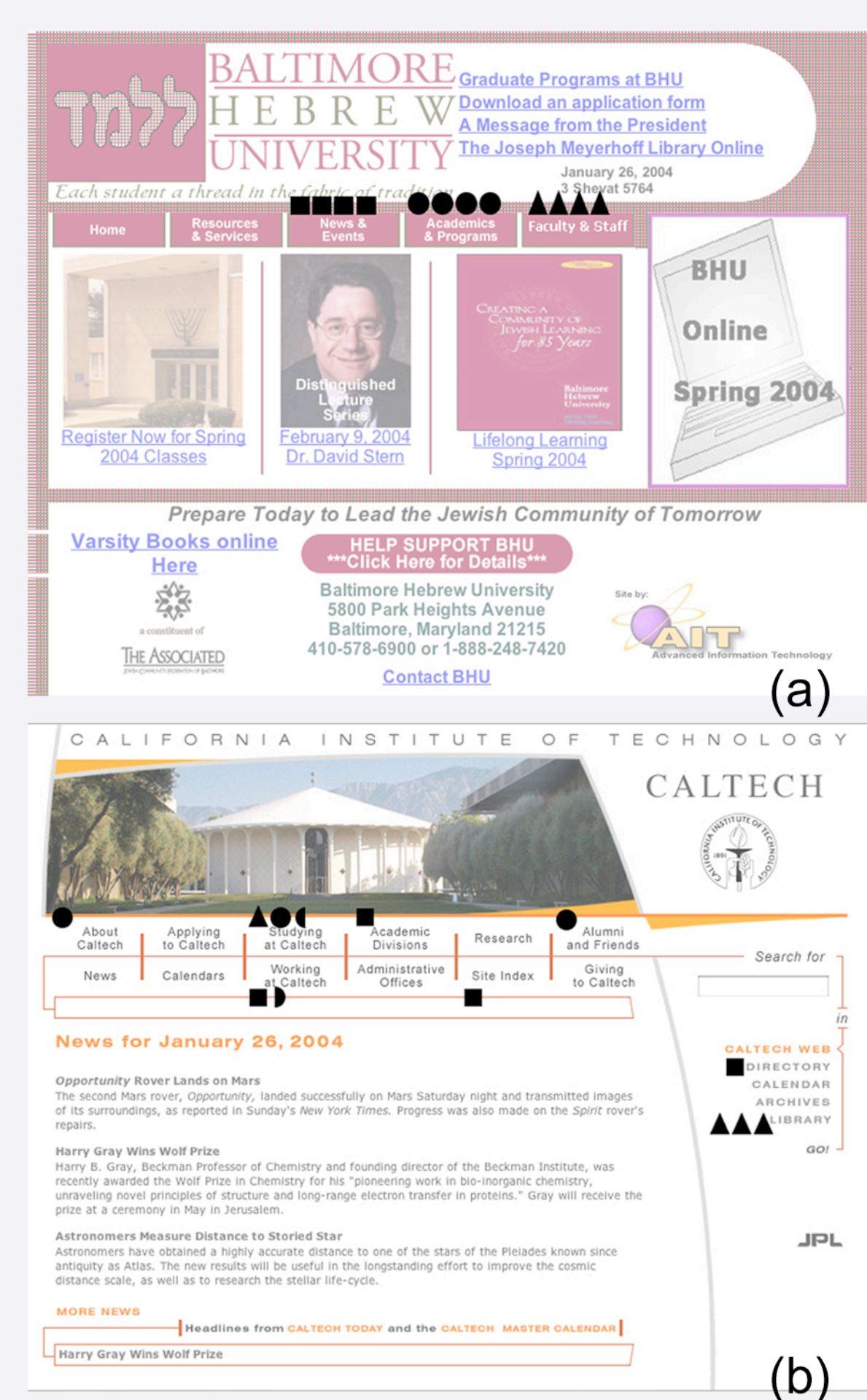
(A) Percent correct, (B) precision and (C) recall as functions of window size for the best measures of the three semantic systems. For all dependent variables, PMI-IR performed best on all the three tests.

SURVEY OF USER BEHAVIOUR

We would like to test a number of semantic systems to compare their performance in predicting user selection behavior. To collect human data that could be used to compare these systems, we conducted a survey in which we asked participants what link they would select on a webpage given an information goal. For example, we asked them where they would click to find a publication from a professor in the Psychology department on a webpage.



Figures on the left shows the responses of the participants plotted on the web pages for both the (a) highest and (b) lowest agreement observed between participants. The differently shaped data points refer to different goal statements. It is evident that a very high agreement between participants exists in the first case, whereas in the second case, the responses are less clustered.



CONCLUSION

We have compared the three semantic systems, WordNet, LSA and PMI-IR, for predicting the link that humans would select given a goal on a webpage, a behavior that we captured in a user survey. Our study suggests that PMI-IR is the best system for making such a prediction. Research in automated display usability evaluation and prediction needs a comprehensive theory incorporating various aspects of visual search. Ecologically valid semantic analysis, such as in this paper, might lead to a more accurate prediction of where people would click on actual webpages.



REFERENCES

- Blackmon, M., Kitajima, M., Lewis, C., & Polson, P. (2002). Cognitive Walkthrough for the Web. *CHI 2002*, 1, 463-470.
- Brumby, D., & Howes, A. (2004). Good enough but I'll just check: web-page search as attentional refocusing. *ICCM 2004*, 46-51.
- Chi, E. H., Rosien, A., Suppattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J., Cousins, S. (2003) The Bloodhound Project: Automating Discovery of Web Usability Issues using the InfoScent(tm) Simulator. *CHI 2003*
- Furnas, G., Landauer, T., Gomez, L., & Dumais, S. (1987). The vocabulary problem in human-system communication. *CACM*, 30(11), 964-971
- Landauer, T. K., & Dumais, S. T. (1997). A solution to the Plato's Problem: The Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211-240.
- Miller, G. (1995). WordNet: A lexical database for English. *CACM*, 38(11), 39-41.
- Pierce, B. J., Parkinson, S. R., & Sisson, N. (1992). Effects of semantic similarity, omission probability and number of alternatives on computer menu search. *International Journal of Man-Machine Studies*, 37(5), 653-677.
- Turney, P. (2001). Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the Twelfth European Conference on Machine Learning*, 491-502