ELECTRONIC PUBLISHING, VOL . 3(4), 227–234 (NOVEMBER 1990)

# Node popularity as a hypertext browsing aid

R. PAUSCH AND J. DETMER
*University of Virginia*
*Thornton Hall*
*Charlottesville, VA 22903-2442, USA*

*e-mail pausch@Virginia.edu*

**SUMMARY**

**We have performed a user study where the popularity of each node in a hypertext database was presented with the links leading to that node. Popularity was computed by counting the number of users who had previously visited the node. Our users clearly incorporated popularity information in their decisions; we compare their browsing patterns with a control group for whom the popularity information was not provided. One possible use of popularity can be to offset the previously documented trait of users to over-select items near the top or bottom of a linear list. We document that popularity information affects user behavior, but we do not necessarily advocate its use. Incorporating popularity information raises other questions of design and ethics which are beyond the scope of this paper.**

KEY WORDS Hypertext Browsing Node relevance User study Popularity

## INTRODUCTION

Designers of hypertext systems are interested in techniques to improve the user's ability to interactively browse through information. Hypertext, originally envisioned by Vannevar Bush [1], now exists in many variations [2], but in any hypertext system the selection of links is the fundamental operation for navigating through information. When traversing linked nodes or frames, the viewer is constantly presented with path decisions. Our general research interest is finding mechanisms that assist the user with these decisions. Some hypertext systems generate icons to show the media type (text, audio, motion video) contained in potential nodes [3]. Other systems dedicate fixed screen areas for optional summaries of potential nodes [4]. Previous studies have measured the effectiveness of various mechanisms [5], and our study continues in this vein.

One extremely valuable piece of information to a browser which cannot be automatically generated is a review of a node's quality. Many existing systems provide annotation mechanisms that could support reviews [6], but they require human effort. A related concept is that of *popularity*, which we define as the relative number of times previous users have visited a node. This can be automatically generated and presented with links to help describe the nodes to which they lead. We use a count in this study, but elapsed time could also be considered.

Previous researchers have established that the need to identify with others is a basic human motivation [7]. The shared knowledge of large groups of people, dubbed *cultural literacy* [8], is essential for social communication involving assumptions. In this paper, we document the effect of this information on users, but do not address the social issues

0894–3982/90/040227–08$05.00 | *Received 16 July 1990*
 *Revised 8 November 1990*

that designers will have to address when deciding whether or not to include popularity information in hypertext systems.

We have performed a user study to determine whether users would alter their behavior if links to nodes contained popularity ratings. Because we believe this information will be more useful to casual browsers than fact-finders [9], we had our users perform a browsing task during the study. We have two basic hypotheses:

1. Browsers will incorporate this information in the decision making process.
2. This effect can offset the previously documented tendency of browsers to pick near the top and bottom of linear lists [10–12].

We display the popularity of a node with a numeric popularity indicator next to links leading to that node. The number is the relative number of times the node has been previously accessed relative to the most popular node in the database. Sample popularity indicators are displayed in Figure 1. The words shown in boldface represent links to other nodes. The **{72}** next to **service** shows that the node reached via the **service** link was visited 72 percent as many times as the most heavily visited node in the database. The **service** node seems to have been of more interest to previous users than the **exploration, development,** or **stratigraphic test** nodes. In a dynamic system, node popularity would be updated as users browse through the database, but our study used databases where the popularity indicators remained static during the experiment. Our subjects were undergraduates browsing a hypertext database in a supervised classroom setting with the specific task of preparing themselves for a quiz based on the contents of the hypertext database.

In the oil and gas industry, an **exploration {04}** well

is only one of many different types of wells, others

include **development {12}**, **service {72}**,

and **stratigraphic test {52}** wells.

*Figure 1. Example of popularity indicators at links*

Our basic result is that the subjects significantly altered their selections by visiting popular nodes more often and unpopular nodes less often than the corresponding control group. We measured how often individual nodes were visited but did not attempt to analyze multiple node paths made by the users [13, 14].

## THE EXPERIMENT

Our two-part experiment used the Hyperties system for the delivery vehicle for primarily textual databases. In the first part of the experiment, the control group browsed a hypertext database containing 248 nodes. All links had dummy popularity indicators shown as **{00}** after the link name to ensure consistent screen presentation with the experimental group. Logging software recorded how many times each of the nodes was visited by the members of the control group. We then normalized these totals against the most popular node in the database, forming a popularity range between 00 and 99. The experimental group browsed the same database with the popularity indicators produced by the control group. The independent variable was the use of popularity indicators and the dependent variable was how often each node was selected. Our hypothesis was that the experimental group would prefer the most popular nodes of the control group at the expense of the less popular nodes.

We also expected the control group to pick items near the top of screens, especially topics presented in linear menu form [15], such as our **Table of Contents** screen. When our experimental group used the database with popularity indicators present, we expected the picking-from-the-top behavior to be reinforced. To determine whether popularity indicators could offset this behavior, we performed a second part to the experiment where a group was given popularity indicators intentionally concentrated away from the beginning or end of the table of contents. This seeded database was generated by having a group of subjects browse with a specific study task that emphasized topics reachable via links in the middle of the Table of Contents. Although we could have seeded a database by inserting artificial numbers, we felt that using an externally motivated seeding group would produce more realistic results. A second experimental group then browsed the seeded experimental group. Our hypothesis was that the centrally located high popularity indicators would counteract the natural tendency to choose links located near the top or bottom of a list. The overall structure of the experiment is shown in Figure 2.

```
                 Dummy Indicators {00}    Real Indicators {00-99}
              ===================================================
                                         |
                   Control Group A            Experimental Group X
       PART 1                           -->
                       Browsing                   Browsing
                                         |
              ---------------------------------------------------
                                         |
               Seeding Control Group B        Experimental Group Y
       PART 2                           -->
                  Directed Task                   Browsing
                                         |
              ===================================================
```

*Figure 2. Overall structure of the experiment*

All groups of subjects browsed for ten minutes through a training database to become familiar with Hyperties and then had 30 minutes to view the experiment database. Groups A and B viewed the document with all popularity indicators set to **{00}.** The experimental groups X and Y browsed databases with popularity indicators produced by groups A and B, respectively.

We used a simple navigation model in the experiment. All navigation was done by selecting links with the mouse. The only exception was a single keyboard command to return subjects to the **Table of Contents** node. This node acted as the *home node* for any subject who became disoriented.

Seventy-one subjects were divided into four groups. Each of the four groups was a class section from an entry-level undergraduate course in computer literacy. The subjects' only prior experience with computers had been four hours of hands-on training on how to use a word processor. Roughly 10 percent were computer information systems majors with the rest having diverse majors including business, health science, and social sciences. Each group contained an approximate balance of males and females. The material on hypertext presented during the experiment was considered part of the course content and attendance was expected but not required. The students were not informed that their actions would be recorded; all data was logged unobtrusively by the software.

At the beginning of the lab period each subject was seated in front of an IBM-PC-AT type microcomputer equipped with a mouse, keyboard, and 14-inch VGA monochrome display. Directions were provided both verbally and in writing. Subjects were informed that they would have 10 minutes to use a training database called the **Hyperties Tour.** The first few minutes of this time included hands-on instructions on how to use the mouse to select links, with an assistant giving any necessary help to subjects. We described the popularity indicators along with other features of the system, and did not inform subjects that their node selections were being recorded. The subjects were informed that they would have 30 minutes to browse the **Hypertext: Hands On!** database, which would be followed by a brief quiz on the contents of the database. The subjects were told they would not be able to view more than a small fraction of the database in the time allotted. During the browsing phase, we were available to answer questions and provide help.

## RESULTS

Our overall result is that experimental groups X and Y were significantly influenced by the presence of popularity indicators. The null hypothesis would be defined in a plot of group A versus group X as a linear relationship with a positive slope of one and intercept at zero. Our expectations were that group X subjects, on seeing the choices of group A, would be influenced to avoid low popularity nodes in favor of high popularity nodes. A curve fitted to this X vs. A visits per node scatter plot would yield a low flat curve in the range of lower A values and a steep upward slope with higher X visit values at the other end of the A axis. We present results both on the specific links presented as a linear list in the **Table of Contents** node, and on the result over all nodes in the database.

### Group X vs. group A on the table of contents nodes

The **Table of Contents** node, shown in Figure 3, contains links to 14 articles and is the one screen guaranteed to be seen by all subjects. Therefore, the data for these 14 nodes was extracted and analyzed for confirmation of our hypothesis for links presented as a linear list.

Figure 4 illustrates the average number of visits to each of these nodes by control Group A plotted against the average visits per node by Group X. By example, the node 'Essential Concepts' was visited 2.3 times per subjects by Group A which was translated

```
  * CHAPTERS *

  Introduction {26}
1 Essential Concepts {71}
2 Applications {59}
3 System Design Issues {45}
4 Implementation Issues {31}
5 Authoring {22}
6 Systems {35}
7 Personalities {56}
8 Possibilities {70}
9 The End Is Just the Beginning {36}
  Appendix: About Hyperties { 0}
  Bibliography {29}
  Hyper Glossary {29}
  Epilog: The Making Of Hypertext Hands On! {14}
```

*Figure 3. Table of Contents node*

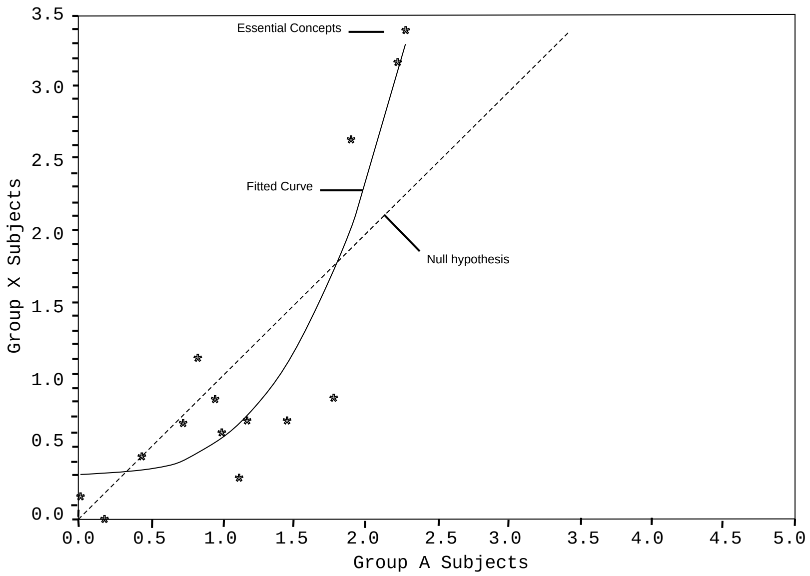


*Figure 4. Table of Contents scatter plot: Group X vs. Group A*

into the {71} marker (see Figure 3) viewed by Group X resulting in 3.5 average visits to this node. We found a 3rd degree polynomial relation curve fit for the data provided a coefficient of determination indicating an 86.104 percent explanation of resulting effect, confirming our hypothesis. This is plotted in Figure 4,along with the superimposed null hypothesis line.

## Group X vs. Group A on all nodes

The next step was to examine group X vs. group A average visits per node data for all 248 nodes. The home node was excluded from this process because it could be reached via a keyboard command. A large portion (230) of the nodes had popularity indicators less than 30 percent. Local analysis of the data from the cluster of nodes at the low end of the range indicated the possibility of a marker sensitivity threshold, but Figure 5,shows that again a polynomial of degree 3 provides a good fit to the data with a coefficient of determination of 0.7605.

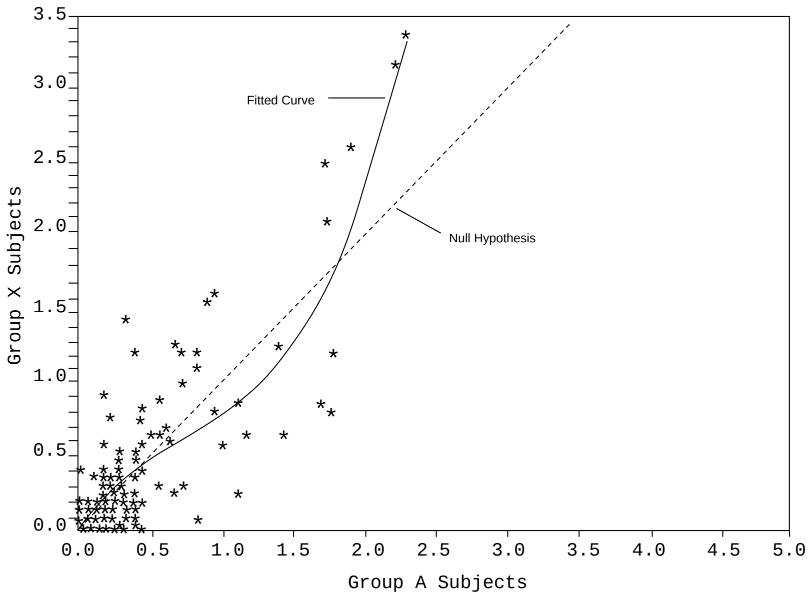


*Figure 5. All nodes scatter plot: Group X vs. Group A*

## Group Y vs. Group A

Group B was used to generate high popularity values for two topics that were generally of low interest to group A and located in the center of the linear **Table of Contents** node. Subjects in Group B were told that the **Implementation Issues** and **Authoring** chapters would be emphasized on an upcoming quiz. These nodes were indeed the most popular (both in number of visits and time spent reading) and also resulted in greatly lowering the popularity of other nodes in the database. When presented with the database seeded with these popularity indicators, group Y subjects showed a much more even distribution between topics than had group A. By considering group Y behavior to be the dependent variable influenced by not only the natural interests of the population as reflected in group A node visits, but also by the group B popularity indicators, we have two independent variables. Multivariate regressions on data sets containing just the the table of contents nodes and on all nodes produced high correlations of 0.94324 and 0.88979, respectively.

## DISCUSSION

The display of popularity indicators consumes screen space and research is needed on how to selectively show this information or display it in less space. We considered using small sliders next to each word, but were prohibited by screen resolution. Asterisks, icons, single-digit numbers, or variation in link color and/or boldness may be better than our two-digit popularity indicators.

We have not addressed many deeper issues indicated by this study. The first is the identification of popular *paths*, or multinode connections, through the document. While in general this presents a difficult problem in the definition of a common sequence, our techniques could easily identify which percentage of the previous users completed particular pre-defined sequences [14]. Another issue for future work is identifying the range, both in terms of group size and period of time the effect lasts, over which this group influence successfully occurs.

## CONCLUSIONS

Our initial hypothesis was that users would include information about popularity in their decision making process when moving from one node to another in a hypertext database. Groups X and Y altered their node selections based on the presence of popularity indicators, confirming the hypothesis. Group Y also established that this information could be used to counterbalance the previously observed tendency of users to over-select objects based on their position near the top or bottom of a linear list.

Because people use popularity information, designers may wish to consider providing it in future systems. While the sociological effects raise important issues, information about other browser's choices have obvious value in applications presenting consumer goods and general interest information. We also hypothesize that researchers may consider it valuable to know the previous usage pattern of reference materials, both to find heavily referenced works and to find materials that may have been previously overlooked.

## ACKNOWLEDGEMENTS

Hyperties is available from Cognetics Corporation, 55 Princeton-Hightstown Road, Princeton Junction, NJ 08550, USA, and we acknowledge their gracious contribution of software. Ben Shneiderman and Gary Marchionini of the University of Maryland supplied the hypertext source database and the custom logging software to make this experiment possible. Their comments on initial design and early drafts of this work were also greatly appreciated. A special thanks to Michael Kubovy of the University of Virginia Psychology department for his concern and valuable help in the design of the experiment. Matt Conway, Rich Gossweilier and three anonymous reviewers provided helpful comments on earlier versions of the paper.

## REFERENCES

1. V. Bush, 'As We May Think', *Atlantic Monthly*, **176** (1), 101–108 (1945).
2. J. Conklin, 'Hypertext: An Introduction and Survey', *IEEE Computer*, **20** (9), 17–41 (1987).
3. N. Yankelovich, N. Meyrowitz, B. Haan, and S. Drucker, 'Intermedia: The Concept and the Construction of a Seamless Information Environment', *IEEE Computer*, **21** (1), 81–96 (1988).

4. B. Shneiderman, 'User Interface Design fo the Hyperties Electronic Encyclopedia', *Proceedings of Hypertext '87* pp. 199–204 (1987).
5. Jakob Nielsen, 'The Matters that Really Matter for Hypertext Usability', *Proceedings of Hypertext '89*, Pittsburgh, PA pp. 27–42 (1989).
6. T. Catin, P. Bush, and N. Yankelovich, 'InterNote: Extending a Hypermedia Framework to Support Annotative Collaboration', *Proceedings of Hypertext '89*, Pittsburgh, PA pp. 365–378 (1989).
7. A. Becker and J. Carroll, 'Ordinal Position and Conformity', *Journal of Abnormal Social Psychology*, **65**, pp. 129–131 (1962).
8. E. D. Hirsch, *Cultural Literacy: What Every American Needs To Know*, Houghton Mifflin Company, Boston, MA, 1987.
9. G. Marchionini and B. Shneiderman, 'Finding Facts vs. Browsing Knowledge in Hypertext Systems', *IEEE Computer*, **21** (1), 70–80 (1988).
10. J. J. Tecce and B. T. Hamilton, 'Random Ordering of Candidates' Names on APA Ballots', *American Psychologist*, **27**, 967 (1972).
11. B. Shneiderman, 'Evaluating Three Museum Installations of a Hypertext System', *Journal of the American Society for Informational Science*, **40** (3), 172–182 (1989).
12. E. E. Wagner and T. O. Hoover, 'The Effect of Serial Positioning on Ranking Error', *Educational and Psychological Measurement*, **34**, 289–293 (1974).
13. R. Furuta and P. D. Stotts, 'Programmable Browsing Semantics in Trellis', *Proceedings of Hypertext '89*, Pittsburgh, PA pp. 27–42 (1989).
14. P. Zellweger, 'Scripted Documents: A Hypermedia Path Mechanism', *Proceedings of Hypertext '89*, Pittsburgh, PA 1–26 (1989).
15. F. R. Campagnoni and K. Ehrlich, *Getting There When You Don't Know Where 'There' Is: Navigation Strategies In A Hypertext Help System,* Human Interface Group, Sun Microsystems Inc., 1989.
16. B. Shneiderman and G. Kearsley, *Hypertext Hands-On!: An Introduction to a New Way of Organizing and Accessing Information,* Addison-Wesley Publishing Company, Reading, MA, 1989.