Testing a Browsable Interface*

Jamshid Beheshti (beheshti@libsci.lan.mcgill.ca)

Valerie Large

Mary Bialek

Graduate School of Library and Information Studies

McGill University

Introduction

Browsing is an integral part of human behaviour. Many life necessities are obtained by "browsing enormous, ever-changing ranges of goods in a multiplicity of locations."[1] It appears in different domains, one of which is information seeking activity. Browsing can occur in individual books, at library shelves or in catalogues. Users can initiate a search in an online catalogue (OPAC) through many access points such as author, title, subject or keywords. The system displays a number of index entries and users begin to browse through the list.

Many OPACs offer both querying and browsing. Querying involves exact keyword or phrase matching by utilizing Boolean logic, proximity and other operators, the result of which is 'all or nothing.'[2] This mode of inquiry is generally used for known item searches and high precision. Browsing consists of scanning lists of index terms, subject headings, shelf lists, or brief

*This research is supported by grants from Social Sciences and Humanities Research Council of Canada (SSHRC) and Fonds pour la Formation de Chercheurs et l'Aide à la Recherche (FCAR).

bibliographic records. It is most effective and practical when the search aim is not specific and the precise subject headings or descriptors are unknown. Querying puts a large amount of cognitive load on users. They have to cope with devising search strategies, conceptualizing appropriate search terms, coping with Boolean and proximity operators and managing large or zero number of hits. Browsing, however, requires very little cognitive effort.

Browsing has been the subject of discussion recently in library and information science literature. Chang and Rice have reviewed over 160 articles and studies to arrive at a conceptual framework for browsing. They define browsing as:

> the process of exposing oneself to a resource space by scanning its content (objects or representations) and/or structure, possibly resulting in awareness of unexpected or new content or paths in that resource space.[3]

They state that the general model of browsing consists of four major components: context, external and internal factors, browsing process, and consequences, or outcomes. They conclude that a clear taxonomy of browsing has not yet been devised.

Browsing in the online evirnoment is primarily *visual* and *is dependent on patterns and shapes* which are presented on the screen.[4] Direct manipulation of objects on the screen conforms to existing syntactic/semantic models of human-computer interaction. These models are based on the notion that when objects are displayed on the computer monitors, the need for learning and using complex commands, the syntax of which might create more obstacles for users, are

substantially eliminated. Each direct manipulation would immediately result in a visible reaction from the system and therefore reduce cognitive burdens. Hence, "the task semantics dominate the users' concerns, and the distraction of dealing with the computer semantics and the syntax is reduced."[5] Visual display of objects also provide clues and along with auditory and tactile ones are crucial factors in human information processing which have been ignored in the information retrieval systems.[6] One method of visualizing information is by simulating physical entities such as books, shelves, and physical layout of the library. The advantage of direct simulation is that users are presented with familiar images with which they have had contact since childhood. Bates states that "creating a virtual physical layout on the screen may make it easier for the searcher to think of moving among familiar categories of resources in an information retrieval system, in the same manner in which they move among resources in the actual library."[7]

PACE

Public Access Catalogue Extension (PACE) is an interface designed specifically for browsing. It is based on simulating images of books and book shelves on the screen. The data for generating images are derived from MAchine Readable Catalogue (MARC) records. TAG 3XX in MARC contains the physical description of a book, including number of pages and height. This information is used to generate an image of a book the dimensions of which are proportional to the actual size. The book spine, title page and verso of the title page are rich

sources of information for library patrons. PACE presents this information in a simulated environment, mimicking users' mental images of books and libraries, and therefore reducing the cognitive burdens of decoding and encoding of displayed information. The spine of each book appears on the screen as a three dimensional image with call number and full or truncated title. Each book is assigned a colour using a hashing algorithm based on the publisher and title in order to distinguish it from adjoining books. About ten books are displayed at a time on a simulated shelf. Depending on their width (number of pages), fewer or more may appear on the screen. A mouse is used for navigation allowing the user to "click" on the spine of a book so that a simulated title page is displayed. This page contains the full title, author, call number, publisher, publication date and LC subject headings derived from the MARC record.

The first phase of this three-year project is now complete. Two different versions of PACE were created for testing the effectiveness of the interface. The first version (PACE1) uses multi-shelf and single shelf displays. It does not include any menus or navigational guides. The second version (PACE2) provides the user with a menu and a single shelf display. The fundamental difference between the two versions lie in their approach to retrieving information; PACE1 relies on a purely sequential approach where the user has to move along the shelves in a linear fashion. PACE2 is a hybrid system providing the user with alternate access points from which the shelves can be browsed. Both interfaces have a menu bar at the bottom of the screen allowing users to zoom in, zoom out, move back (left arrow), move forward (right arrow) and go back to the opening

screen (HOME).

Currently, the second phase of the project is under development and will be tested in early Fall 1994.

Methodology

The main objective for testing PACE was to observe and explore users' behaviour in this browsable information environment. Information retrieval system (IRS) testing and evaluation is a difficult and complex task, particulary in the context of a new interface. Recently, attempts have been made to develop more effective methodologies than the traditional measures of recall and precision for testing IRSs. The focus of these new methodologies is on the user rather than the system. Meadow suggests that the "most useful of the current measures if the objective is overall assessment of an IR process" is user evaluation.[8] Tague and Schultz propose a model based on 'informativeness' of a system, which they define as the "subjective judgment by the user" with regard to the relevance of information for his/her particular need. Informativeness depends on the "user's personal space-time context."[9] Hancock-Beaulieu, Robertson and Neilson place users at the center of any evaluation methodology for assessing online catalogues.[10] Shneiderman's "five measurable human factors" also rely on the user's evaluation of the system. He recommends that retrieval tests consist of open-ended and closed questions, measurement of time to learn the new system and speed of performance, rate of errors by users, and assessment of subjective satisfaction.[11]

In this study, a number of different factors relating to the interaction between users and IRSs have been measured and evaluated. An experimental design was used to observe user behaviour. Four groups of students searched IRSs; two used a traditional online catalogue and two groups utilized PACE.

a) Database

For testing purposes a small database based on the Library of Congress "Z" collection at McGill University was created. The library system's online catalogue, called MUSE, uses the NOTIS system and contains over 1.5 million records. The "Z" collection which deals with libraries and librarianship, contains approximately 14,000 records in the Z100 to Z999 section. Since it was not logistically feasible to use these particular records, a sample of 2,048 records was selected from two cataloguing utilities to simulate the actual records in MUSE. The range of the sample was from Z100 to Z999, each category proportionally representing the same number of books as in the McGill Libraries.

b) Subjects

The Graduate School of Library & Information Studies at McGill University offers a two-year MLIS program. Students who had just completed their first year of study were chosen for testing for several reasons. First, they are familiar with online systems and OPACs as they are required to take an introductory course in *Database Development and Information Retrieval* in their first year of study. Second, MLIS students are expected to be more critical than

students from other disciplines when evaluating an OPAC interface, since evaluation of online systems is part of their training. Third, their common training in the first year of MLIS ensures a basic knowledge of computers and online searching. Fourth, they are considered to be a relatively homogeneous group which means that prior experience effect should be minimal.

A total of 26 students were selected based on their availability and divided into four groups. Volunteers were compensated for their time with a $20 fee.

c) Procedure

Of the original 26 students, two were chosen for a pretest of the procedures and questionnaires, and were not included in the results. The remaining students were divided into four groups: MUSE1, MUSE2, PACE1, and PACE2. The first group, MUSE1, used the MUSE system with all its searching capabilities to retrieve information. This group had access to full Boolean and proximity operators, author, title, subject, keyword searching and call number browsing. The second group, MUSE2, also used MUSE, but was restricted to call number browsing only. Groups PACE1 and PACE2 used PACE versions 1 and 2 respectively. PACE1 consists of two screens: the first screen displays several shelves, about fifty books in all; the second screen displays one shelf of about ten books. On the first screen, students can move back and forth fifty books at a time by clicking on an arrow. Once they have reached the desired section of the "Z" collection, they can zoom in to view the second screen. PACE2 also consists of two screens; the first screen displays a menu of

subcategories of the "Z" class, i.e., Z100, Z200, Z300...Z900; the second screen displays one shelf with approximately ten books. Students can choose individual subcategories on the first screen by clicking them and move directly to the beginning of that particular shelf on the second screen.

Each student was interviewed before and after the experiment by one of the two research assistants involved in the project. A structured interview schedule was used to record students' previous online experience, their library use habits and demographic data. Each student conducted a search to retrieve information to answer a single question which was presented to them within a specific context. Students were asked to retrieve up to 15 books on "What is Library Science." This was a vague or "fuzzy" question designed to accommodate browsing. Although they were all familiar with the topic, none had had the occasion to research it for their courses. They were asked to record the call numbers of books which they found to be relevant to the topic. Each student was timed and video taped. Another structured interview schedule was used after the search to solicit students' subjective evaluation of and comments about the particular system they had used.

Each student was given a general outline of the LC "Z" schedule indicating the sub-categories and their headings. A few students in the MUSE1 and MUSE2 groups who had never used the call number browsing feature of MUSE were given brief instructions. While the PACE1 group required less than two minutes of training, students in the PACE2 group needed a few minutes more of instruction. In general, the learning curve for PACE can be assumed to be

insignificant. The time to find the first desirable call number was recorded for each student.

Results

Table 1 shows the results of all the searches by all the subjects. For the MUSE1 group, three additional columns appear in the table indicating the initial search strategy used (*first search*), how many different strategies were used (*# of search*), and whether the student made any simple mistakes such as missing brackets or inappropriate use of Boolean and proximity operators (*basic mistake*). Two columns need additional explanation. RP is relative precision and is defined as the ratio of the number of relevant books selected by a student to the total number of relevant books selected by all students using the same system. RPT (relative precision total) is defined as the ratio of the number of relevant books selected by a student to the total number of relevant books selected by all students regardless of the system used.

Relevance is a very illusive concept and has been the subject of much discussion in the literature of library and information science.[12] In this study, relevant books were defined as those books which were selected more than once. This measure of relevance is based on the assumption that if a book is chosen by more than one student in small experimental samples, then it might be considered relevant to the topic. Figures 1 and 2 show the frequency of books chosen by all the students and by experimental groups. Except for the PACE2 group, less than 40% of all the books chosen were selected more than once.

Figure 3 shows the frequency rank distribution of books by experimental groups. Books chosen by the three groups have logarithmic rank distributions ($R^2 > 0.990$), while PACE2 group's distribution is exponential ($R^2 = 0.977$). The concentration of books chosen by the latter group is different from the other three experimental groups. Analysis of variance also confirms the above result; PACE2 is different from other groups for both measures RP and RPT ($F(3,23) = 5.96$, $p = 0.005$, and $F(3,23) = 3.80$, $p = 0.026$ respectively).

In spite of their differences, as Figure 4 indicates, most of the top ranking books chosen by students are distributed evenly among the experimental groups. All the call numbers have the same prefix, Z665, and deal with librarianship and information science. There are a few exceptions, however, which can be attributed to the differences between the interfaces. Whereas the PACE1 group was confined to sequential browsing, the PACE2 group used a hybrid method of searching. The MUSE2 group was limited to call number browsing, while the MUSE1 group used keyword and subject searching. Each interface can compliment and enhance the other - no one of them provides a complete retrieval system.

Speed of performance was measured by recording the total time for the search and the time it took each student to retrieve the first call number. Both analysis of variance (ANOVA) and Kruskall-Wallis tests show no significant difference among groups with respect to the amount of time spent conducting the experimental task ($F(3,20) = 2.39$, $p = 0.099$ and $H = 4.72$, d.f. $= 3$, $p = 0.194$). On average, the MUSE2 group spent marginally more time than the other groups,

while the PACE2 group utilized less time (Figure 5). No time limits had been set for searching, and the range was very wide from a low of 11 minutes to a high of 76 minutes. Only a few students chose 15 books which was the maximum allotted number of choices. For all the subjects, search duration does not have a strong correlation with the number of books chosen ($R^2=0.07$). Search time does not show a definite correlation with RP ($R^2=0.10$), or RPT ($R^2=0.03$). Therefore, total search time is not directly related to the number of relevant books retrieved. It should be noted that the PACE database was a subset of McGill University's online catalogue, and therefore direct comparisons between the two systems may be misleading.

Performance was also measured in terms of the time taken to locate the first required call number. This is an important measure because it might be directly related to users' level of frustration. Figure 6 shows the amount of time each group spent at the terminal to find the first call number. ANOVA and Kruskall-Wallis show no significant difference among groups ($F(3,20)=1.53$, $p=0.238$ and $H=5.16$, d.f.$=3$, $p=0.161$). The times ranged from a low of just under two minutes to as high as 11 minutes. Again, the PACE2 group performed marginally better than the other groups, with an average retrieval time of just over three minutes. These times do not show strong correlations with RP ($R^2=0.06$), or RPT ($R^2=0.08$).

Error rate for the PACE groups was insignificant. Three out of seven students in the MUSE1 group committed some basic errors, such as not including brackets or using incorrect Boolean operators. One student in the MUSE2 group

used an imprecise call number to browse through the database and wasted 11 minutes before she realized her error.

Students attitudes towards the systems they used were recorded on a five-point Likert scale, with the higher numbers indicating more positive attitudes (Figure 7). Some significant differences ($0.01 < p < 0.1$) were observed among the four groups in terms of how easy they found the system to use, how easy they found the display to read, and how friendly they found the system. Attitude towards the usefulness of the system for the particular experimental task was evenly distributed among all groups. PACE1 was seen as the easiest system to use with a friendly display. MUSE1 had the worst scores except in terms of its usefulness for this search.

An example from each experimental group is used to illustrate the characteristics of searches, based on the data from Table 1. Figure 8 is a flow chart of one of the student's searches in the MUSE1 group (Subject #7). This student began the search by choosing subject searching and used a total of seven different strategies. Call numbers were recorded during the fifth search strategy after spending six minutes and 11 seconds on the system (*first time* column in Table 1). The student chose 9 books, the first one of which was Z665 E37. Total search time was 18 minutes and 19 seconds and the student did not make any basic mistakes. The student's relative precision (RP) compared to other subjects in the two MUSE groups is 44%, and the relative precision total (RPT) is 67%. A flow chart for Subject #10 in the MUSE2 group would be much simpler showing an input box, a process box and a decision box. This student

used a call number search and after four minutes and 49 seconds retrieved the first book, Z665 C7465. The entire search took 48 minutes and resulted in recording 15 call numbers, 40% of which were relevant compared to other students in the MUSE groups, and 60% relevant compared to all students. Subject #17 in the PACE1 group began the search by "clicking" on the forward arrow at the bottom of the screen and sequentially moving through stacks of 50 books to stop at the beginning of Z665 range. The student then zoomed in to scan shelves of ten books. After seven minutes and 23 seconds, the first book, Z665 C7465, was recorded. The student spent 38 minutes on the system, retrieving 11 books with RP of 45% and RPT of 90%. Subject #23 in the PACE2 group began the search by selecting the option "Z662 - Z680" from the opening menu. The student moved quickly to the Z665 section of the catalogue and chose the first call number, Z665 B476. Eleven minutes later, this student had chosen seven books with RP of 86% and RPT of 86%.

The results and the examples provided here suggest that the PACE2 group used marginally less time and chose less books with higher relative precision. All subjects in the PACE2 group recorded a book in the Z665 B section of the database. These findings suggest that a hybrid system consisting of a menu and graphical interface may lend itself to a more concentrated search, resulting in fewer, more precise number of items chosen.

Conclusions

PACE is a graphical interface based on simulating images of books on the

screen and is designed for browsing. PACE is an interface based on the existing mental model of users. Although this study is limited to a particular group of users and a specific database, it shows that PACE performs as well as, if not better, than an existing online catalogue. The interface's performance is enhanced by the addition of a menu, indicating that a hybrid system may be a more effective and efficient retrieval system than a purely visual one. Retrieval performance as measured by the amount of time spent to obtain the first call number and the total search time were not significantly different among the experimental groups. Students' subjective attitudes show their preference for a visual display of information based on their existing mental models. Although PACE1 was considered the easiest system to use, it was also deemed the least useful for retrieving specific information.

Some of the limitations of this study such as the limited number of records in the PACE system and the sample of specific user group, will be addressed in the second phase of the project. A new browsable retrieval engine is added to PACE and it will be tested in a small college library with a more diverse population and utilizing the library's entire database.

References

1. Alan Poulter. Towards a virtual reality library. *Aslib Proceedings*. 45(1), 1993:11-17.

2. Charles R. Hildreth. *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems*. Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1989.

3. Shan-Ju Chang and Ronald E. Rice. Browsing: A multidimensional framework. In: *Annual Review of Information Science and Technology* (ARIST), 28, 1993. Martha E. Williams, ed. Medford, N.J.: Learned Information, Inc., 1993.

4. J. F. Cove and B.C. Walsh. Online text retrieval via browsing. *Information Processing & Management* 24(1), 1988:31-37.

5. B. Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Don Mills, Ont.: Addison-Wesley Publishing Company, 1987.

6. M. E. Rorvig. Psychometric measurement and information retrieval. *Annual Review of Information Science and Technology*. Martha E. Williams, ed. 23, 1988.

7. Marcia J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13(5), 1989:407-424.

8. Charles T. Meadow. *Text information retrieval systems*. Toronto: Academic Press, Inc., 1992. p.292.

9. Jean Tague and Ryan Schultz. Evaluation of the user interface in an information retrieval system: A model. *Information Processing & Management* 25(4), 1989: 377-389.

10. M. Hancock-Beaulieu, S. Robertson and C. Neilson. *Evaluation of online catalogues: An assessment of methods*. British Library research paper 78. British Library Research and Development Department. The British Library Board, 1990.

11. B. Shneiderman.

12. see *Journal of the American Society for Information Science* 45(3), 1994.

Table 1. Characteristics of searches

group	sub#	first search	first time	first call number	# of Search	total time	books	basic mistake	RP	RP total
MUSE1										
1	1	sub	3:35	z665 B43	3	19:31	10	N	70	90
1	2	key	5:00	z665 A28	7	23:41	9	N	56	67
1	5	key	9:50	z669 A6	21	52:11	12	Y	67	83
1	6	key	7:34	z674.8 E44	9	33:38	15	Y	60	60
1	7	sub	6:42	z665 E37	7	18:19	9	N	44	67
1	8	key	6:11	z665 C7465	11	21:23	5	Y	20	40
1	11	sub	3:34	z665 C7465	4	20:09	15	N	47	87
MUSE2										
2	13		2:01	z665 A544		10:42	14		43	64
2	9		11:47	z665 A544		20:23	6		67	67
2	10		4:49	z665 C7465		48:01	15		40	60
2	12		5:18	z665 A28		41:59	15		60	87
2	14		2:14	z665 B43		76:20	15		47	73
PACE 1										
3	15		6:01	z665 C538		15:28	8		75	100
3	16		7:40	z658 U5 F694		31:36	15		20	33
3	17		7:23	z665 C7465		38:04	11		45	90
3	18		2:28	z721 L635		16:57	13		69	69
3	19		4:05	z665 B476		17:34	13		54	54
PACE 2										
4	20		2:30	z665 B43		13:09	12		67	92
4	21		3:52	z665 B476		17:30	6		100	100
4	22		4:13	z665 B43		25:02	8		75	88
4	23		1:58	z665 B476		11:12	7		86	86
4	24		2:30	z665 B43		21:12	12		92	100
4	25		3:26	z665 B476		15:00	8		75	100
4	26		2:00	z665 B87		20:00	9		78	100

Figure 1. Frequency of all books chosen

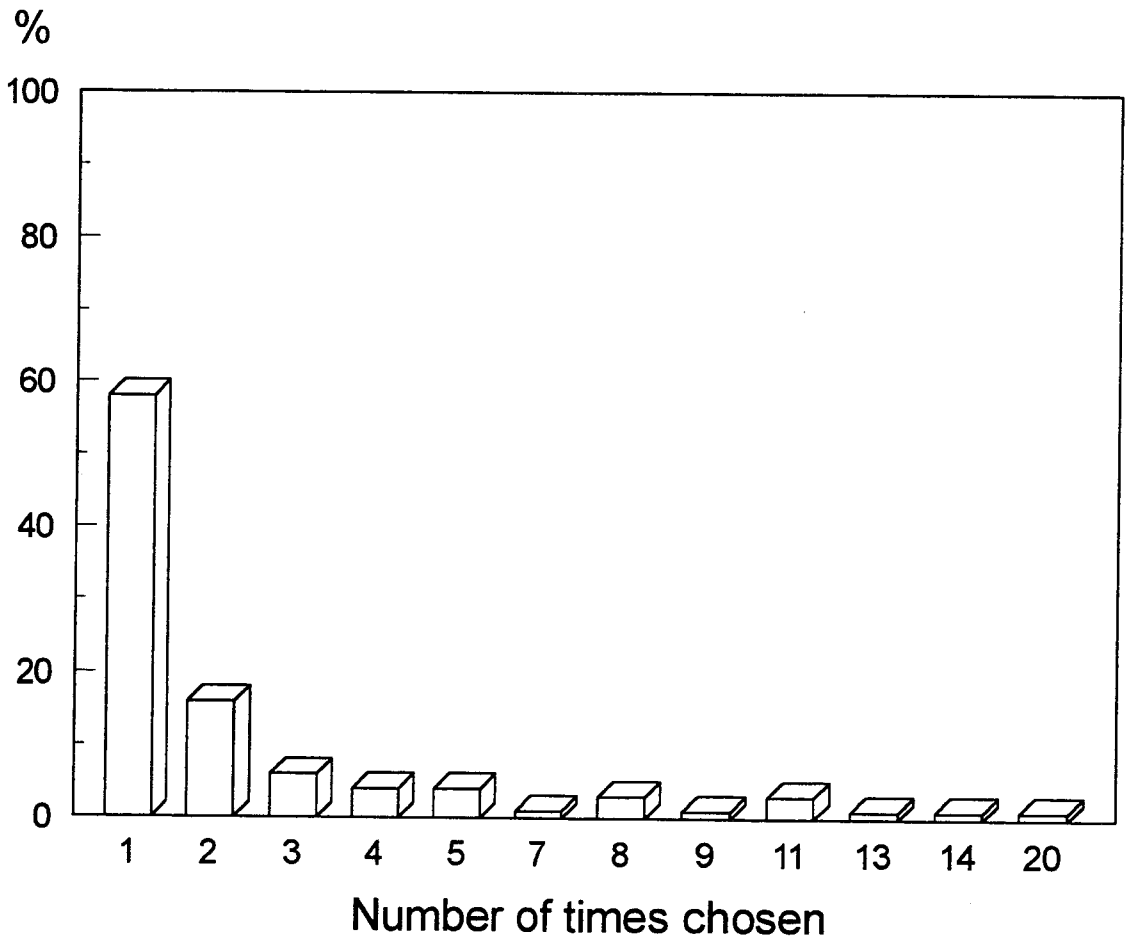

Figure 2. Frequency of books chosen

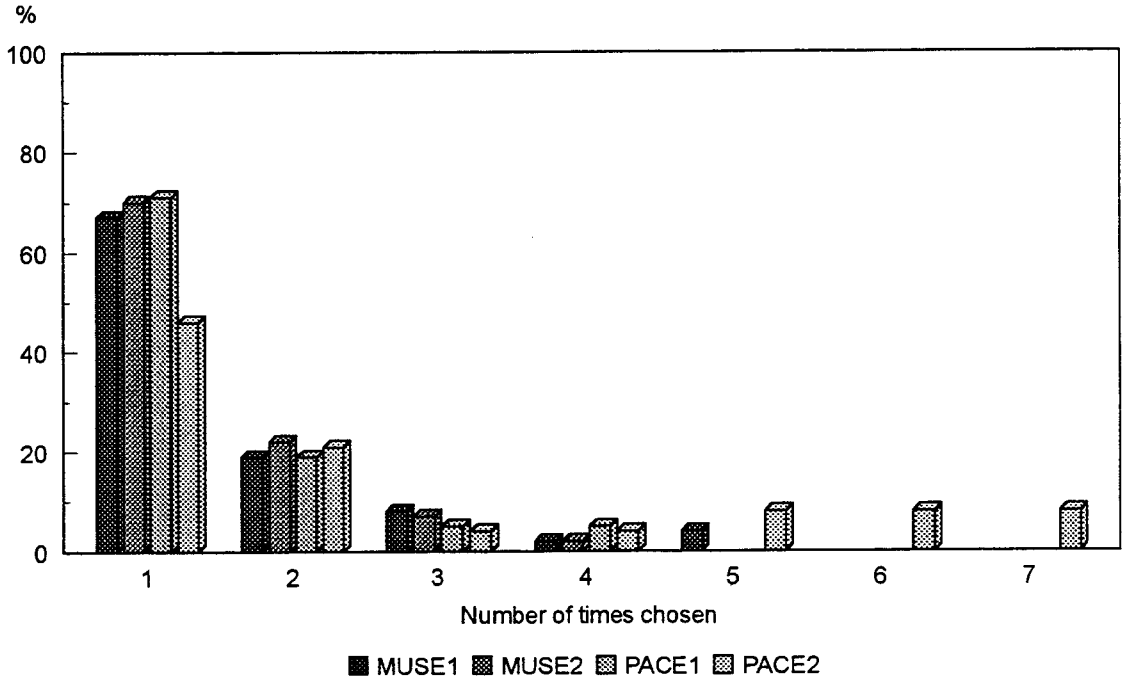

Figure 3. Frequency rank distribution

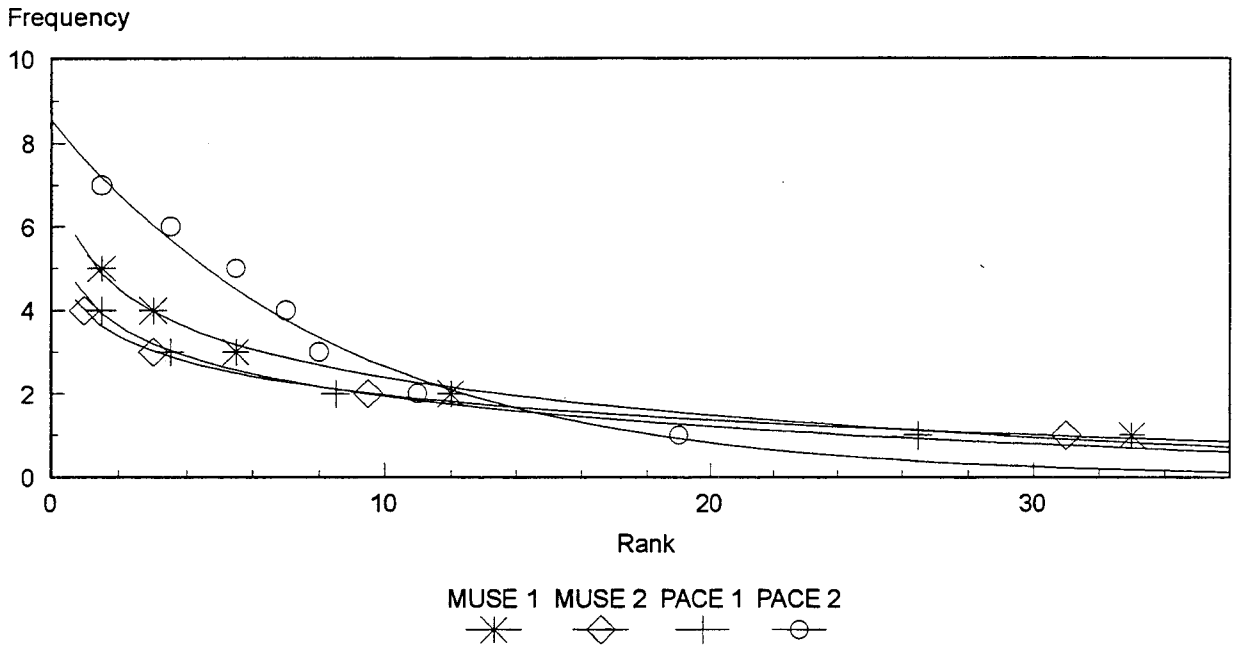

Figure 4. Frequently Chosen Books

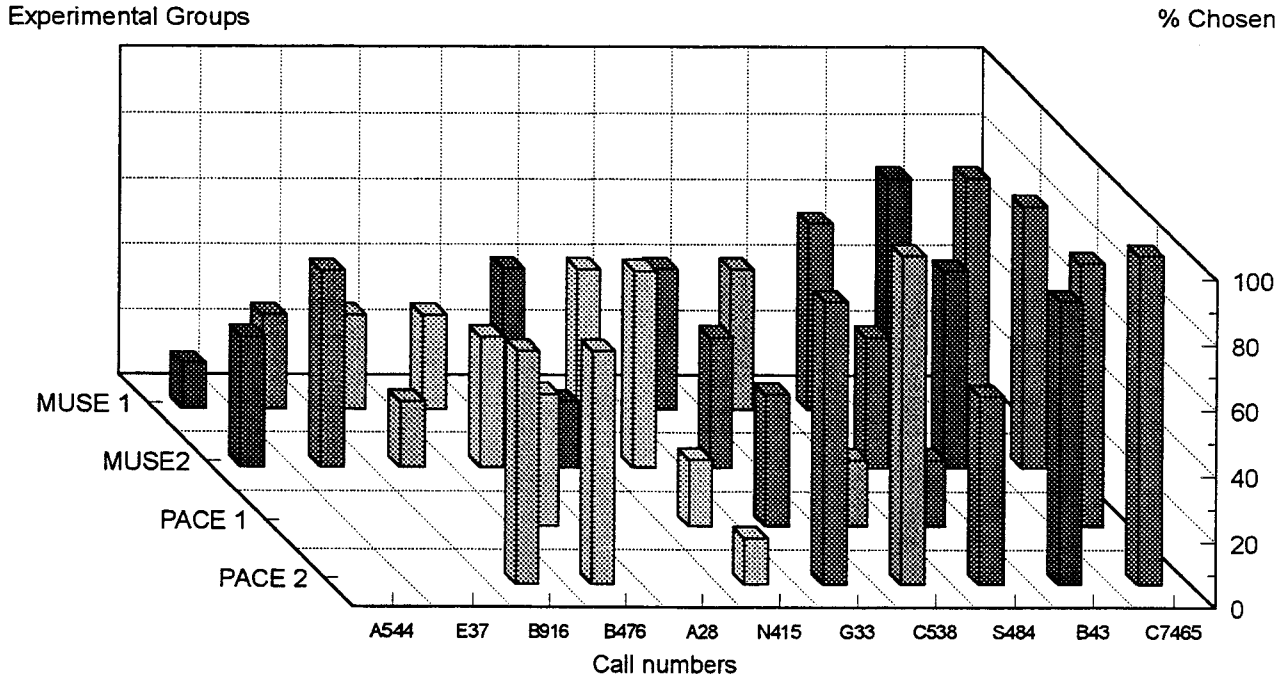

Figure 5. Total search time

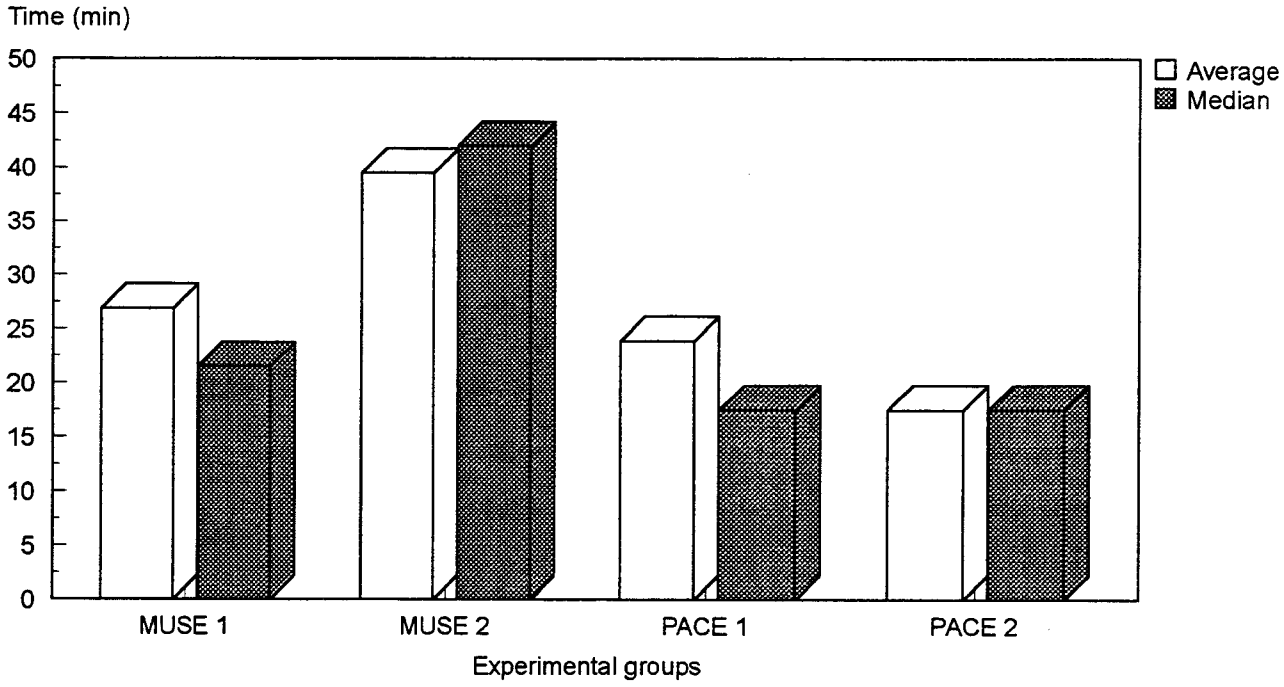

Figure 6. Time to locate first call number

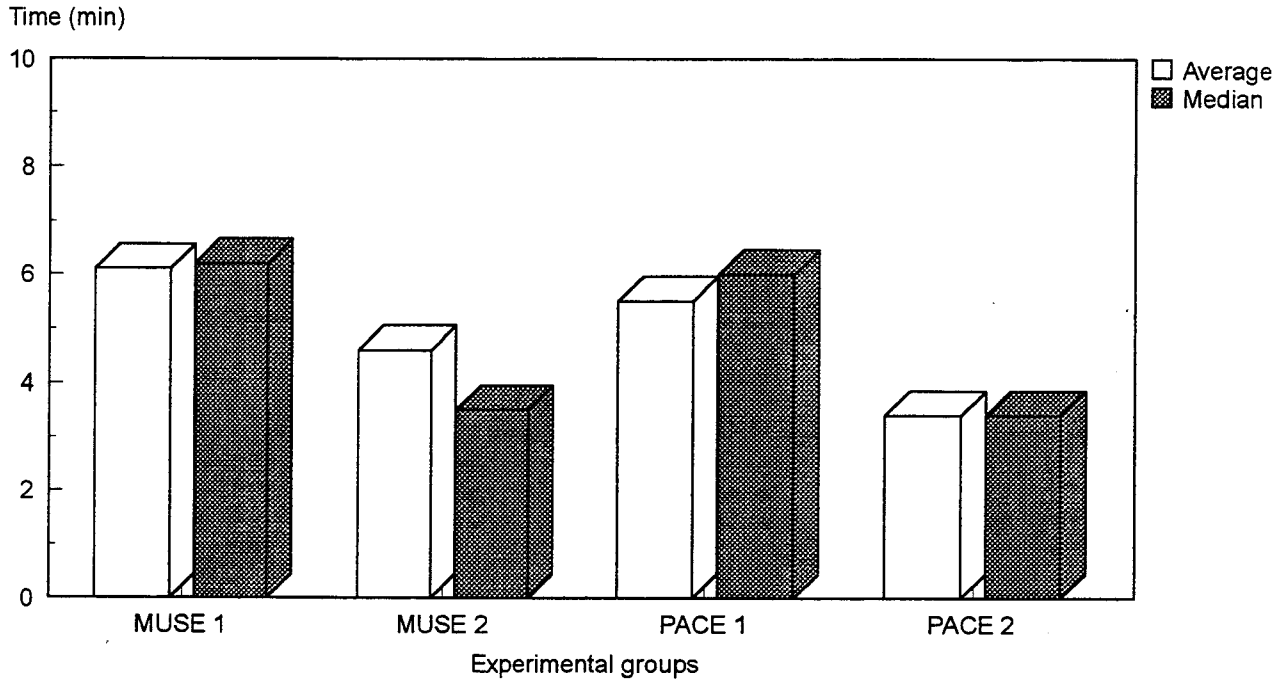

Figure 7. Attitudes of Experimental Groups

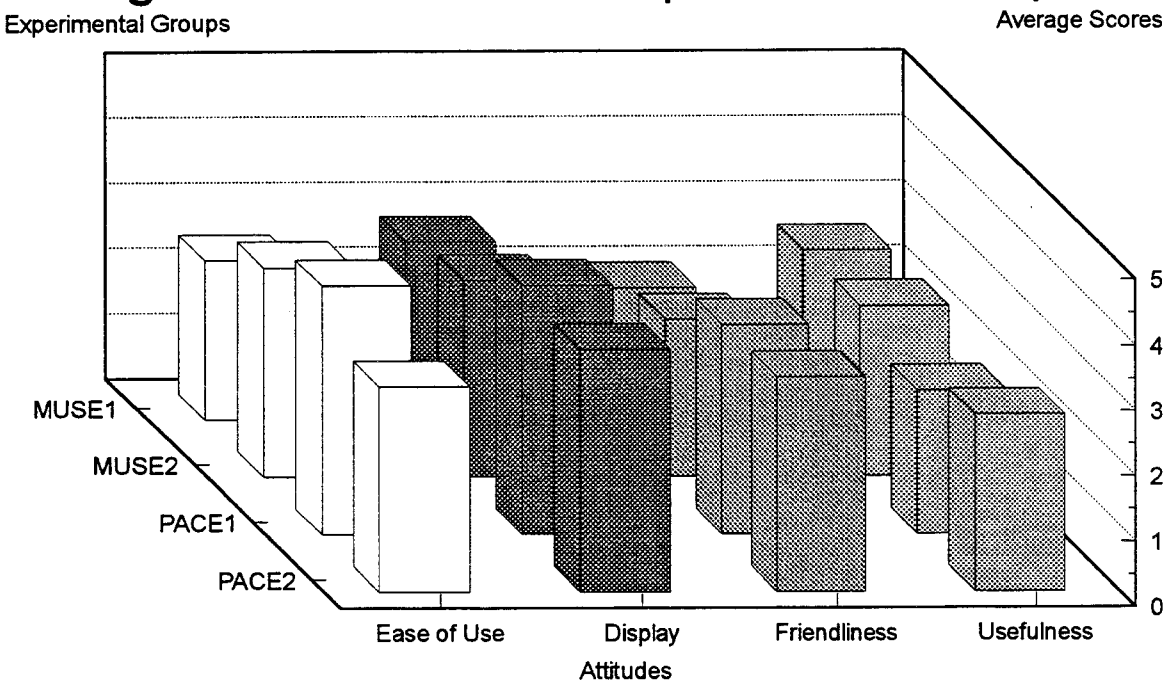

Figure 8. Flow Chart of Search Characteristics for Subject # 7