TELIDON

TELIDON BEHAVIOURAL RESEARCH 2

THE DESIGN OF VIDEOTEX TREE INDEXES

Behavioural Research and Evaluation
Department of Communications

TELIDON BEHAVIOURAL RESEARCH 2
THE DESIGN OF VIDEOTEX TREE INDEXES

May 1981

TELIDON BEHAVIOURAL RESEARCH 2
DEPARTMENT OF COMMUNICATIONS
OTTAWA
MAY 1981

THE DESIGN OF VIDEOTEX TREE INDEXES

Preface

Videotex data bases are being developed using hierarchical or tree-structured indexes. It is clear from the reports in this volume that a poorly designed index can make it difficult, or in some cases impossible, for users to find information even if it does exist in the data base.

There has been very little study of this retrieval method and guidelines to assist index designers in constructing easy to use indexes have not yet been developed.

These reports from the Department of Communications' Behavioural Research laboratories are a first step toward developing a set of guidelines for constructing tree-structured indexes. The reports indicate that people do indeed make a substantial number of errors in choosing index items when they are searching for specific information. Some recommendations for improving the tree-index can be made based on this work. As well, the methods used for testing tree indexes can be adapted for testing commercially developed indexes. Further work on these issues is continuing at the Department of Communications. Four reports are presented here.

1. "The Effectiveness of a Tree-Structured Index when the Existence of Information is Uncertain"

   explores the search process when the participants are uncertain that information exists in the database to answer their questions, a situation that would resemble normal home use. Results showed that people do make errors in their index choices (using a simulation of the Telidon demonstration data base index) and most errors occur on the first two levels. People also stop searching before they find existing information.

2. "The Use of Tree-Structured Index which Contains Three Types of Design Defects"

reports an experiment in which design defects were purposely added to the database. Again, a simulation of the Telidon demonstration database index was used. Design defects were miscategorization of information, two synonymous labels on a page and vague category labels. Miscategorization was found to be the most serious defect leading to longer search times. The other defects also impaired performance.

3. "An Investigation of User Search Performance on a Telidon Information Retrieval System"

reports on an experiment for the first time using an actual Telidon data base (again the Telidon demonstration data base at the Department of Communications). People made many errors in finding information; indeed, error rates were higher than in the previous experiments where simulations were used. As with previous experiments, many errors occurred in the first two levels of the tree.

4. "The Design of Videotex Tree Indexes: The Use of Descriptors and the Enhancement of Single Index Pages"

reports on two experiments designed to evaluate the effect of adding descriptors (brief descriptions of each index term) to index items, to test the "first" or basic index page in a number of ways and to demonstrate methods for testing and modifying tree indexes.

It should be noted that the experiments in this report were conducted using the index of the Department of Communications' demonstration Telidon data base when it was in a very early stage of development. Many changes have been made since then. This means that the absolute figures reported, such as error rates, shold not be taken to generalize to other Telidon indexes. However, the experiments do allow the authors to draw conclusions about the principles of design of a tree structure index. A summary of the conclusions is incorporated into the guidelines below.

Guidelines for Constructing Tree Indexes: Summary of Conclusions from these Experiments

1. Testing a tree index with naive people (not familiar with the data base) can improve it significantly. Methods used in these reports could be adapted for use by commercial data base providers.

   - testing the top two levels of the tree is most efficient if resources are limited
   - testing should be done with the target population
   - testing can be done with the "single-page method" (i.e. devising several versions of a page and measuring performance and preference)
   - both preference and performance measures are required as far as we know with present evidence

2. Index pages can be improved by avoiding errors and taking account of user preferences.

   - miscategorization of information is a most serious design defect, leading to errors or user's inability to find the information at all
   - ambiguous, vague category labels or two synonymous labels on a page lead to errors in searching
   - users prefer that equal amounts of information be placed under each label (not 80% under one category)
   - users prefer more than 4 index items on a page and probably find information faster than when only 2 or 3 index items appear on a page
   - users prefer some rational ordering of index items, i.e. alphabetical, or in order of most frequently used

Dorothy Phillips
Behavioural Research and Evaluation
Department of Communications
300 Slater Street
Ottawa, Ontario K1A 0C8
May 1981

TELIDON BEHAVIOURAL RESEARCH 2

MAY 1981

THE DESIGN OF VIDEOTEX TREE INDEXES

Contents   Page

CHAPTER I

THE EFFECTIVENESS OF A TREE-STRUCTURED INDEX WHEN THE EXISTENCE OF INFORMATION IS UNCERTAIN

Thomas Whalen and Susane Latrémouille

THE EFFECTIVENESS OF A TREE-STRUCTURED INDEX WHEN THE EXISTENCE OF INFORMATION IS UNCERTAIN

SUMMARY

Eight people were each asked to find 16 items of information with a hierarchical index system implemented on a time-shared computer. They were told that some of the information that they were asked to find was not available. Overall, the probability of the subject making an incorrect selection on an index page was 0.15. Most of these errors occurred on the first two levels of the tree. These results agreed with those of previous studies. In addition, this experiment showed that most searches ended when the desired information was found after accessing a mean of 4.52 index pages. If the information was not found after a mean of 8.78 pages, the participants terminated the search without finding the information. Thus, it was concluded that people are very likely to stop searching before they find existing information, rather than searching the database extensively. This implies that errors in the design of the index structure will disrupt a search of the Telidon database.

INTRODUCTION

One of the first uses for Telidon is information retrieval from a general database by untrained individuals. While it is generally agreed that constructing a hierarchical index system is one possible way to implement this, there is very little data showing how easy it is to use. This is particularly true when the user does not know if the information he desires is in the database or not.

The first section of this paper describes the results of an experimental investigation of the use of a heirarchical index system when some of the information that the user is trying to find does not exist. The second section provides a comparison of these results with a previous experiment. A detailed description of the experimental procedure is included in the Appendix.

PROBABILITY OF A SUCCESSFUL SEARCH

Each of the eight participants in this experiment was presented with 16 tasks, each requiring finding a specific item of information by making selections from hierarchical index pages sequentially displayed on a computer terminal. Of these problems, 25% were insolvable because the information did not exist on the database. The participants found the information on 86.5% of the solvable problems. Thus, the probability that the participants would find the information was very high, even when they knew that a prolonged search might prove fruitless.

It is surprising that these people so seldom terminated a search before finding the information. The reason, however, is that they were successful in finding the information quickly when it was available. For those problems which were solvable, a perfect search required accessing a mean of 3.58 pages. When the information was found, they had accessed a mean of 4.52 pages. Thus, the participants found the information almost as quickly as was possible. On the average, they looked only at one extra page.

On the first problem that each participant terminated unsuccessfully, they accessed a mean of 21.87 pages. This dropped to 10.12 pages on the second. This suggests that a learning process was occurring. In order to draw conclusions about the participant's normal search strategy, this first unsuccessfully terminated problem was not included in the following analysis. Ignoring the first problem that was terminated unsuccessfully by each participant, they accessed a mean of 8.78 pages before terminating an unsuccessful search. Thus, the participants searched about twice as far when they were unable to find the information as when they did.

Overall, then, the index system is practical when the existence of information is uncertain because people are able to find information easily, not because they are willing to conduct exhaustive searches of the database. This implies that the design of the index pages is important. If people could not find information with a poorly designed index system, they would quickly conclude that the information is unavailable.

PAGES WITH HIGH ERROR RATES

Of all the instances in which a correct choice was available to a participant, that choice was selected 85.3% of the time. This seems to be a high probability of selecting a correct item, but it must be considered in the context of a search of a hierarchical index system. To obtain an item of information from the database, a number of index pages must be accessed. As the database grows in size, this number increases. If the probability of making an error is 15% on each page, the probability of correctly selecting 10 index items in a row is less than 20%.

This analysis assumes that the errors are distributed uniformly throughout the index pages. Most of the errors, however, occur on a small number of pages. In this experiment, there were 25 pages on which an error could have occurred but errors occurred only on 11 of these pages. There were only one or two errors on most of these, so that, in all, 45 of the 57 errors occurred on only 16.0% of the pages. One of these pages was the root of the tree, and the other three were directly below it. This suggests that increasing the depth of the tree by adding index pages to the bottom may not further increase the difficulty in finding information in the database. This possibility must be verified by further experimental research.

COMPARISON WITH PREVIOUS RESEARCH

A previous Technical Memorandum by Lee and Latremouille (1980) describes how people were able to find information when searching a tree-structured arrangement of index pages. Their study was conducted with menu pages typed on 5x7 inch cards. Each participant in their experiment was asked to find information using this index in order to evaluate the tree structure. They also collected subjective statements from the participants in order to determine the causes of the errors when they occurred.

There are differences in the way people search for information between this experimental procedure and that used with Telidon. First, the participants in the experiment knew in advance that the information for which they were searching actually existed in the database. When using Telidon, however, most often it is not known initially if the information desired exists in the database or not. Thus, the user must also make a decision about continuing the search for the desired information after seeing the outcome of each choice.

Second, the participants in their experiment received immediate feedback about whether they were correct or not after each choice. This provided the participant with information that was required to solve the problem at hand. This ensured that participants did not make the same error twice in a row, though they could make different errors of the same type repeatedly. A Telidon user, however, will consistently make the same incorrect choice on a page if he misunderstands that page.

Thus, Lee and Latremouille's experiment was conducted under different circumstances than occur in normal Telidon use. To better represent this situation, the present study used a simulation of the Telidon demonstration database on a large, time-shared computer. The index pages were displayed on a standard video terminal and items were selected by the participants through a keypad on the terminal keyboard. The index pages used in Lee and Latremouille's experiment were essentially the same as were used in the present experiment, though the problem tasks they used were unrelated to those chosen for the present experiment.

Considering the magnitude of the differences between the two experiments, there was a surprising degree of similarity in their results. Lee and Latremouille reported that the probability of an error was 14%, whereas the present experiment found a probability of 14.7%. It is undoubtedly accidental that these two values are as close as they are, but it does indicate that the difference in experimental methods did not have a major effect on the results. This is important in validating Lee and Latremouille's procedure: index pages can indeed be tested by typing them on cards and asking people to make choices from them; the expense of incorporating them into a complete database is not necessary.

These two experiments also showed substantial agreement in locating the majority of the errors in the first two levels of the tree. Overall, Lee and Latremouille found that 80% of the errors occurred on 6 of 79 menu pages, and that 58% of the errors occurred on the first two levels of the tree.

Thus, the general conclusion of the two studies is the same: a hierarchical index system may only be useful if it is carefully designed. As the database increases in size, it becomes more important to ensure that there is the lowest possible probability of a user making an error on each page.

REFERENCE

Lee, E. and Latrémouille, S. Evaluation of Tree-structured Organization of Information Telidon. Telidon Behavioural Research I. Department of Communications, February, 1980.

APPENDIX: THE EXPERIMENTAL METHOD

SUBJECTS

None of the 8 volunteers who served in this study had used Telidon previously.

PROCEDURE

Each participant was instructed how to use the simulation of Telidon. First, he was shown the solution of a practice problem, then asked to solve 16 problems without the experimenter's intervention. The experimenter did not intervene while the participant was attempting to solve these 16 problems.

Solving each of the 16 problems required that the participant find a specific item of information in the database. The information for 12 of these problems was determined by selecting an item at random from consecutive index pages until arriving at a leaf of the tree. The actual problems used are listed in Table 1. The problems were worded so that there was no reference in the problem to the phrases found on the index pages. The information was not available for the other 4 problems. The problems were presented to each participant in a different order in such a manner that sequential effects of the presentation of the problems would cancel between participants.

The simulation consisted of the display of successive index pages on a Volker-Craig 404 video computer terminal. Each index page consisted of a list of up to nine numbered phrases. Each phrase described the information that would be found in documents subsumed under that label. Selecting one of these labels by pressing the appropriate number on the terminal's keypad resulted in the erasure of that index page from the screen and the display of another index page. This following index page comprised items which were more specific categorizations of the information which was described by the label on the previous page. Thus, a series of choices of descriptions of the information would produce ever more specific descriptions until the participant had arrived at the most specific description possible. Selecting one of these most specific labels would either result in a message informing the participant that he had found the requested information, followed by the presentation of the next problem, or would return the participant to the most general index page so that he could try again.

In addition to the digits which appeared on the index pages, the participant could press a zero which would return to the last index page that he had viewed. Repetitive presses of zero would take him further back through the pages that he had already viewed. If the participant forgot the problem that he was trying to solve, pressing "P" would redisplay it. Pressing "S" stopped the search and presented the next problem.

TABLE 1

This is a list of the problems presented to each participant. The number preceding each problem indicates the number of pages which would be accessed in a perfect search. If this number is zero, then the data was not present on the database.

5	What are the latest scores in the National Football League?
4	What is the weather like in Manitoba?
4	Find the second international news story listed.
0	Find a recipe for fried chicken.
3	What are the names of the ministers in the Government of Saskatchewan?
4	What is the horoscope for Scorpios today?
0	What are the recent winning numbers for Loto Canada?
3	Find a bedtime story.
3	Find out where to borrow money.
3	Find out about Diefenbaker when he was Prime Minister.
0	What drug stores will be open this Sunday?
2	Find the telephone number of the Police.
4	Find out how pages are numbered on Telidon.
4	Play the Star Trek game.
4	Find a list of craft fairs.
0	Find a map of Switzerland.

TABLE 2

These are the four pages on which participants made the most errors. The first page is the root of the tree, and the other three are obtained by selecting items 1,3 and 6 from the first page. The first number which precedes each item is the total number of errors participants made when selecting that item.

12	1	PERSONAL ACTIVITIES
1	2	BUSINESS
1	3	GOVERNMENT
5	4	NEWS, WEATHER AND SPORTS
1	5	EMERGENCY
1	6	UNDERSTANDING TELIDON

3	1	WHERE TO GO FOR ENTERTAINMENT OR RECREATION
0	2	PERSONAL MEMOS
0	3	TRAVEL
3	4	PERSONAL HELP AND ADVICE
0	5	THINGS TO DO AT HOME
0	6	SCHEDULE OF EVENTS
0	7	EDUCATIONAL OPPORTUNITIES
0	8	TABLES AND STATISTICS

0	1	HOUSE OF COMMONS
0	2	THE SENATE
0	3	THE JUDICIARY
0	4	FEDERAL REPORTS
7	5	CANADIAN PRIME MINISTERS
0	6	MORE PRIME MINISTERS
0	7	PROVINCIAL GOVERNMENT
0	8	MUNICIPAL GOVERNMENT

4	1	HOW TO USE TELIDON
0	2	TELIDON EXPLANATIONS
7	3	INDEX OF TELIDON PAGES

CHAPTER 2

THE USE OF TREE-STRUCTURED INDEX WHICH CONTAINS THREE TYPES OF DESIGN DEFECTS

Thomas Whalen and Candy Mason

THE USE OF TREE-STRUCTURED INDEX WHICH CONTAINS THREE TYPES OF DESIGN DEFECTS

SUMMARY

The present experiment examined how different classes of defects in the design of a hierarchical index system affected the retrieval of information. An understanding of the relative importance of these defects will assist information providers in designing an index for a videotex system.

A tree-structured index system, very similar to that used in the Telidon videotex demonstration database, was altered so that it contained four instances of each of three types of design defects. The first type of design defect was the miscategorization of information under an inappropriate category label. The second type of design defect was the inclusion of two synonomous category labels on the same page. The third type was the inclusion of a vague category label on a page, which was uninformative to the user.

Each of 21 civil servants were asked to find 16 different items of information using a computer program which simulated a Telidon tree-structured index system. Comparisons of several measures of the efficiency of these searches on this defect-laden index structure with a previous study revealed the same consistent pattern of results. While all of the defects impaired the efficiency of the searches, the effects of the synonomous and ambiguous labels were much smaller than the effects of a miscategorization.

These results prompted the following recomendations. Each of the three types of defects should be avoided, because each had a detrimental effect on every measure of search effectiveness used in this study. In situations where the risk of a defect is unavoidable, however, a more general category label is preferred to an overly specific label, because

this will lead to synonomous or vague category labels which have less detrimental effects on information retrieval than miscategorizations. In all cases, an index system must be tested empirically to discover the presence of defects, because they may not be obvious to the index designer but may have severe detrimental effects if left in the index structure.

INTRODUCTION

The present experiment measures the decrease in the probability of successfully retrieving information and the increase in the length of searches caused by defects in the design of a database index system.

The index system studied consists of the arrangement of a series of menu pages in a hierachical structure. This type of index system, often called a tree-structured index, has been proposed as the best index system for information retrieval by naive, infrequent users of a database, and, therefore, best for Telidon-based and other videotex systems.

In any indexed database, the efficiency of a search for information will depend upon the design of the index system. Any part of the index system which causes difficulty for the user is a defect by definition, regardless of how logical or reasonable it is from the designer's point of view. This situation is likely to arise because the semantics of natural languages are not fixed, but vary from one individual to another. Thus, something that is a good example of a category for one person may not even be in that category for another person (Rosch, 1975).

The possiblity that some parts of an index system may be reasonable in principle but still contain defects by this criterion may make the design of index systems very difficult. The only way to know if some part of an index system contains a defect for most users is to actually test whether a sample of users can find information using that part of the index system.

In previous studies (Lee and Latrêmouille, 1979; Whalen and Latrêmouille, 1980), it was inferred that defects were present in the index system because the users had difficulty finding the information. It was possible to guess what the defect was in these cases, but impossible to be certain that that guess was correct. Thus, it was impossible to tell which types of defects were most detrimental to information retrieval. The only way to ensure that the greatest part of an observed effect is the result of a particular design defect is to deliberately design a defect into the database which is more severe than the other, incidental defects. Therefore, the present study examines the severity of the effect of deliberate design errors in an experimental setting.

Three types of defects were tested. The first type of defect is a miscategorization. This occurs when the selection of one item on an index page retrieves a second index page with an item which the user would expect to appear under a different item on the first index page. For example, if the selection of "MARKETPLACE" produced another index page which did not contain any items about craft fairs, but the selection of "ADVERISING/MARKETING METHODS AND SERVICES" produced a page with an item marked "CRAFT FAIRS", we would say that "CRAFT FAIRS" had been miscategorized.

The second type of defect is synonymous labels. If two items on an index page have the same meaning to the user, then he will not have any reason to select one over another on that page. For example, if the items "AMUSEMENT" and "ENTERTAINMENT" occur on the same page, most people would say they were synonomous labels.

The third type of defect is a vague label. If an item on an index page does not have any clear meaning to the user, that item is vague. For example, the label "MISCELLANEOUS INFORMATION" is vague because it does not clearly indicate the type of information that is found beneath it. It is important to note that this is not the same thing as a label which means something different than the user expected. In that case, the label does have a meaning for the user, but all of the items under it were miscategorized.

Clearly these three types of defects are not mutually exclusive. It is possible, for example, to have "MISCELLANEOUS INFORMATION" and "GENERAL INFORMATION" as synonomous, vague labels on a page.

In order to study these defects, four examples of each type were deliberately incorporated into an index structure which had been studied previously by Whalen and Latrêmouille (1980). This variant of the index structure was then tested in the same way as previously. The problems presented to the participants were the same, as were all other aspects of the study, except for the number of subjects. This allowed easier comparison of the outcome of various searches with the results of the previous study.

In both this and the previous experiment, the subjects were told that some of the information for which they would be asked to search would not be present in the database. Thus, they would terminate the search for an item of information if they were unable to find it. In the previous study it was found that the probability of making an error was high in the course of a search for information. If the information was present, however, most participants found it, rather than concluding that it was not in the database.

METHOD

PARTICIPANTS

Twenty-one employees of the Department of Communications each participated in individual experimental sessions lasting approximately one hour.

PROCEDURE

First, each participant was instructed in the use of the Telidon simulation, then, following a practice problem, was asked to solve sixteen experimental problems.

These problems were the same as those used in the previous experiment by Whalen and Latrémouille (1980). They required that the participant find a specific item of information from the database, but were worded so that the problem did not contain terms used in any index page except on the bottom level. For 12 of the problems, the desired information could be found in the database, but the remaining 4 problems required information that was not present in the database. In the instructions, the participants were told that some of the problems asked for information which was not available, but were not told which problems, nor how many there were. Problems were presented in one of four different orders so that the effects of presentation order were minimized. The wordings of the 16 problems are presented in Appendix A.

The simulation itself involved a display of successive index pages on a Volker-Craig 404 video computer terminal. Each index page consisted of nine or fewer numbered alternatives. Except for the changes required to introduce the four instances of each type of defect, these index pages were the index for the Telidon demonstration database as of November 1979, which was also used in the previous study by Whalen and

Latrémouille (1980). The exact location of each defect may be found in Appendix B.

When an alternative was chosen by pressing the selection number on the terminal's keypad, another page of index items subsumed under that selection would appear. In addition to the digits which accompanied selection alternatives on the index pages, the participant could press "0", which would return him to the last index page that he had seen. Repetitive presses of "0" would return him to the trunk of the tree, as would selecting an item at a leaf of the tree. Pressing the letter "P" would redisplay the problem for the participant, and the search could be terminated by pressing "S".

RESULTS

The results of this experiment are summarized and compared with the results of the previous experiment by Whalen and Latrémouille (1980) in Table 1.

OVERALL PERFORMANCE

Participants found the requested information for 150 out of the total of 252 problems for which a solution existed. For the other 84 problems, the information requested did not exist in the database, so the participants correctly terminated these problems. Thus they correctly terminated or solved 234 out of 336, or 70%, of the problems. They incorrectly terminated 102 of 252, or 40% of the problems which had a solution available. These are different from the results of a previous experiment by Whalen and Latrémouille (1980). In that experiment, the probability of an incorrect termination was 13.5%. Thus, there was a greatly increased probability that a problem would be incorrectly terminated when defects were built into the database.

Of all instances where a participant made a selection from a page with a correct choice available, they selected it 60% of the time in the present experiment. Thus, there was a 40% mean chance of an error per page. In two previous experiments (Whalen and Latrémouille, 1980; Lee and Latrémouille, 1979) the mean probability of an error per page was about 15%. Thus, overall, there was a greatly increased probability that the user would make incorrect selections when the index structure contained defects.

A measure of how long the partcipant looked for the information is the number of times that he selected an index item to produce another index page. For problems in which the participants were able to find the information, they accessed a mean of 10.6 pages. On these problems

TABLE 1

Summary of the comparisons in performance in the present experiment and that of Whalen and Latrémouille (1980).

	A	B	C	D
Overall, present experiment	40.0%	40.0%	10.6 pages	58.5%
previous experiment	13.5	15.0	8.3	24.2
Miscategorizations,				
present experiment	95.2	0.01	14.0	94.3
previous experiment	16.6	-	4.4	28.7
Synonomous labels,				
present experiment	17.8	45.0	7.0	9.7
previous experiment	8.4	-	5.7	7.3
Vague labels,				
present experiment	14.2	44.0	3.7	17.2
previous experiment	12.5	-	3.0	8.7

Column A: Probability that a problem will be terminated when the information is available

Column B: Probability that an incorrect selection will be made on a page

Column C: Number of pages searched when the requested information was found

Column D: Proportion of total pages accessed during searches which terminated without finding the information available perfectly errorless searches could have required retrieving a mean of only 3.58 pages.

Because the participants in this experiment had had no prior experience in searching through the database, changes were found in the in the number of pages searched as the participants became more familar with the task and the database. The mean number of pages accessed by each participant in searching for the information on the first problem that he was able to solve correctly was 14.0 On all subsequent problems in which the information was correctly found, the mean number of pages searched per problem decreased to 5.65. Thus, the participants learned to find information faster in the database after their first success.

On the first problem in which each participant incorrectly terminated his search, that is, he terminated his search without finding the information when it existed in the database, he searched a mean of 14.81 pages. On the second terminated problem, however, this mean dropped to 10.24 pages, on the third the mean was 9.6 and the fourth question the number dropped to 6.9 pages of search.

In the correctly terminated search instances, when the data was not available, participants first searched a mean of 16.81 pages, but this dropped to a mean of 11.81 over following search attempts. Thus, in both terminated searches and those in which data was found, the participants learned to search through fewer pages after they had aquired some familiarity with the database.

PERFORMANCE AT THE SITES OF THE DEFECTS

When the correct choice was miscategorized under a label, the probability of selecting the correct alternative was less than 1%. When a problem required the choice of a synonymous category, the probability of selecting the correct alternative was 45%. When the first synonomous category of the pair was the correct choice, this probability was further reduced to 37%. When the page contained an vague label that was not to be chosen in a correct search path, the probability of selecting the correct alternative was 44%.

As expected, the most severely disruptive of the defects was the miscategorization of information. When a choice is found under a label which clearly does not identify it, a user is very unlikely to choose it. Also as expected, the probability of correctly choosing a synonomous category label is less than 50%. It would be exactly 50% if one of the labels was always chosen, but on occasion, the user will fail to choose either of the synonomous category labels. It is interesting that the participants choose the second synonomous category label more often than the first. This suggests that the user does not simply read the selections until he sees one that fits, but reads past a selection that fits, at least until he finds another one.

COMPARISON WITH PREVIOUS DATA

In order to assess the effects of the defects on searching more completely, those searches which required an encounter with a defect are compared to searches for the same information in the previous experiment (Whalen and Latrémouille, 1980) which used this index system before the errors were inserted. In the present study, three of the problems required finding information which had been miscategorized, six required finding information which had been placed under a synonomous category heading, and two required finding information which had been placed under a vague heading.

On those problems which required finding information which had been miscategorized, the participants in the present experiment terminated 60 of the 63 searches. Thus, the probability of terminating the search without finding the information was 95.2%. For the same problems, the participants in the previous experiment terminated 4 of the 24 searches. Thus, the probability of terminating the search without finding the information was only 16.6%. When a miscategorization was inserted into the path of the search, the probability of terminating the search without finding the information increased by 78.6%.

On those problems which required finding information which had been categorized under one of two synonomous headings, the participants terminated 15 of the 84 searches or 17.8%. In the previous experiment, when the information required by these problems was not categorized under an ambiguous heading, the participants terminated 3 of 32 searches or 9.4%. Thus, inserting synonymous category headings in the path of the search increased the probability of an incorrect termination by 8.4%.

On those problems which required finding information which had been placed under a vague category heading, the participants terminated 6 of 42 searches on these problems or 14.2%. In the previous experiment, the participants terminated 2 of 16 searches or 12.5%. Thus, inserting vague category headings in the path of the search increased the probability of an incorrect termination by 1.7%.

On those problems in which the information was found, a measure of the difficulty of the search is the number of pages that the participant accessed before finding the information. In the present experiment, the participants accessed a mean of 14.0 pages before finding the information on those problems in which the information had been miscategorized. In the previous experiment, the participants accessed a mean of 4.4 pages on these same problems. Thus, the insertion of a miscategorization resulted in a mean increase of 9.6 pages in the search for the desired information. This was a 122.7% increase in the length of the search.

When the information was categorized under one of a pair of synonymous headings, the mean number of pages searched before the information was found was 6.99, whereas, in the previous experiment, the mean number of pages searched had been 5.72. This was a 22.2% increase in the length of the search.

When the information was categorized under a vague heading, the mean number of pages searched before the information was found was 3.75, whereas, a mean of 3.00 pages had been searched in the previous experiment on these problems. This was a 25% increase in the length of the search.

OVERALL EFFICIENCY OF THE DATABASE

A general measure of the efficiency of the database is the proportion of the total number of pages accessed which result in incorrect terminations. This is effort expended which should have resulted in a successful search, but was wasted instead. In the previous study, 24.2% of the pages accessed during attempts to find information that was present in the database were accessed in searches which were terminated unsuccessfully. When the design errors were inserted into the database, this increased to 58.5%. Thus, far more than half of all of the pages accessed in searches for information which was present, failed to culminate in the retrieval of that information when design errors were present in the database.

Considering only those problems which involved miscategorizations, this proportion rose from 28.7% in the previous experiment to 94.3% in the present experiment. For those problems involving synonomous category headings, this proportion rose from 7.26% to 9.73%, and for those problems involving vague category headings, it rose from 8.70% to 17.17%. Thus, once again, the miscategorization of information is by far the most severe type of design error.

DISCUSSION

These results show a very consistent pattern. Any type of defect decreases the effectiveness of a search for information by any reasonable measure that one cares to apply. This is not surprising considering the severity of the defects which were incorporated into the database.

The defects are not all equally disruptive, however. The miscategorization of information is by far the most serious type of defect. The miscategorization of information makes a search for desired information practically impossible. On the other hand, category headings which are synonomous or ambiguous are surprisingly benign. They do cause some increase in the length of a search and in the probability of a failure to find the information, but these effects are very small, particularly considering the severity of all the design errors built into the database.

The classes of design defects considered here are related to each other. Once information has been clustered into a category, the selection of a label to describe that category will tend to produce miscategorizations if the label is more specific than the category merits, while it will be ambiguous and tend to overlap with other category labels if it is more general than required.

This observation in the context of the present results leads to the following recomendations:

1) The information must be gathered into as distinct and logical categories as possible. If it is not, one of the above defects will be inevitable, and the presence of any one of these defects will decrease the effectiveness of the index structure.

2) Once a category has been created, it must be labeled with the most accurate description possible, because, once again, any deviation from an accurate description will lead to a defect.

3) In the selection of a category label, if a perfect label cannot be found, it is important to ensure that the label chosen subsumes all of the information placed under it. This may increase the risk of making the label ambiguous or having it overlap with the label given to another category. These defects, however, are less severe than miscategorization.

4) When an index structure has been created, it must be tested empirically in order to determine the severity of the defects that may have been introduced. The present results show that the effects of a design defect may be severe. The design of an efficient index system will likely require several repetitions of these steps.

These recomendations only apply to the design of menu driven index systems in which any particular incorrect decision may be easily corrected by obtaining the previous menu if the decision is discovered to be an error. In other systems, such as keyword search, in which the consequences of selecting an overly general term may be an overwhelming volume of information, the relative importance of these types of defects may be very different.

Another possible constraint on the generality of these recommendations is that they are based on the most severe possible cases of each type of design defect. Lacking an empirical finding to the contrary, it must be assumed that the relative importance of each type of defect will be the same for less severe cases. Considering the magnitude of the effects found, this is a reasonable assumption, but the possibility remains that it may be shown to be incorrect by a future empirical study.

REFERENCES

Lee, E. A., and Latrémouille, S., Evaluation of tree-structured organization of information on Telidon. (Technical Memorandum No. BRG 79-12). Department of Communications, Ottawa. December, 1979. Also in Telidon Behavioual Research I. Department of Communications, Ottawa, 1980, pp.231-242.

Rosch, E., Cognitive representations of semantic categories. Journal of Experimental Psychology: General. 1975, 104, 3, pp 192-233.

Whalen, T. E., and Latrémouille, S., The effectiveness of a Tree-structured Index when the Existence of Information is Uncertain. (Technical Memorandum No. BRIC80-3). Department of Communications, Ottawa. 1980.

APPENDIX A

These are the problems which were presented to each subject. The numbers in parentheses indicate the minimum number of pages required to access the information. If that number is zero, then the information did not exist in the database.

1. What are the latest scores in the National Football League? (5)
2. What is the weather like in Manitoba? (4)
3. Find the second international news story listed. (4)
4. Find a recipe for fried chicken. (0)
5. What are the names of the ministers in the government of Saskatchewan? (3)
6. What is the horoscope for Scorpios today? (4)
7. What are the recent winning numbers for Loto Canada? (0)
8. Find a bedtime story. (3)
9. Find out where to borrow money. (3)
10. Find out when Diefenbaker was prime minister. (3)
11. What drug stores will be open this Sunday? (0)
12. Find out the telephone number of the police. (2)
13. Find out how pages are numbered on Telidon. (4)
14. Play the Star Trek game. (4)
15. Find a list of craft fairs. (4)
16. Find a map of Switzerland. (0)

APPENDIX B

This is a list of the design defects of each type which were inserted into the index structure.

I. MISCATEGORIZATIONS

1. "Sports scores" was moved from "News, weather and sports" to "Emergency".

2. "Map selection" was moved from "Travel" to "Personal help and advice".

3. "Craft fairs" was moved from "Marketplace" to "Advertising/marketing methods and services".

4. "Ministers of Saskatchewan" was moved from "Provincial government" to "House of Commons".

II. SYNONYMOUS LABLES

1. Both "Government" and "Industry" were changed to the label "Government and industry".

2. Both "Amusements" and "Entertainment" appeared on the same page.

3. The category for local news was divided into "First section" and "Second section".

4. Both "Leisure at home" and "Home leisure" appeared on the same page.

III. AMBIGUOUS CATEGORIES

1. "Miscellaneous information" appeared on a page.

2. "Other information" appeared on a page.

3. "Other news stories" appeared on a page.

4. "Miscellaneous Emergencies" appeared on a page.

CHAPTER 3

AN INVESTIGATION OF USER SEARCH PERFORMANCE ON A TELIDON INFORMATION RETRIEVAL SYSTEM

Scott A. McEwen

AN INVESTIGATION OF USER SEARCH PERFORMANCE ON A TELIDON INFORMATION RETREIVAL SYSTEM[1]

ABSTRACT

The development of computerized information retrieval systems has generated considerable interest and research into these systems in terms of how people will interact with them. The present experiment, which was the first search experiment to employ an actual Telidon database, investigated search performance with 24 naive users asked to locate information from the database pertaining to 16 search questions. The major dependent variables were search time and number of errors made. The results of the study showed that people experienced some difficulty in locating information from the hierarchically organized database. Problem areas within the tree structure were also revealed. These findings concur with the results from previous studies. The conclusion also was in agreement with these other studies: people may experience difficulties in searching for information from hierarchically organized databases similar to the database tested. On the basis of the present data, it is possible to identify where errors occur in the present tree and to suggest how to improve the tree.

[1] The author would like to thank the following people from the Behavioral Research Group of the Department of Communications for their comments and assistance: Eric Lee, Paul Muter, Thomas Whalen, and Susane Latrémouille.

INTRODUCTION

Within the past few years a number of videotex information retrieval systems have been developed. Some of the largest and most well known systems include Telidon, developed by the Department of Communications Research Centre in Ottawa, the British Prestel system, and the Antiope system from France. An increasing amount of interest and research into these systems has occurred as the day nears when people will use such systems in their homes.

Apart from all the recent technical advances that have made the building of these systems possible, the human factors issue (that is, the man-machine interface) remains an integral component in their development. If these systems are to be successful (meaning a sufficient number of people making use of them) they must be made compatible with human cognitive processing and responding. It is quite reasonable to assume that if a user encounters problems in the functioning of a particular system, he/she will not be inclined to use the system. For this reason it is essential to undertake research to investigate potential user-oriented problems and correct them before the products go on the market.

The present experiment was designed to examine several man-machine issues of using a Telidon database. A fundamental question to ask is "How easy is Telidon to use?" One can begin to answer such a question by having people search through a Telidon database to locate requested items of information. This procedure was chosen to simulate actual Telidon usage as closely as possible.

As well as determining how easy Telidon is to use, a search experiment of this kind enables one to examine the hierarchical tree structure which Telidon employs in some data bases. Is this the best type of structure for organizing information? What kinds of problems do people encounter when searching through information bases? What changes can be made to improve or eliminate problem areas in the tree?

These questions, and others, have been investigated in earlier studies with other hierarchically organized information bases. Van Ness & Tromp (1979) conducted an exploratory investigation using a videotex system (Prestel) of limited size to gather some information on the human factors of videotex. Their participants, who had no previous experience with videotex, performed 7 search tasks on a database containing 75 index pages and 900 information (document) pages. An experimenter recorded the time taken for each search task and occasionally he assured the user that the requested information was indeed stored in the database.

The results from this study indicated that, averaged over all tasks, people required about twice the minimum number of pages necessary to find the information. The averaged search time per task (extrapolated from a figure in their article) was approximately 2.7 minutes (162 seconds). Search time ranged from a low of about 1.5 minutes (90 seconds), to a high of about 16 minutes (960 seconds). From these results van Ness & Tromp cast some doubt on the contention that viewdata can be used easily.

A study by Lee & Latrémouille (1979) had participants choosing items from successive menus (index pages) presented on 13 x 18 cm cards to locate requested information. The purpose of the study was to evaluate the hierarchical tree structure and database which Telidon employs in some databases and to estimate how well people will search on an actual Telidon system. The users were told in advance that the information for which they would be searching would be contained in the database. They also received immediate feedback about whether they were correct or not after each choice. Subjective statements from the users were collected in order to determine the causes of errors when they occurred.

The results indicated that there were problems in the tree and in the way people search for information. For example, the probability of a user making an error on any given selection was 0.14. It was also revealed that 53% of the errors were made on the first two levels of the tree organization, and each person made mistakes on approximately half of the problems attempted. The conclusion expressed in that study suggested that people, in general, are going to have difficulties in finding information in hierarchical databases similar to the one tested.

An extension of the above study was conducted by Whalen & Latrêmouille (1980). Instead of information being organized on cards, a large, time-shared computer was used to simulate a Telidon system and database. Some procedural differences were incorporated by Whalen & Latrêmouille to make the situation more like an actual Telidon system. Unlike both of the previously mentioned studies, the users in this study were told that some of the items that they were requested to find would not exist in the database. When people use an actual Telidon system, it will not be known if the information desired exists in the database or not. Therefore, in their simulation the users were permitted to discontinue a search and select a new problem at any time. When people make choices on an actual Telidon system, they will not be given immediate feedback as to whether their choices are the correct ones leading to the desired information. Thus, in their simulation the users were not told whether their choices were correct or incorrect.

The results of the Whalen & Latrêmouille study again pointed to difficulties in the tree organization as well as user performance. For example, the probability of a user making an error on any given selection was 0.15. Approximately 80% of the total number of errors occurred on four pages within the first two levels of the tree. The conclusions reached are similar to those of Lee & Latrêmouille: people using Telidon are very likely to make errors at some point in their search for information in a tree-like arrangement of index pages.

The significance of the present study is that it is the first search experiment using an actual Telidon system. The procedural shortcomings of the van Ness & Tromp and the Lee & Latrêmouille studies were overcome by informing the users that some of the requested information might not be stored in the database and by not providing experimenter feedback on the choices made. The two main dependent measures were search time and number of errors made. The present study is also the first to measure search time on a Telidon system. The purpose of the present experiment was to evaluate the existing database of an actual interactive (videotex) Telidon system (the demonstration database at the Department of Communications, November, 1979). A major objective was to determine the average amount of time that people require to find information on an actual Telidon system (search time). Of interest are whether the type and number of errors made by the users on an actual Telidon system will be the same as the type and number of errors made by users in the previous studies. The experiment was designed to identify problem areas in the tree so that suggestions could be made for improving the database. A secondary purpose was to obtain verbal descriptions of the methods users employed to locate the information. A final reason for the present study was to stimulate interest and further research into this new area.

METHOD

SUBJECTS

The participants employed for this task were 24 volunteers from the Federal Government Public Service Commission. Of the 24 participants, 14 were male and 10 were female. The average age was 32.7 years for the male users and 31.6 years for the female users. All the users were naive with respect to both Telidon and the use of electronic information retrieval systems.

APPARATUS

The experiment took place in a room designed to simulate a "living room" atmosphere. The users sat on a couch facing the television screen approximately 2 meters away. A telephone line and a modem were employed as the direct link to the database computer. A Norpak Mark I Telidon terminal received data from and sent page requests to the computer via the modem. The received data were displayed on a modified 47.5 cm Electrohome colour television. A Telidon keypad (about the size of a pocket calculator) provided a means of communication between the user and the system. Search time was recorded on a Micronter LCD stopwatch, while a tape recorder was used to record the responses of some of the users.

PROCEDURE

Two instructional conditions were used in this experiment: a verbal condition and a non-verbal condition. Twelve users were randomly assigned to each condition. In both conditions each user was introduced to Telidon through an explanation of the types of display pages and the kinds of information available, as well as being informed of the search task (See Appendix A). In the verbal condition the users were asked to verbalize their search process by giving their reasons or strategies for the choices made while they were performing the search. A tape recorder was used to record the users' responses.

The verbal condition was employed to obtain verbal protocols which are intended to be used in the construction of a model of cognitive processing in the search process. This model will attempt to explain how people search for information and why they make mistakes. Such a model might prove useful for improving tree structures for retrieving information. (The verbalized data has not been analyzed as of the writing of this paper; therefore a report on the findings of this data will be included in a later paper).

Since all of the users were naive with respect to Telidon usage, each user was given a demonstration on how to use the Telidon keypad, and they were provided with two example search questions to ensure that they knew exactly what was required of them. The total time for instructions and demonstration was approximately 20 minutes.

In previous studies and in pilots to the present experiment, it was obvious that people were using Telidon in quite different ways. Sometimes people seemed to be searching for quite specific kinds of information. Other times they were searching for quite general kinds of information which might require looking in several different parts of the database. Browsing was another frequent activity. The present experiment attempted to employ tasks representative of the type likely to be employed by the typical user.

The Telidon database of November 1979 which was used in the present experiment contained approximately 1400 items of retrievable information. There were 900 document pages (pages containing information) and 500 index pages. From this pool, 16 items were randomly chosen. Questions were then created to correspond to the chosen items. These questions were intended to make up a representative sample of the kinds of searches a user might perform on Telidon. (See Appendix B for a list of the search questions). The questions were of three different types. Of the 16 questions, ten were of the "Specific" variety (that is, there was only one specific answer to the question contained in the database). Three questions were of the "General" form, meaning that there was more than one area in the tree where information could be found for the particular question. The remaining three were "No-Answer" questions for which there was no information in the database pertaining directly to the question. The users could determine precisely whether the information for the three no-answer questions existed by following a specific path. Thus, these questions were of the specific type except that the information was not contained in the database.

The rationale for including the no-answer questions was to simulate more precisely actual Telidon usage since not all of a user's requests will be answerable in a commercial Telidon system. The questions also allow one to investigate the differences (if any) in user behaviour between these types of questions and ones where the answer is contained in the information base.

The users were told that the answers to some of the questions might not be contained in the information base. They were instructed to inform the experimenter if they believed that an answer to a particular question was indeed not in the information base. The onus was on the user to decide whether the information was there. Thus the searches could end in either of two ways. If the user found the correct information, a successful termination of that particular search was recorded. If they incorrectly believed that the information was not in the database (excluding the no-answer questions), then an incorrect premature termination resulted.

The experimenter recorded the search time for each of the 16 questions. The search time commenced when the user began the search for a question and terminated when the correct page appeared on the screen. The search time could also be terminated when a user reported that the answer was not contained in the information base. The users were unaware that they were being timed so that they would not feel intimidated or try to rush their searches. The average length of the experimental sessions, including instructions and demonstrations, was 75 minutes.

At the end of the experiment the users were asked to fill out a short questionnaire about the experiment and Telidon usage in general. Thus some feedback information from actual Telidon users could be provided. Finally, the users were debriefed about the purpose of the experiment.

RESULTS

A split-plot factorial analysis of variance performed on the data for the users in the verbal and non-verbal treatment groups showed that there were no significant differences in terms of either search time or total number of errors per user between the two groups, F (1,22) less than 1 and F (1,22) = 1.43, p greater than .20, respectively. Therefore the data for both groups were collapsed resulting in one group of 24 users. (See Appendix C for analysis of variance tables.)

In terms of total search time, users spent approximately 160 seconds per question on the average. Broken down into question type, the 10 specific questions required an average of 160 seconds, the 3 general questions took 142 seconds on the average, with an average of 177 seconds needed for the 3 no-answer questions. The specific search time totals may have been inflated by the inclusion of specific questions in which the users gave up without finding the information. Search time for the successfully answered specific questions can be obtained by subtracting the unsuccessful specific search times from the specific search time totals. In doing so, the average search time for the successful specific questions becomes 137 seconds, with 205 seconds required on the average for the unsuccessful specific questions.

An analysis of variance on the four search time means (specific successful, general, no-answer, specific unsuccessful) revealed significant differences between the means, $F(3,23) = 6.29$, p less than .001. A Newman-Keuls test indicated that the unsuccessful specific questions required significantly more time than either the successful specific questions or the general questions.

It is possible that during answering of the 16 questions the users' performance could have improved with practice or through learning. Although both the average search time and the average number of errors made between the first three specific questions and the last three specific questions decreased, these differences were not significant.

Because of procedural and conceptual differences between previous experiments and the present experiment, different methods of defining errors have been developed. One method, employed in the present experiment, considered all of the users' choices as the basis for computing errors. However, procedural differences between the present and previous experiments precluded the comparison of previous error rates with an error rate that is based on all choices for the present experiment. Therefore, to be able to compare the present results with the earlier results it will be necessary to define errors in two additional ways.

The first method employed in the present experiment is somewhat different from methods of recording errors in previous studies. When a user made an "incorrect" choice, that is, one which would not directly lead him in the direction of the correct information, it was considered an error. Every successive choice made from an "incorrect" choice was also considered an error.

For example, should a user choose to investigate the "Business Guide" for information pertaining to the "Arts", an error would be recorded. Every successive choice within the business guide branch not in the direction of the arts category would be considered an error. Such choices can be considered errors because they take the user further and further away from the information requested. In such a situation, the correct choice is to select the previous index page (by pressing the retrace, or back arrow, key). Similarly, if the user chose to recall the previous index page when in fact his present index page actually contained a correct index term, then it was counted as an error. Such a choice was considered to be an error since it took the user further away from the target information.

By this first definition of errors, users in the experiment made a total of 1282 errors. The average number of errors per user was 53.42 for an average of 3.34 errors per problem per user. The 24 users retrieved and reviewed a total of 3654 pages. Thus, the probability that a user would make an error of this type on any given page was 0.351. In other words users had a probability of only 0.65 of making the right choice on any given presentation of an index page.

Errors can be defined in a second way. It can be argued that the first method of defining errors identifies as errors certain choices made by users which are not really errors. For example, the first method defines as an error any choice that takes the user further away from the target information. Thus, once the user is on a path that does not lead directly to the target information choosing any of the written alternatives on the menu page is defined to be an error (the correct choice by this definition is to retrace one's path). The second method differs by stipulating that for an error to be counted there must have been a correct written choice on each menu page which the user did not choose. According to this method of defining errors, there are neither correct nor incorrect choices once the user is on the wrong path. This method is analogous to that employed by Whalen & Latrémouille (1980).

To calculate errors by this second method, the shortest path (that is, the shortest sequence of pages) to the target information was identified for each "specific" problem. All pages not on the shortest path were excluded from further analysis. Considering only pages on the shortest possible path, a subject could make either a correct choice or an error. A correct choice was defined as a choice which retrieved the next successive page on the shortest path to the target information (such a page would necessarily be on the next level of the tree). An error was defined as any other choice, that is, an error was any choice which either retrieved a page not on the shortest path or retrieved a page on the shortest path that was further away from the target information than the current page (by pressing the retrace key).

By this second definition of errors, users made a total of 443 errors. The average number of errors per user was 18.46 for an average of 1.85 errors per specific problem per user. The 24 users retrieved and reviewed a total of 1312 pages on the shortest possible path. Thus, the probability that a user would make an error of this type on any given page (on the shortest path) was .338.

A third method of defining errors was employed in a previous experiment by Lee & Latrémouille (1979). This method was a variation on the second method just described. The users in the Lee & Latrémouille study were told whether their choices were correct or incorrect. Therefore, once they had made a correct choice on a given page, they could not go back up to the previous level. They also could not choose the same incorrect page more than once. In an attempt to present the results as though this method had been used, choices in which users retrieved the previous index page after a correct choice were eliminated from the error analysis in the present study. Also, when a user accessed an incorrect page more than once it was also eliminated from analysis. This third method of computing errors is comparable to the methods employed by Lee & Latrémouille, making it possible to compare error rates across experiments. The present method, however, is not identical to that employed in the earlier experiment. For example, users in the present experiment received no immediate feedback (being told correct or incorrect) whereas they did receive this type of feedback in the earlier experiment.

By this third definition of errors, users made a total of 275 errors. The average number of errors per user was 11.46 for an average of 1.15 errors per specific problem per user. The 24 users retrieved and reviewed a total of 1063 pages on the shortest possible path. Thus, the probability that a user would make an error of this type on any given page (on the shortest path) was .258.

The ratio of the number of pages actually retrieved to the minimum number of pages needed to locate the information for the 10 specific questions was 2.00:1. Therefore, the users required twice the minimum number of pages to find the information. Broken down into question type, an average of 9.53 pages were accessed for the specific questions. For the general and no-answer questions, 7.82 and 11.58 pages were chosen, respectively. The average minimum number of pages required for the specific questions was 4.7.

On 72% of the searches (excluding the no-answer questions) the users successfully terminated their searches. Conversely, 28% of the searches resulted in incorrect premature terminations. No question was solved perfectly (that is, with a minimum number of pages retrieved and no errors) by all the users.

In terms of the percentage of all errors that occur on a given index page, the single worst page was the "first", or top, index page in the tree on which 22% of the 1282 errors were committed. Over 47% of the errors occurred on the first two levels of the tree. This measure of the quality of a given page is particularly useful in deciding where the greatest effort should be expended to improve the tree index structure. Thus, for example, since the greatest percentage of errors were made on the first index page, the greatest potential for improving the general performance of users lies in improving the first page. However, the first page must also be retrieved by users far more often than any other page in the tree. This is because the retrieval of any information in the database requires the retrieval of the first page. Since the first page must be retrieved very frequently, there is much greater opportunity for users to make errors than with, say, a page at a lower level of the tree which is accessed relatively infrequently. Therefore, such a measure does not reflect the difficulty of one page relative to another. This is better measured by the percentage of all presentations of a given page that result in errors. Thus, the first page was accessed 730 times by the users and a choice was made on each presentation. Users made 278 errors on 730 choices for an error rate of 38%. One page on the second level of the tree was presented 67 times, and users made a total of 47 errors for an error rate of over 70%. By this measure this particular page on the second level is much worse than the first page. This measure is useful in identifying the most poorly and the best designed index pages. (See Appendix D for a complete listing of all the statistics.)

DISCUSSION

The results of the present experiment are in many ways similar to those of Lee & Latrémouille (1979), Whalen & Latrémouille (1980), and van Ness & Tromp (1979). For example, in each experiment it was evident that users experienced problems in locating the requested information. Most errors occurred in the first two levels of the tree structure (Lee & Latrémouille, 1979; Whalen & Latrémouille, 1980). A 2 to 1 ratio between the actual and minimum number of pages accessed in retrieving information and a comparable search time in the order of 2-3 minutes were found by van Ness & Tromp.

In answer to the question, "Is Telidon Easy to Use?", the collected data suggests that in fact the Telidon demonstration database, in its initial form, resulted in some problems for its users. Measures of user difficulty include the number of pages retrieved, the percentage of incorrect premature terminations, the number of problems solved without any errors, and error rate. That people were required to retrieve twice the minimum number of pages to find the target information could be looked upon as a general lack of efficiency in the search process. On the average, users prematurely terminated their searches 28 percent of the time when the answer was in the database. None of the questions were solved without any errors by all of the users. The three methods of calculating error rates produced error probabilites of 0.35, 0.34, and 0.26. The error rates are greater than those found by Lee & Latrémouille and by Whalen & Latrémouille and indicate a measure of difficulty for the user in searching for information. The discrepancies in error rates exist even though attempts were made to compare the errors in a similar way to the previous experiments. Several reasons can be hypothesized to account for these differences. One reason might be that the present study was conducted on an actual Telidon system whereas the others were simulations of Telidon. The differences might also be attributable to the fact that different databases were employed in each experiment as well as different search questions. A final reason may be that, in the

Lee & Latrémouille study, feedback was given to the users on the correctness of their choices.

Almost half of all errors occurred in the first two levels of the tree. This result is similar to both the Lee & Latrémouille and Whalen & Latrémouille studies. Since the pages on the first two levels must be accessed much more frequently than pages at the lower end of the tree (because to retrieve any piece of information one must retrieve the first page, etc.), one would expect a larger number of errors to occur there.

The category names used to identify different subject headings may have been responsible for some confusion among the users. Comments from the post-experimental questionnaire revealed a general concern that the subject headings were either too broad (for example, "General Interest Guide") or too ambiguous (for example, "Canadian Living"). Comments from users in the Lee & Latrémouille study revealed that the first page had titles that were too general. At the second level they indicated that there was some overlap between categories and some terms were judged be ambiguous.

The problems with the database tested in the present experiment indicated that some effort was necessary to improve the database. One way in which a Telidon demonstration system was improved was by removing ambiguous subject headings and replacing them with more lucid ones. If users become confused by ambiguous subject headings in the first or second level, it may compound the error rates by misguiding them to incorrect branches. Since the majority of errors occurred in the first two levels of the tree, the changes were focused here. A recent experiment by Latrémouille & Lee (1981) showed that the first page could be considerably improved by testing empirically several alternative "first" pages. This same experiment also showed that the use of descriptors (brief descriptions of the content of each index term) greatly facilitated performance and increased user preference.

Improvements may also result from empirically testing alternate retrieval methods. Among the different retrieval systems are keyword, keyword-tree hybrids, trees with cross-reference, and directory-tree hybrids. These alternative systems will be discussed in future reports.

The present data showed that experience using Telidon decreased search time and the number of errors made between the first and the last three specific questions, although these results were not significant. It is possible that more experienced users would have performed the task more efficiently, but further investigation is required to answer this question definitively.

If there is a difference between "beginner" and "experienced" user performance, a different mode of access to the information may be required. For if the retrieval method is too difficult, many first-time users may reject the system because of poor performance. On the other hand, if the retrieval system is over-simplified, many experienced users will find it too inefficient for their needs. What is needed is either a compromise between the two extremes or perhaps one method for novices and another method for experienced users. For example, a combination keyword-tree retrieval system which provides a tree index for naive users and a keyword system for experienced users may satisfy the needs of both types of users. Such combination retrieval systems are currently under investigation.

It is quite clear that more research is required along these lines. For example, a 35 percent error probability may or may not be too high for people to tolerate. What is the tolerance level of users? Moreover, performance on tree-structured retrieval systems must be compared with performance on alternative retrieval systems. The only way to answer questions like these is to test them empirically. As was mentioned earlier, the implementation of these systems on a large scale is not far off in the future. Therefore our efforts should be concerned with empirical tests of the major variables involved. As well, more effort should be expended on investigating methods for optimizing the information retrieval process. For example, the effect of adding cross references and descriptors to the tree index should be explored systematically.

REFERENCES

Latrémouille, S., & Lee, E. The Design of Videotex Tree Indexes. Telidon Behavioral Research 2. Department of Communications, Ottawa, May 1981.

Lee, E., & Latrémouille, S. Evaluation of tree structured organization of information on Telidon. (Technical Memorandum No. BRG 79-12). Department of Communications, Ottawa. December, 1979. Also in Telidon Behavioural Research I. Department of Communications, Ottawa, 1980, pp.231-242.

van Ness, F.L., & Tromp, J.H. Is viewdata easy to use? Institute For Perception Research. Eindhoven, Netherlands: IPO Annual Progress Report, No. 14, 1979.

Whalen, T., & Latrémouille, S. The effectiveness of a tree-structured index when the existence of information is uncertain. (Technical Memorandum No.BRIC 80-3). Department of Communications, Ottawa. September, 1980.

APPENDIX A

INSTRUCTIONS

During this experiment you will be introduced to Telidon. Telidon is an interactive visual communication system which permits public access to computer-based information sources. Pages of information from the information base can be selected by commands issued by the user via a keypad. The page is then delivered on the screen of a televison receiver.

A wide variety of information now exists in the Telidon information base. Examples of the types of information found include lists of restaurants in Ottawa, business news, real estate listings, etc. The information base is organized in a hierarchical tree structure which resembles an upside-down tree. This tree structure branches downward from a single root at the top to form a descending hierarchy of levels, each containing nodes that branch to another level.

There are two kinds of display pages. Index pages divide a category into subcategories to tell the user making a choice at each level of the tree where to go looking for the next appropriate page. A document page contains the kind of information that is of direct interest to the user. Selecting a particular item from an index page at a node brings the user to a node at the next level down. As one moves down the tree, the information becomes more and more specific until the desired document is located.

Your task will be to locate specified documents from the information base. We will ask you 16 questions which will inform you of the information that is to be located.

In the majority of instances the answer to the question will be contained in the information base. But in some cases there may be no answer to the given question. Your search will be concluded when you actually locate the desired information or when you believe the information is not contained in the information base.

As you search through the information base you will notice on certain index pages that one or several of the alternatives will be blocked off. This means that you cannot access any information from that particular alternative. The reason for this is that there are currently no pages of information contained under that alternative.

For your task some of the answers to the search questions may be contained in a blocked alternative. If this is the case and you believe the final answer to be included in that alternative, then you may "choose" that alternative by calling out its name or number. Should you choose a blocked alternative which does not contain the right answer, you will be told that the information that would be under the alternative is not correct for the given question and that you may continue your search.

While you are searching for the desired information we would like you to verbalize your search process. By this we mean that we want you to tell us what you are thinking of as you perform the search. For example, we would like to know why you choose one alternative over another, or why you have decided to search a particular part of the tree. In other words we would like you to think out loud.

We would also appreciate it if you would tell us your reasons or strategies before you actually press the buttons on the keypad. Please make every effort to give a reason for your decisions, even if you feel that is is trivial or unimportant. We are interested in every reason. As you are performing this task, we will tape record your responses.

You will probably make some mistakes while you are searching through the information base. There are no penalities for making an error. In fact, the mistakes may not neccessarily reflect on you but rather on the design of the system. So don't be nervous or get upset if you happen to make an error.

Before the experiment begins we will instruct you in the operation of the Telidon keypad and we will provide you with an example of what you are to do. If you have any questions, please feel free to ask.

APPENDIX B

Search Questions

1. What are the major land divisions of Switzerland?

2. Your car needs some work done on it. Find out some information on car repairs.

3. Where can you find out about purchasing some property outside of Ottawa?

4. How many radio stations are there in Ottawa?

5. You are interested in purchasing a refrigerator. What will you need to know before you buy one?

6. Someone has told you about an Italian restaurant called Calabria. What is the address of this restaurant?

7. You have just moved to Ottawa. You need to know what schools you can send your children to this fall.

8. What is the geography of Newfoundland like?

9. What does the Winnipeg provincial electoral division of Riel look like?

10. What does your horoscope say for today?

11. What kind of information is available on the Arts?

12. After winning the Wintario lottery you decide to find out about investing some of your winnings.

13. What's happening in sports on Telidon?

14. How do you play the Math trigonometry game?

15. What portfolio does Robert Smith hold in the Saskatchewan Cabinet?

16. You would like to see a list of the 1980 cars for sale.

APPENDIX C

ANALYSIS OF VARIANCE TABLES

SPLIT-PLOT ANALYSIS OF VARIANCE TABLE
VARIABLE: SEARCH TIME

Source	SS	df	MS	F	SIG
1. BETWEEN SUBJECTS	1060755.13	23	48459.13		
2. VERBALIZATION	255.13	1	255.13	.01	.943
3. ERROR	1060500.00	22	48204.00		
4. WITHIN SUBJECTS	5912310.00	360	94851.00		
5. QUESTION	967460.00	15	64498.00	4.52	.001
6. VERBALIZATION X QUESTION	233650.00	15	15577.00	1.09	.363
7. ERROR	4711200.00	330	14276.00		
8. TOTAL	6973000.00	383			

VARIABLE: NUMBER OF ERRORS

Source	SS	df	MS	F	SIG
1. BETWEEN SUBJECTS	1062.11	23	110.33		
2. VERBALIZATION	65.01	1	65.01	1.43	.244
3. ERROR	997.10	22	42.32		
4. WITHIN SUBJECTS	9137.31	360	194.58		
5. QUESTION	2267.70	15	151.18	7.66	.001
6. VERBALIZATION X QUESTION	354.91	15	23.66	1.20	.271
7. ERROR	6514.70	330	19.74		
8. TOTAL	10199.00	383			

RANDOMIZED BLOCK ANALYSIS OF VARIANCE TABLE
VARIABLE: SEARCH TIME

Source	SS	df	MS	F	SIG
1. BETWEEN TREATMENTS[1]	98047.00	3	32682.00	6.29	.001
2. BETWEEN SUBJECTS	408710.00	23	17770.00		
3. ERROR	358320.00	69	5193.10		
4. TOTAL	865080.00	95			

[1]Type of question: Specific successful, Specific unsuccessful, General, No-Answer.

APPENDIX D

SEARCH EXPERIMENT RESULTS

MEAN SEARCH TIME

n=16 QUESTIONS = 159.99 s
n=10 QUESTIONS(SPECIFIC) = 160.42 s
n= 3 QUESTIONS(GENERAL) = 141.77 s
n= 3 QUSETIONS(NO-ANSWER) = 176.81 s

SP (successful) = 137.30 s
SP (unsuccessful) = 204.93 s

ERROR METHOD 1

TOTAL ERRORS = 1282
AVERAGE NO. ERRORS PER USER = 1282/24 = 53.42
AVERAGE NO. ERRORS PER USER PER QUESTION = 3.34
TOTAL NO. PAGES RETRIEVED = 3654
PROBABILITY OF AN ERROR = 1282/3654 = .35 (.3508)

ERROR METHOD 2

TOTAL ERRORS = 443
AVERAGE NO. ERRORS PER USER = 443/24 = 18.46
AVERAGE NO. ERRORS PER USER PER SPECIFIC QUESTION = 1.85
TOTAL NO. PAGES RETRIEVED = 1321
PROBABILITY OF AN ERROR = 443/1321 = .34 (.3377)

ERROR METHOD 3

TOTAL ERRORS = 275
AVERAGE NO. ERRORS PER USER = 275/24 = 11.46
AVERAGE NO. ERRORS PER USER PER SPECIFIC QUESTION = 1.15
TOTAL NO. PAGES RETRIEVED = 1063
PROBABILITY OF AN ERROR = 275/1063 = .26 (.2587)

TOTAL SPECIFIC CHOICES = 2277
TOTAL SPECIFIC QUESTIONS = 239
2277/239 = 9.53 PAGES PER SPECIFIC QUESTION
TOTAL GENERAL CHOICES = 555
TOTAL GENERAL QUESTIONS = 71
555/71 = 7.82 PAGES PER GENERAL QUESTION
TOTAL NO-ANSWER CHOICES = 822
TOTAL NO-ANSWER QUESTIONS = 71
822/71 = 11.58 PAGES PER NO-ANSWER QUESTION

RATIO NO. PAGES ACTUALLY RETRIEVED TO MINIMUM NO. PAGES
RETRIEVED (n = 10 SPECIFIC QUESTIONS)
TOTAL SPECIFIC CHOICES = 2277
TOTAL MINIMUM NO. SPECIFIC PAGES =1138
2277/1138 = 2.00

AVERAGE NO. SEARCHES TERMINATED SUCCESSFULLY (n = 13 N-A QUESTIONS EXCLUDED)
TOTAL NO. TERM SUCCESSFUL = 222
N = 24
222/24 = 9.25 SUCCESSFUL TERMINATIONS PER SUBJECT

PERCENTAGE OF SEARCH TERMINATED SUCCESSFULLY (n = 13 N-A QUESTIONS EXCLUDED)
TOTAL NO. TERMINATED SUCCESSFUL = 222
TOTAL SPECIFIC+GENERAL QUESTIONS = 310
222/310 = 72% (.7161)

AVERAGE NO. SEARCHES TERMINATED PREMATURELY (n = 13 N-A QUESTIONS EXCLUDED)
TOTAL NO. INCORRECT PREMATURE TERMINATION (I.P.T.) = 88
N = 24
88/24 = 3.67 INCORRECT PREMATURE TERMINATION (I.P.T.) PER SUBJECT

PERCENTAGE OF SEARCHES TERMINATED PREMATURELY (n = 13 N-A QUESTIONS EXCLUDED)
TOTAL NO. INCORRECT PREMATURE TERMINATION (I.P.T.) = 88
TOTAL SPECIFIC+GENERAL QUESTIONS = 310
88/310 = 28% (.2839)

AVERAGE NO. OF ERRORS PER QUESTION

QUESTION	MEAN ERROR 1	MEAN ERROR 2	MEAN ERROR 3
1	1.417	0.500	0.458
2	4.292	1.833	0.875
3	7.083	1.958	1.333
4	2.458	1.292	0.583
5	9.917	2.583	1.667
6	3.917	2.917	1.917
7	1.250	0.792	0.750
8	4.458	3.167	1.417
9	5.052	2.333	1.625
10	2.875	1.083	0.958
11	0.500		
12	1.042		
13	1.625		
14	1.708		
15	4.542		
16	1.792		

NO. QUESTIONS SOLVED PERFECTLY BY ALL SUBJECTS = 0.00

NO. QUESTIONS SOLVED PERFECTLY = 63

PERCENTAGE OF QUESTIONS SOLVED PERFECTLY = 63/239 = 26% (.2636)

PERCENTAGE OF QUESTIONS HAVING ONE OR MORE ERRORS = 59% (.5936)

PERCENTAGE OF ERRORS ON A GIVEN PAGE:

LEVEL	PAGE	% ERROR	
1	FRONT PAGE	22	278/1282 = .2168
2	1.0-9.0	26	330/1282 = .2574
2	1.0	13	164/1282 = .1279
2	2.0	07	86/1282 = .0671
2	3.0	04	47/1282 = .0367
2	4.0	01	10/1282 = .0078
2	5.0	0.5	7/1282 = .0055
2	6.0	0.2	3/1282 = .0023
2	7.0	0.3	4/1282 = .0031
2	8.0	0.2	2/1282 = .0016
2	9.0	0.5	7/1282 = .0055

LEVEL 1 + 2 (FRONT PAGE + 1.0-9.0) = .47 608/1282 = .4743

PERCENTAGE OF THE CHOICES MADE ON A GIVEN PAGE THAT ARE ERRORS:

LEVEL	PAGE	% ERROR	
1	FRONT PAGE	38	278/730 = .3808
2	1.0-9.0	38	330/879 = .3754
2	1.0	34	146/476 = .3445
2	2.0	61	86/142 = .6056
2	3.0	70	47/67 = .7015
2	4.0	33	10/30 = .3333
2	5.0	09	7/79 = .0886
2	6.0	30	3/10 = .3000
2	7.0	19	4/21 = .1905
2	8.0	05	2/39 = .0513
2	9.0	47	7/15 = .4667

LEVEL 1 + 2 (FRONT PAGE + 1.0-9.0) = .38 608/1609 =.3779

CHAPTER 4

THE DESIGN OF VIDEOTEX TREE INDEXES: THE USE OF DESCRIPTORS AND THE ENHANCEMENT OF SINGLE INDEX PAGES

Susane Latrémouille and Eric Lee

ABSTRACT

At the present time, most videotex systems employ some variation of a tree structure or menu-selection approach in the retrieval of information. The field of information science provides no guidelines for constructing tree-like indexes. Without such guidelines, information providers have been forced to construct tree indexes primarily on the basis of intuition. The purpose of the present study is to develop such recommendations and guidelines. The objectives of the two experiments described herein were: (a) to demonstrate the application of an alternative method for modifying tree indexes called the single-page method; (b) to evaluate the effect of adding descriptors (brief descriptions of each index term) on user preferences and performance; (c) to determine whether there is any degree of consistency among experts or among naive users in their preferences for alternative "first" pages; (d) to determine the relationship between user performance and user preferences for the "first" pages; (e) to determine the best first page from among 12 alternative first pages for the Department of Communications' experimental Telidon database.

In the first experiment 10 videotex experts rank ordered 12 different "first" or top index pages for an experimental database. Six of the 12 index pages were generated by a number of experts working independently (not the same experts that participated in the experiment). The other pages included the original index page, the page produced in an earlier experiment by modifying the original index page, and the page produced by a professional indexer. There were two versions of each of the last three pages: one with and one without descriptors. There was virtually no agreement among the experts in preference. Each expert had a completely different view of what was best. None of the 12 pages was reliably judged better than any of the others.

In the second experiment, 20 videotex-naive users performed two tasks: a search task followed by a preference ranking of the same 12

index pages employed in Experiment 1. In the search task, each naive user was required to search for the information requested in 10 questions for each of the 12 index pages (for a total of 120 questions). The user responded to each question by choosing an index term on the index page displayed. Naive users were consistent both in their preference rankings of the 12 pages and in the number of selection errors made on each page. Some pages were reliably superior to the others. Pages with descriptors had fewer errors and were preferred over pages without descriptors. The addition of descriptors to an index page reduced errors by up to half.

RECOMMENDATIONS AND CONCLUSIONS

1. Testing a tree-like index can be used to improve significantly the ease of retrieval of information by naive users. Testing the top index pages is recommended if resources are scarce.

2. Descriptors significantly reduce user errors and enhance user preferences.

3. Since there was virtually no agreement among experts on what constituted a good index page, index pages should be tested on a sample of the target population (in this case, naive users).

4. Preference judgments and the number of user errors are two measures of the quality of tree index pages. For naive users, these two measures are highly correlated when preference judgments are elicited after performance. Naive users were highly consistent both in their performances and in their preference judgments. Furthermore, there were large and reliable differences both in the number of errors and in performance among the alternative index pages. Further research is required to determine whether preference judgements alone can be used as the sole measure of index-page quality for improving tree indexes.

5. Avoid the use of over-general, ambiguous index terms such as "General Interest" and "Miscellaneous".

6. Avoid the use of fewer than 4 index terms on a page.

7. Attempt to use index terms that are approximately equivalent in terms of amount of information classified under each term on a page.

8. Use some form of intra-page organization (eg., alphabetical or frequency).

INTRODUCTION

At the present time most videotex systems employ some variation of a tree structure or menu-selection approach in the retrieval of information. Contemporary information science offers little assistance to the designer of tree-structured indexes. Traditional information science has focused on developing keyword retrieval systems for books and articles, not general information of interest to the public which is not in books. Moreover, such keyword retrieval systems have usually been designed for the librarian specialist and not for the untrained or the general public.

Without general guidelines from information science, information providers and others have been forced to construct tree-like indexes primarily on the basis of intuition. Recent experiments (Lee & Latrémouille, 1979; Whalen & Latrémouille, 1980; McEwen, 1981; van Ness & Tromp, 1979) have shown that members of the general public can experience considerable difficulty in retrieving information of interest on such first-generation tree indexes. Methods and techniques for improving such first-generation tree indexes must be developed to realize the potential of tree-like indexes for information retrieval on videotex.

The purpose of the series of experiments to be described herein was to investigate methods and procedures for enhancing the tree structure from the user's point of view. This research was directed primarily towards the objective of developing recommendations for information providers and others who will be constructing tree indexes for information databases. These recommendations should enable the indexer to develop a more efficient, effective, and satisfactory retrieval system from the user's perspective.

Previous investigations of tree structures (Lee & Latrémouille, 1979; Whalen & Latrémouille, 1979, van Ness & Tromp, 1979; McEwen, 1981) have shown that most errors (50% or more) committed by naïve users when searching for information using a tree index are made at the two highest levels of the tree. In fact, the number of errors per page is greatest the highest level of the tree and decreases as one goes down the tree. That is, most errors are made on the top page, fewer on pages at the second level of the tree, fewer at the third level, and so on. In part, this can be attributed to the greater frequency with which pages higher in the tree must be accessed in retrieving information; the first page must be accessed at the beginning of every session whereas a page at the second level is accessed only on a small fraction of the sessions (specifically those sessions in which the information requested is actually categorized under one of the alternative choices on the given second-level page). If resources for improving a given tree index are scarce, it can be argued that such resources should be concentrated where they will have the biggest impact in terms of improving the tree from the user's perspective. Index pages yielding the most errors have the greatest potential for improvement. It follows, therefore, that resources should be directed towards improving the index page with the most errors first, then the index page with the second highest number of errors, and so on. Since most errors occur on the top index page, this is the page that should be improved first. Thus, the strategy to be employed in the present experiments is to concentrate entirely upon improving a single index page rather than attempting to improve the entire tree as was the case in Lee & Latrémouille (1979). This method of attempting to improve the tree structure will be referred to as the single-page method in the present report.

A major cause of user errors in searching through a tree index is the ambiguity of the index terms themselves (Lee & Latrémouille, 1979). Index terms on the same index page often overlap considerably in the kinds of information that people expect to find categorized under each term. Similarly, some kinds of information do not seem to fall under any of the available index terms on the present index page, but the information definitely falls under the general category label, or title page, which subsumes all the index terms on the page. In either case, the linguistic ambiguity causes people to make errors because it is not always obvious from a consideration of the category name (index term) alone what information is actually subsumed under a given category name. Ambiguity in the interpretation of index terms could be reduced by providing a complete description of all information subsumed under each category label (that is, index term). Such a complete description would require users to read virtually all the information in the entire database each time a new search for information is initiated. Complete descriptions are clearly impractical even though they could eliminate all ambiguity in the interpretation of index terms. It is possible, however, that brief descriptions of the information categorized under each index term could greatly facilitate the naive user in his search for information by reducing significantly the inherent ambiguity of the index terms. Two kinds of brief descriptions are possible: a short, abbreviated description of the contents of each category (that is, index term) or a list of the index terms which are subsumed under a given category label and which are found on the next lower level of the tree. Short descriptions of either type will be referred to as descriptors. There is, however, a price to pay for the inclusion of descriptors on each menu page. Even short descriptions after each alternative on a menu page can increase both the amount of time required to transmit the information from the central computer and display it on the user's terminal and the time required by users to read this extra information on each menu page. The addition of descriptors to index pages might also increase the costs of both transmission and data entry. Such costs, however, do not affect the user directly and are unlikely to be very large relative to the costs due to failed searches.

Several different predictions of the effects of descriptors are possible. One possible prediction is that as the amount of description added to the index terms increases, the ambiguity of the index terms is likely to decrease (thereby reducing the amount of time wasted searching incorrect paths to the information). At the same time, there could be a concomitant increase in reading time and display time required. From the user's perspective, therefore, there could be a tradeoff between shorter search times due to fewer user errors and longer search times due to longer reading and display times. The relationship between user search time (and user satisfaction as well) and amount of description could be described by a curvilinear function such as that illustrated in Figure 1. From the graph, it can be seen that search time should be minimized and user satisfaction maximized with some intermediate degree of description. Too much or too little description produces much longer search times and less user satisfaction.

On the other hand, descriptors may allow the user to stop reading sooner than on a comparable index page having no descriptors. Descriptors may facilitate rapid selection by the user. By clarifying the content of information subsumed under each index term, descriptors should enable the user to locate more quickly the specific alternative for which he is looking. Without descriptors, users may have to read all the alternatives on an index page and, even then, may require considerable time to choose among essentially ambiguous alternatives. Thus, descriptors could reduce both errors and search time. A third possibility is that users will reject the use of any amount of description, preferring the use of index terms alone in the tree structure. (However, descriptors can be displayed after all index terms. Thus, experienced users should be able to read the index terms before the descriptors are displayed.) Only empirical research can answer the question.

A major purpose of the present study was to determine the effect of short descriptions on user performance and user preference. A secondary objective was to assess the degree of consistency or agreement among naive users in their preferences for and performance on first pages as well as to assess consistency among "experts" in their preferences. A lack of agreement among naive users in their preferences or variable performances on the different first pages would be a strong argument that no tree could be designed that would meet the needs of most people. Conversely, high agreement among users in both preference and performance would argue strongly that (a) some index pages are much better than others, and (b) index pages can be designed that could meet the needs of most users.

Previous experiments (Lee & Latrémouille, 1979; Whalen & Latrémouille, 1980; McEwen, 1981) have indicated that people often choose the wrong alternative when using the present (Sept. 1980) first page for the Telidon experimental database. A practical issue is whether a better first page can be found for this particular database. Therefore, a practical objective of the present experiments was to find a new first page that would minimize errors. An improved first page would be particularly useful in Telidon demonstrations since the first page is almost always presented to the audience.

Another issue in information retrieval by menu selection is the relationship between users' preferences and their performance. If preferences and performance are highly related, then it may be possible to substitute preference judgments for performance as a measure of the quality of index pages. Since preference judgments are usually easier and less costly to collect than performance measurements, the process of testing and modifying tree indexes could be considerably simplified. Is user preference strictly an inverse function of the number of errors? What factors other than performance affect user perceptions of tree index pages? If user preferences are unrelated to performance, then how should tree structures be modified -- on the basis of preferences or on the basis of performance? The present study was designed to address these questions.

To summarize, the objectives of this study were: (a) to demonstrate the application of an alternative method for modifying tree indexes, the single-page method; (b) to evaluate the effect of adding descriptors (brief descriptions of each index term) on user preferences and performance; (c) to determine whether there is any degree of consistency among experts or among naive users in their preferences for alternative "first" pages; (d) to determine the relationship between user performance and user preferences for the "first" pages; (e) to identify the best possible first page from among 12 alternative first pages for the Department of Communications' experimental Telidon database.

Two experiments were designed and conducted to meet the objectives described above. On the basis of this empirical research, recommendations were derived for information providers and others who must design and construct tree indexes.

EXPERIMENT I

Since most errors have occurred on the top (or first) index page in the tree index (Lee & Latrémouille, 1979, McEwen, 1981), the present experiment was directed towards finding a more satisfactory first page for the experimental Telidon database. Modifying the "first" index page should, if successful, result in a considerable overall improvement in user performance and user satisfaction with the tree retrieval method.

Several alternative techniques were employed in generating 12 different versions of the "first" index page. The original "first" page was included as a baseline, both because any first page substituted for it would have to be better and because user performance data had already been collected on it (Lee & Latrémouille, 1979). For similar reasons, the first page derived by modifying the original first page was also included. This modified first page was obtained by identifying where typical members of the general public chose incorrect index terms in the original tree, and then having another group of similar people reclassify and relabel index terms wherever errors had been made previously in the original tree (Lee & Latrémouille, 1981). The rationale for using typical members of the general public to modify the tree was that the language and classifications employed by people from the general public should match the language and natural classifications of typical users more than would the language and classification schemes of an expert. Moreover, a tree based on the classifications, etc., of a sample of people should be more representative than one based on the classifications of a single individual. The assumption was that Telidon users would select fewer incorrect alternatives if the terminology and classifications employed in the tree structure were maximally similar to their own.

A third version of the first page was produced by a professional indexer (librarian). Three additional versions of the first page were produced by adding descriptors to each index term of the original, modified, and professional index pages. These six alternative first pages constituted the elements of a design embedded within the overall 12-page design. The embedded design consisted of two factors: type of first page (original, modified, or professional) and descriptors (presence or absence). This embedded design was specifically constructed to test the usefulness of placing descriptors on tree index pages. The remaining six first pages were produced by experts. The conceptual design and organization of each page differed considerably from page to page.

The purpose of the first experiment in this study was to examine the preferences of Telidon experts - people very familiar with Telidon - for the 12 different versions of the first index page just described. Each expert was required to rank the 12 versions of the first page from best to worst.

METHOD

SUBJECTS

Ten people who knew the Telidon experimental database quite well were asked to participate in this experiment. This sample consisted of 5 English-speaking people and 5 bilingual (French and English) people. All of these people worked with Telidon on a regular basis and, therefore, could all be considered experts in videotex usage.

STIMULI

The stimuli consisted of 12 "first" pages (see Appendix A), each of which was typed on a separate white file card (12.5 X 20.5 cm). Each of these 12 pages was designed to serve as the first index page in a tree for retrieving information from the same experimental database of information. For comparison purposes, the top index page from the original tree and the corresponding top page derived, through experimentation, by modifying the original top page were included as controls (see Lee & Latrémouille, 1979). A third top page was produced by a professional librarian in the field of indexing. None of these three pages included descriptors. (Descriptors are brief descriptions of the contents of index terms.) A corresponding set of three pages was constructed by adding descriptors to each page. An additional 6 top pages were constructed by experts working independently of one another. These "experts" were not professional indexers but people very familiar with the existing videotex tree structures. The experts included human factors specialists as well as computer database specialists. Experts creating pages Expert #1:ND, Expert #4:1D, and Expert #5:ND constructed their index pages using their own intuitive judgments together with user comments and the results from earlier studies of tree structures. The page Expert #2:ND was constructed by systematically fixing those parts of an earlier first page where users made many errors. The page was modified by reclassifying and renaming problem items. The page Expert #6:ND was derived in quite a different way from the others. Rather than employing index terms that described the contents of a category of information, this page employed index terms that specified actions or time.

PROCEDURE

The 12 first pages were randomly arranged before presentation to each person. People were instructed "to rank-order them (the 12 first pages) according to your preference. We would like to know which pages would be adequate to use with Telidon and which ones would not". To make the scaling task easier, participants were asked first to separate the 12 pages into two piles, the good pages and the poorer pages. Then they rank ordered the stimuli within each group from the best to the worst.

RESULTS AND DISCUSSION

Table 1 gives the average and the normalized ranks in terms of user preferences for each of the first pages. (See Appendix C for ranks for each individual.) There was no agreement among the subjects in their preferences as measured by Kendall's coefficient of concordance, $\underline{W} = 0.1682$, $\underline{X}^2$ (9) = 18.05, not significant (Siegel, 1956). Moreover, the average Spearman rank correlation between all possible pairs of users was only 0.08, indicating negligible agreement among subjects in their preference rankings.

A Friedman two-way analysis of variance by ranks (Kirk, 1968) indicated no significant differences among the first pages in terms of judged preferences, $\underline{X}^2$ (11) = 18.50, $\underline{p} = 0.07$. Thus, not even one of the "first" pages was significantly preferred over any of the other "first" pages. Separate Friedman analyses of variance on the rankings for the 5 English participants and the 5 bilingual participants were also nonsignificant.

Both the Friedman and Kendall analyses indicate that there was virtually no agreement among the experts on the perceived goodness of first pages. Either the alternative "first" pages are really not very different from one another, or experts have quite different perceptions of what a good "first" page should include. The next experiment provides information relevant to choosing between these two interpretations.

EXPERIMENT 2

The first experiment suggested that there was virtually no agreement at all among experts on what constitutes a good tree index page. Although the experts disagreed, it is plausible that naive users may, as a group, have quite similar perceptions of the tree. Two alternative interpretations of the results of Experiment 1, however, are that all people will perceive the tree in different ways or that the 12 pages tested are not sufficiently different from one another that people can differentiate between the pages reliably. (A cursory glance at the 12 index pages tested in Experiment 1 is almost sufficient to reject the latter interpretation.) Experiment 2 served to choose among these three competing interpretations.

The purpose of the second experiment in this study was to determine the preferences and the performance of Telidon-naive users for the same 12 versions of the first index page employed in Experiment 1. Users in the present experiment performed two tasks: first, each user had to search for information using each of the 12 index pages and, second, each user ranked the 12 pages in terms of preference. An objective of the experiment was to determine the relationship between performance and preference in tree indexes.

METHOD

SUBJECTS

Twenty civil servants who were unacquainted with Telidon were asked to participate in this study. Ten people were bilingual (English and French-speaking) and 10 spoke English only. The names of these people were drawn from a list of volunteers.

STIMULI

For the second task (preference judgments) of the present experiment, each stimulus card displayed the alternatives for a single "first" page. The stimuli were identical to those employed in Experiment 1 except for some very minor alterations. The alterations included (a) capitalizing all letters in each alternative while leaving all descriptors in lower case, (b) adding the words "for the General Interest Guide" to the end of the first alternative on the professional page, and (c) indenting descriptors six spaces.

For the first task (search and retrieval) of the present experiment, the first pages appeared on the video terminal screen with exactly the same format and wording as that employed on the stimulus cards for the first task.

TASK

All subjects performed two tasks: a search and retrieval task followed by a preference ranking task. In the first task, participants were asked to find specified information by selecting the correct alternative on each index page. The second task for all subjects was to rank order the 12 first pages in terms of their preference.

The database of information under study in the present experiment was the same one used in the study by Lee & Latrémouille (1979). To obtain a representative sample of the information contained in this Telidon database, 10 different documents were randomly selected from among the 900 documents contained in the database. The information contained in each document was used to phrase one question. Each question was worded carefully to avoid giving the participants inadvertent clues about the location of the information (see Appendix B for a list of the questions). In the present study, participants did not search through the entire tree. Instead, only the first page of each index was presented. They were only allowed to choose one alternative on a given page for each question. Ten questions were asked for each page for a total of 120 questions per participant.

Participants were not told whether they were right or wrong in their choices. Thus, this task most closely approximates usage by completely naive users; users who have no knowledge of videotex and who have received no feedback about the actual location of information. Feedback (being told right or wrong) would have allowed people to learn from their mistakes. Having learned through feedback that information was categorized under a different index term, a user might compensate for an inadequately designed tree and avoid the same type of error on future occassions. With no feedback, the experimental design is maximally sensitive to design deficiencies in the index pages. The experimental test between the pages would be considerably weakened if feedback permitted users to avoid choosing incorrect alternatives even though such alternatives appeared to be the correct ones to the user.

PROCEDURE

All subjects performed the search task first and then the preference ranking task. At the beginning of the session, participants were given a brief description of Telidon. Then the 12 index pages were presented one at a time on the display screen in a different random order for each user. Users were asked to find the information requested in the question appearing at the top of the screen by choosing the appropriate alternative on the index page displayed below. For each first page, 10 questions were presented to each user in a different random order. The questions are reproduced in Appendix B. Users responded to each question by typing in the number corresponding to the alternative selected on the index page. After a brief delay a new question was presented. No feedback was given to the user.

After completing the search task, users were required to rank order the 12 pages on preference. The procedure for the ranking task was essentially the same as that employed in Experiment 1.

RESULTS

While all participants in Experiment 2 were fluent in English, the mother tongue for half the participants was French and for the other half it was English. A split-plot factorial analysis of variance (Kirk, 1968) with repeated measures on the index page factor (12 levels) and independent groups for the mother tongue factor (2 levels - French and English) showed that no effect of either mother tongue (F (1, 18) = 2.62, NS) or the interaction of mother tongue by index page (F (11, 198) = 1.76, NS) on performance. Moreover, the Spearman rank-order correlation between the average preference ranking of the 12 pages by those whose mother tongue was English with those whose mother tongue was French was 0.92, p less than 0.01. Since about 85% of the variance was common between the French and English preference rankings, both French and English users were probably ranking the 12 pages in basically the same way. Since mother tongue had no appreciable effect on the responses of the participants, the data for all 20 subjects were combined for all further analyses.

PREFERENCE

Table 2 gives the average and the normalized preference ranks (Garner & Creelman, 1970) for each index page. (See Appendix C for preference rankings made by each individual.) There was a high degree of agreement among the users in their preference judgments as measured by Kendall's coefficient of concordance, W =0.5195, $X^2(19)$ = 114.29, p less than 0.001 (Siegel, 1956). The average Spearman rank order correlation between the preference rankings of different users was 0.49, indicating a relatively high degree of agreement and uniformity among naive users in their preferences for various index pages.

A Friedman two-way analysis of variance by ranks (Kirk, 1968) indicated significant differences among the index pages in terms of preference judgments, $\underline{X}^2(11) = 132.95$, $\underline{p}$ less than 0.001. Table 3 presents the results of a distribution-free multiple-comparison test based on the Friedman rank sums (Hollander & Wolfe, 1973). This analysis shows that the 6 highest-ranked pages did not differ significantly from one another, nor did the 5 lowest-ranked pages differ from one another. However, the top-ranked page (modified: with descriptors) was judged to be significantly better than the 6 lowest-ranked pages. (See Table 3 for details).

Six of the 12 index pages differed systematically from one another. There were two versions of each of the index pages labelled original, modified, and professional. One version of each page had descriptors after each index term whereas the second version was identical but without the descriptors. The sign test (Siegel, 1956) provides a statistical test of the impact on user preferences of adding descriptors to an index page. By the sign test, significantly more users preferred each index page with descriptors to its counterpart that did not have descriptors. For the original index page, all 20 users preferred the version with descriptors over the version without descriptors, $\underline{p}$ much less than 0.001. For the modified page, 16 of the 20 users preferred the version with descriptors, $\underline{p} = 0.006$. For the professional page, 16 of 20 users preferred the version with descriptors, $\underline{p} = 0.006$.

PERFORMANCE

Table 4 gives the mean number of errors made on each index page. There was a high degree of consistency among the users in the number of errors that they made on each page, Kendall's coefficient of concordance $\underline{W} = 0.5205$, $\underline{X}^2(19) = 114.50$, $\underline{p}$ less than 0.001. The average Spearman rank-order correlation for all possible pairs of subjects between the rankings of the number of errors per page was relatively high just as it was for user preferences, Spearman $\underline{r} = 0.50$.

Users made an average of 46.85 errors in total for a mean of 3.90 errors per page by each user. The maximum number of errors a user could make was 10, since each user made 10 choices on each page. Therefore, users made errors on 39% of their choices for a probability of 0.39 for making an error given that a user makes a choice on an index page.

The 12 index pages differed significantly in terms of the number of errors per page, $F(11,209) = 22.48$, p less than 0.001, see Table 4. The Geisser-Greenhouse conservative F test (Geisser & Greenhouse, 1958; Greenhouse & Geisser, 1959) with modified degrees of freedom of 1 and 19 was also highly significant, p less than 0.001. The Neuman-Keuls multiple-comparison test revealed that the mean number of errors for the 7 index pages with the fewest errors did not differ significantly from one another, but all 7 pages were significantly better than the 5 remaining index pages. The 2 index pages with the highest number of errors per page were significantly worse than the other 10 pages (see Table 5).

Embedded within the overall 12-page design was a 2 x 3 repeated measures analysis of variance design for the two repeated factors: type of index page and descriptors versus no descriptors. The descriptors-by-page interaction was significant, $F(2,95) = 7.10$, p less than 0.01 and Geisser-Greenhouse conservative $F(1,19) = 7.10$, p less than 0.05 (see table 6). Although the interaction was significant, it accounted for only 4.2% of the total variation. Moreover, it was also a disordinal interaction (see Figure 2).

Because the interaction was significant, a simple main effects analysis was performed on the data (see Table 7). From the table it can be seen that the addition of descriptors to each of the three index pages resulted in a much improved (fewer errors) index page. The degree of improvement varied considerably from page to page but was greatest for the original index page (see Appendix A). The addition of descriptors to an index page reduced errors by over half for the original index page and by almost half for the modified page. There was no systematic relationship between the amount of improvement and the quality (as measured by the number of errors) of an index page without descriptors. Thus, it is not possible to predict the amount of improvement that can be expected from adding descriptors to a page. However, the results of the present experiment provide strong support for the conclusion that the addition of descriptors to any index page can significantly reduce the number of user errors.

The sign test gave similar results to the analysis of variance of the 2 X 3 design. Users performed significantly better on a page with descriptors than on the corresponding page without descriptors. For both the original and the modified pages, 18 of the 20 users made fewer errors on pages with descriptors than on those without, $\underline{p}$ much less than 0.001 (one tie). For the professional librarian page, 13 of the 20 users made fewer errors on pages with descriptors, $\underline{p} = 0.048$ (two ties).

The average ranking of the 12 index pages in terms of the number of errors per page correlated very highly with the average preference ranking, Spearman $\underline{r} = 0.72$, $\underline{p}$ less than 0.01. For each subject, the Spearman correlation was computed between that individual's own preference ranking and the number of errors committed by him on each page; the average correlation across all 20 users was 0.38, which is highly significant with $\underline{p}$ less than 0.001. The average preference ranking by the experts in Experiment 1 correlated significantly with the average preference ranking by naive users (Spearman $\underline{r} = 0.53$, $\underline{p}$ less than 0.05) but not at all with performance, as measured by the number of errors, by naive users, Spearman $\underline{r} = 0.22$, NS.

In the post-experimental session, each person was asked to describe the criteria they used in making their preference judgments. Table 8 lists the frequency with which each criterion was used by the participants to rank order the pages. A total of 7 different criteria were reported by one or more people in the two experiments:

1. Amount of description: Participants commented on the fact that some pages had descriptions beside a title. Some participants commented that descriptors would slow down the search process because more reading would be involved. Other participants said that descriptors would decrease the number of errors in the search process and at the same time would decrease the number of times one had to go back to the beginning of the tree.

2. Generality of titles: People judged a page on the generality of its titles. A title was judged too general when it did not convey a concept or an idea. For example, "Miscellany" was such a title, since anything can be classified under miscellany.

3. Complexity of terminology: People commented on the choice of words used as index terms. Sometimes a title might represent a good category, but the choice of words could be too technical or too complex. An example would be "Field trial DATABASES".

4. Order of titles: The order in which items were presented on a page seemed to be important to some people. Two people from Experiment 2 reported that the order of the items on a page helped in the logic of finding information. One said that items at the top should be more general and items at the bottom should be more precise. This person added that it was easier to eliminate the more general categories first. Other participants said that the items accessed most frequently and the items that should be found quickly (such as "Emergency") should be placed at the top of the page.

5. The number of titles: the number of alternatives on a page seemed to be important for some participants. The reason profferred was that searches are more efficient when there are many alternatives on a page because the user of the system would have to access a smaller number of pages to find his information.

6. Distribution of information: Pages that had a wide variety of categories to choose from were judged better than pages on which most of the information could be found under one title. For example, on one page, 90% of the answers to the questions were under the item "General Interest Guide".

7. Content of categories: Some participants judged a page on the content of the categories. For example, some thought it was better to have only one idea per category than to link many categories under one title. For example, "Business and Government" should not be part of the same title. Some people did not rank page Expert #6 favourably because they thought it confusing to choose when categorization was by action rather than by content. For example, one might want to look at restaurants but not necessarily to go out. One could want to do a survey of restaurants in a region of the city with the purpose of starting a restaurant business. Table 4 lists the frequency with which each criterion was used by participants to rank order the pages.

DISCUSSION

Experiment 1 showed that there was virtually no consistency among experienced users in their preferences for alternative "first" index pages. In contrast, Experiment 2 showed that there was considerable consistency among naive Telidon users both in preferences and in performance on index pages. High consistency implies that different users tend to base their subjective judgments of the "goodness" of index pages on the same factors. The high correlation between performance and preference judgments suggests that subjective preference judgments are based, at least in part, on the problems experienced in trying to choose the right index term. The implication is that if a tree index is improved by reducing the number of errors, the improvement should also result in an increased level of user satisfaction with the index.

The difference between experienced and naive users in the consistency with which they ranked the 12 index pages on preference could be attributable to two different factors. First, as a result of experience on Telidon, experienced users may have developed quite different perceptions of tree structures. Second, it is possible that the experienced users would have been much more consistent in their preferences if they had had the opportunity to look for specific information on each page as the naive users had. The present study does not provide the information necessary to decide between these two alternative interpretations.

Overall, the best "first" index page was the modified page with descriptors. This page was most preferred and had the second-fewest errors. (The index page with the fewest errors - expert#4 with a single descriptor - was ranked very low - sixth - on user preference. In fact, it was ranked significantly lower than the modified page with descriptors.) The superiority of the modified page with descriptors can probably be attributed to two factors. First, the modified page represents an empirical improvement over the original index page that was developed by Lee & Latrêmouille (1981). In the Lee & Latrêmouille (1981) series of experiments, the original tree structure was tested under more realistic conditions, the location of errors identified, and the errors were corrected. Therefore, the modified index page could be expected to be an improvement. Second, the addition of descriptors to the modified page significantly improved the page. Five of the 12 index pages were among the top pages on both measures and were not significantly different from one another on either preference or performance: modified:D, original:D, expert #2:D, expert #3:1D, and expert #1:ND. Four of these five pages have descriptors which provides further support for the superiority of descriptors over no descriptors. Moreover, given the results of the present experiment, it is very probable that the fifth page (expert #1:ND) could be markedly improved by adding descriptors.

An important finding in the present study is the demonstrated value of adding descriptors to an index page. Not only did adding descriptors to a page produce a significant improvement in every case, but the amount of improvement was marked. Index pages without descriptors had two to three times as many errors as corresponding pages with descriptors. Similarly, pages without descriptors were ranked much lower on user preferences than corresponding pages with descriptors. Another indication of the magnitude of the effect of descriptors on user preference can be seen from Figure 3. In a previous study (Lee & Latrêmouille, 1981), the original page with no descriptors was tested (along with the rest of the original tree structure), errors were identified and than corrected to produce the improved "modified" tree index. Figure 3 indicates that the modification of the original page with no descriptors to produce the modified page with no descriptors resulted in a considerable and significant improvement in the average preference ranking. However, the magnitude of improvement in average preference ranking of the original page by simply adding descriptors was much greater than it was by correcting the problems identified in the previous Lee & Latrêmouille (1981) study.

The comments made by naive users on their criteria for ranking the 12 index pages on preference provide useful insights, for those constructing videotex tree indexes, into the factors affecting user satisfaction with the tree structure. The most frequently reported criterion was the amount of description. Virtually all users who mentioned this criterion (65% of users) preferred descriptors to be added to index terms. A few of the users added the proviso, however, that such descriptors should be brief and not too wordy. Only a single user indicated a preference for no descriptors.

The second most frequently reported criterion (60% of the naive users) was the generality of index terms. Terms such as "General Interest" and "Miscellaneous" were judged to be far too ambiguous and overly general. To naive users such terms are meaningless for the term itself tells the user nothing about the type of information classified under it. The order or sequence of index terms on a page was mentioned by 45% of the naive users. However, users differed somewhat in their preferences for particular kinds of sequences. Several users felt that most frequently accessed index terms should be at the top of the page to minimize the time spent searching the list. Other users felt that more general terms should be placed at the top of the list of terms on a page with less general terms at the bottom. Finally, 35% of users reported that the number of index terms on a page should be kept relatively high, three or four index terms per page was considered to be wasteful because it would require retrieval of more pages. This provides support for the recommendation, made by Lee (1979), that placing only 2 or 3 index terms on a page should be avoided because of the excessively long search times required.

Based on the comments made by naive users, the following recommendations can be made to those constructing videotex tree structures: first, use descriptors; second, avoid the use of over-general, ambiguous index terms such as General Interest and Miscellaneous; third, avoid the use of too few (2, 3, or 4) index terms on a page; fourth, if possible, use index terms that are roughly equivalent in the amount of information classified under each (in particular, avoid the use of terms such as General Interest which contain 80-90% of all information retrievable from a given index page); and fifth, indexers should try wherever possible to arrange index terms on a page in sequence from the most general to the least general index term or from the most frequently accessed to the least frequently accessed index term.

Based on the results of the present study, therefore, it is recommended that those in charge of videotex databases consider the addition of descriptors to index pages, particularly near the top of the tree. The addition of descriptors to index pages should significantly enhance user satisfaction as well as reducing the number of errors considerably. If descriptors are useful to naive users because they clarify the type of information categorized under the index terms, then descriptors should be useful at all levels of the tree structure. Further experimentation will be necessary, however, to determine whether descriptors should be added to all levels of the database. Further research will also be required to determine the effect of descriptors on transmission, display, reading, and search times.

REFERENCES

Garner, W.R., & Creelman, C.D. Problems and methods of psychological scaling. In G.F. Sumers (Ed.) Attitude measurement. Chicago, Rand McNally, 1970.

Hollander, M., & Wolfe, D.A. Nonparametric statistical methods. New York: Wiley, 1973, Chapter 7.

Kirk, R. Experimental design: Procedures for the behavioural sciences. Belmont, California: Brooks/Cole, 1968.

Latrêmouille, S., & Lee, E. The design of Telidon tree indexes: Improving tree indexes by testing naive users. Submitted for publication as a Technical Memorandum, Dept. of Communications, Canada, 1981.

Lee, E., & Latrêmouille, S. Evaluation of tree structured organization of information on Telidon. (Technical Memorandum No. BRG 79-12). Department of Communications, Ottawa. December, 1979. Also in Telidon Behavioural Research I. Department of Communications, Ottawa, 1980, pp.231-242.

McEwen, S.A. An investigation of user search performance on the Telidon information retrieval system. Submitted for publication as a Technical Memorandum, Department of Communications, Canada, 1981.

Siegel, S. Nonparametric statistics for the behavioural sciences. New York: McGraw-Hill, 1956.

van Ness, F.L., & Tromp, J.H. Is view data easy to use? Institute For Perception Research. Eindhoven, Netherlands: IPO Annual Progress Report, No. 14, 1979.

Whalen, T., & Latrêmouille, S. The effectiveness of a tree-structured index when the existence of information is uncertain. (Technical Memorandum No.BRIC 80-3). Department of Communications, Ottawa. September, 1980.

TABLE 1

PREFERENCE RANKING OF THE INDEX PAGES BY
EXPERIENCED USERS: EXPERIMENT 1

RANK	INDEX PAGE[a]	NORMALIZED RANK[b]	MEAN RANK[b]
1	Modified:ND	14.82 (0.69)	2.90 (2.08)
2	Modified:D	9.40 (1.07)	5.10 (3.87)
3	Expert #2:D	6.25 (1.28)	5.80 (4.49)
4	Expert #1:ND	4.47 (0.76)	6.40 (2.76)
5	Expert #3:1D	4.37 (0.92)	6.40 (3.50)
6	Original:ND	4.03 (0.78)	6.20 (2.97)
7	Expert #5:ND	3.78 (0.77)	6.95 (3.32)
8	Expert #4:1D	3.40 (0.74)	6.90 (3.11)
9	Librarian:ND	2.60 (0.70)	7.25 (2.66)
10	Original:D	1.47 (0.93)	7.50 (3.57)
11	Expert #6:D	1.37 (1.24)	8.05 (4.26)
12	Librarian:D	0 (0.54)	8.55 (2.29)

[a] D = index page with descriptors; ND = no descriptors; 1D = index page having a descriptor for only one index term.

[b] Entries in parentheses represent the standard deviation of scores for the corresponding mean.

TABLE 2

PREFERENCE RANKING OF THE FIRST INDEX PAGES
BY NAIVE USERS: EXPERIMENT 2

RANK	INDEX PAGE[a]	NORMALIZED RANK[b]	MEAN RANK[b]
1	Modified:D	43.65 (0.40)	2.40 (1.05)
2	Expert #2:D	42.38 (0.92)	3.05 (3.14)
3	Original:D	40.28 (0.52)	3.00 (1.72)
4	Expert #1:ND	27.34 (0.48)	5.40 (2.01)
5	Modified:ND	26.93 (0.86)	5.85 (3.01)
6	Expert #3:1D	24.53 (0.82)	5.60 (2.85)
7	Expert #4:1D	22.81 (0.64)	6.25 (2.49)
8	Librarian:D	15.26 (0.46)	7.95 (1.73)
9	Expert #6:D	10.48 (0.67)	8.65 (2.21)
10	Original:ND	8.25 (0.40)	9.40 (1.50)
11	Librarian:ND	3.76 (0.56)	10.05 (1.96)
12	Expert #5:ND	0 (0.65)	10.40 (2.06)

[a] D = index page with descriptors; ND = no descriptors; 1D = index page having a descriptor for only one index term.

[b] Entries in parentheses represent the standard deviation of ranks for corresponding means.

TABLE 3

MULTIPLE-COMPARISON TESTS OF THE DIFFERENCE IN PREFERENCE RANK SUMS FOR INDEX PAGES: EXPERIMENT 2[a]

		Index Page[b]										
		10	5	4	6	2	7	12	9	1	3	8
Index Page[b]	11	12	13	60	64	69	77*	*	*	*	*	*
	10	-	1	48	52	57	65	99*	*	*	*	*
	5		-	47	51	56	64	98*	*	*	*	*
	4			-	4	9	17	51	65	80*	*	*
	6				-	5	13	47	61	76*	*	*
	2					-	8	42	56	71	84*	*
	7						-	34	48	63	76*	*
	12							-	14	29	42	49
	9								-	15	28	35
	1									-	13	20
	3										-	7
	8											-

[a] Entries in the table represent differences between rank sums for index pages indicated in the corresponding row and column.

[b] Page 1 = original:ND; 2 = modified:ND; 3 = librarian:ND; 4 = expert#1:ND; 5 = expert#2:D; 6 = expert#3:1D; 7 = expert#4:1D; 8 = expert#5:ND; 9 = expert#6:D; 10 = original:D; 11 = modified:D; 12 = librarian:D.

*Entries marked with a star are significantly different at <u>p</u> less than .05.

TABLE 4

MEAN NUMBER OF ERRORS COMMITTED BY NAIVE USERS ON EACH FIRST INDEX PAGE: EXPERIMENT 2

RANK	INDEX PAGE[a]	MEAN NUMBER OF ERRORS[b]
1	Expert #4:1D	2.25 (2.00)
2	Modified :D	2.50 (1.76)
3	Original :D	2.65 (2.43)
4	Expert #2:D	2.75 (1.52)
5	Expert #3:1D	2.75 (1.89)
6	Expert #1:ND	3.05 (2.16)
7	Expert #6:D	3.25 (1.59)
8	Expert #5:ND	4.60 (1.96)
9	Modified:ND	4.85 (1.39)
10	Librarian:D	5.35 (2.30)
11	Original:ND	6.35 (2.28)
12	Librarian:D	6.50 (1.76)

[a] D = index page with descriptors; ND = no descriptors; 1D = index page having a descriptor for only one index term.

[b] Entries in parentheses represent the standard deviation of ranks for corresponding means.

TABLE 5

NEUMAN-KEULS MULTIPLE COMPARISON ANALYSIS OF MEAN NUMBER OF ERRORS PER INDEX PAGE: EXPERIMENT 2[a,b]

Index Page	11	10	5	6	4	9	8	2	12	1	3
7	0.25	0.40	0.50	0.50	0.80	1.00	2.35*	2.60*	3.10*	4.10*	4.25*
11	-	0.15	0.25	0.25	0.55	0.75	2.10*	2.35*	2.85*	3.85*	4.00*
10		-	0.10	0.10	0.40	0.60	1.95*	2.20*	2.70*	3.70*	3.85*
5			-	0	0.30	0.50	1.85*	2.10*	2.60*	3.60*	3.75*
6				-	0.30	0.50	1.85*	2.10*	2.60*	3.60*	3.75*
4					-	0.20	1.55*	1.80*	2.30*	3.30*	3.45*
9						-	1.35*	1.60*	2.10*	3.10*	3.25*
8							-	0.25	0.75	1.75*	1.90*
2								-	0.50	1.50*	1.65*
12									-	1.00*	1.15*
1										-	0.15
3											-

[a] Page 1 = original:ND; 2 = modified:ND; 3 = librarian:ND; 4 = expert#1:ND; 5 = expert#2:D; 6 = expert#3:1D; 7 = expert#4:1D; 8 = expert#5:ND; 9 = expert#6:D; 10 = original:D; 11 = modified:D; 12 = librarian:D.

[b] Entries in the table represent differences between means for index pages indicated in corresponding row and column.

*Entries marked with a star are significantly different at p less than 0.05.

TABLE 6

REPEATED MEASURES ANALYSIS OF EMBEDDED
2 X 3 DESIGN (PAGES X DESCRIPTORS): EXPERIMENT 2

SOURCE	df[a]	SS	MS	F	% TOTAL VARIATION
Subjects Within	19	248.53			
Pages	2(1)	103.65	51.82	22.62**	13.4%
Descriptors	1(1)	172.80	172.80	75.42**	22.3%
Pages X Descriptors	2(1)	32.55	16.28	7.10*	4.2%
Error	95(19)	217.67	2.29		
TOTAL	119	775.20			

[a] The degrees of freedom for the Geisser-Greenhouse conservative *F* test are specified in parentheses.

**p* less than 0.01 for the conventional *F* test and *p* less than 0.05 for the conservative *F* test.

***p* less than 0.001 for both the conventional and conservative *F* tests.

TABLE 7

SIMPLE MAIN-EFFECTS ANALYSIS OF THE EMBEDDED
2 X 3 DESIGN (DESCRIPTORS X PAGES): EXPERIMENT 2

SOURCE	df	SS	MS	F
SUBJECTS WITHIN	19	248.53		
Pages (P)	2	103.65	51.82	22.62**
P at Descriptor	1	102.90	102.90	22.47**
P at No Descriptor	1	33.30	33.90	7.27**
Descriptors (D)	1	172.80	172.80	75.42**
D at Original	1	136.90	136.90	59.78**
D at Modified	1	55.22	55.22	24.11**
D at Professional	1	13.22	13.22	5.77*
Descriptors X Pages	2	32.55	16.28	7.10**
Error	95	217.67	2.29	
TOTAL	119	775.20		

* *p* less than 0.05.
***p* less than 0.01.

TABLE 8

FREQUENCY OF USE OF CRITERIA FOR PREFERENCE RANKING OF THE 12 INDEX PAGES: EXPERIMENTS 1 AND 2

	Experiment 1 (N=10)	Experiment 2 (N=20)
1. Amount of description		
In favour of:	5	10
In favour but titles Should not be too wordy:	-	2
Not in favour of:	5	1
2. Generality of titles	7	12
3. Complexity of words	2	6
4. Order of titles	5	9
5. Number of titles on a page		
Should be low	1	-
Should be high	3	7
6. Distribution of information	2	5
7. Content of categories	6	9

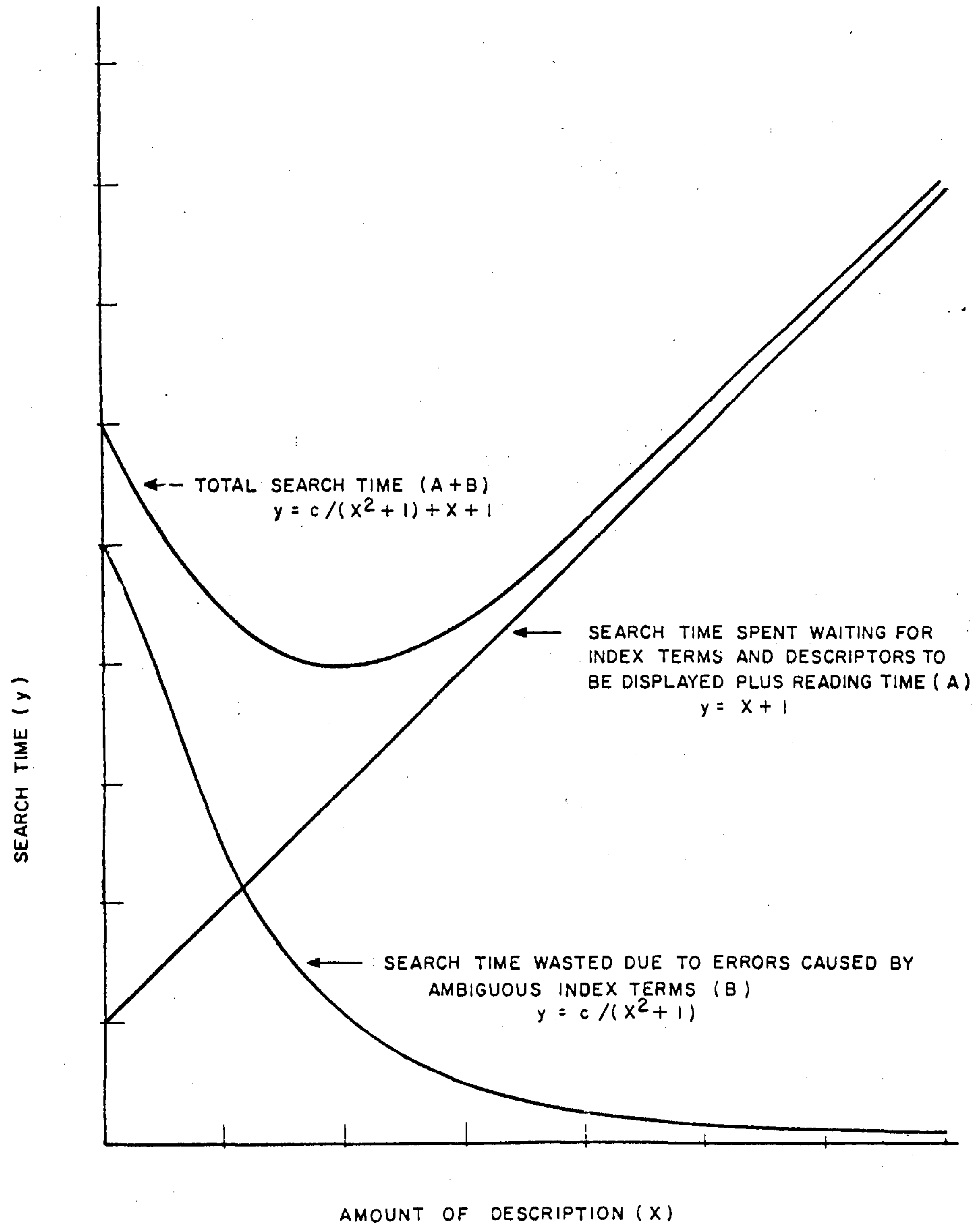

FIG. I PREDICTED RELATIONSHIP BETWEEN THE AMOUNT OF DESCRIPTION AND USER SEARCH TIME

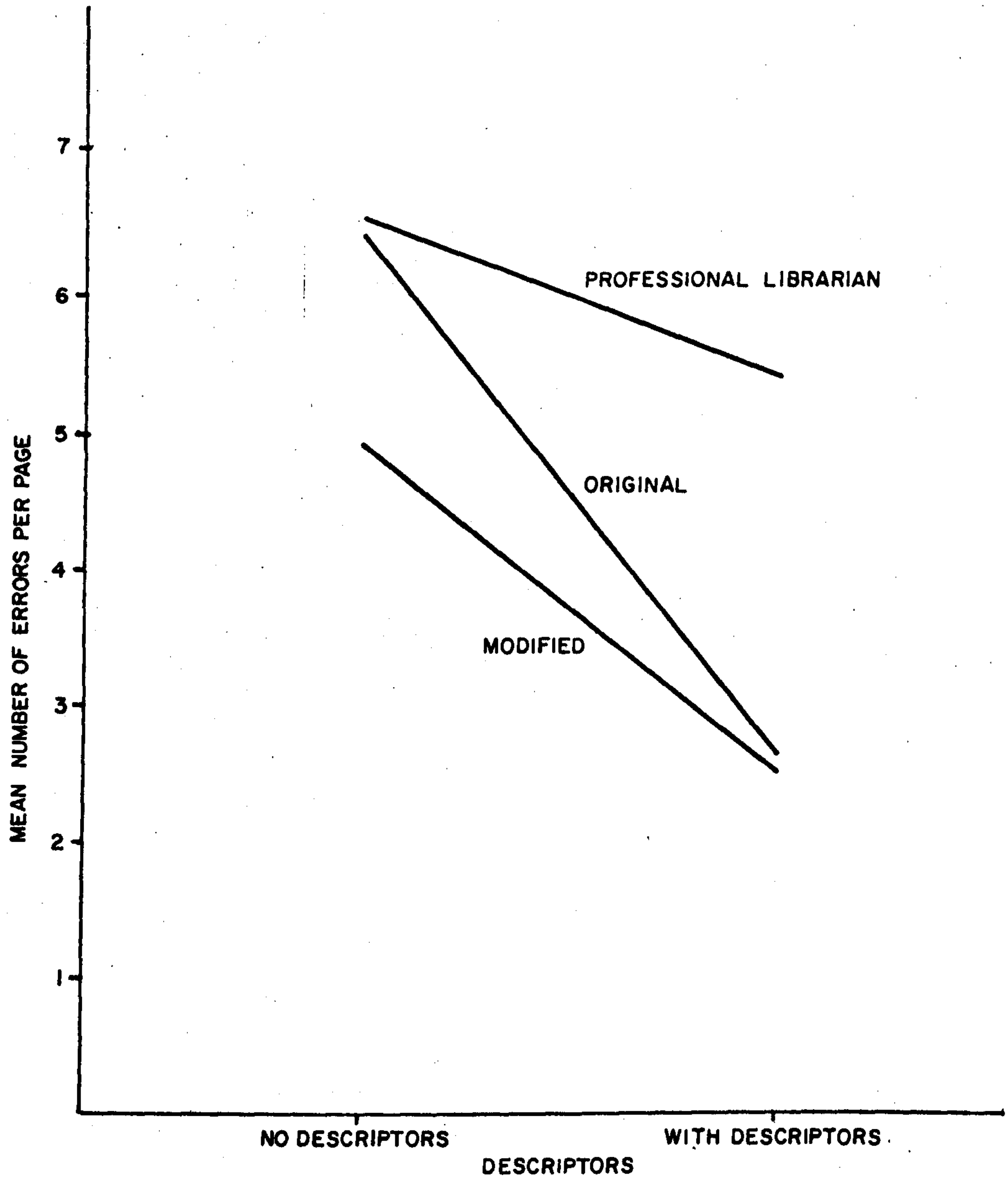

FIG. 2 DESCRIPTOR BY PAGE INTERACTION, EXPERIMENT 2.

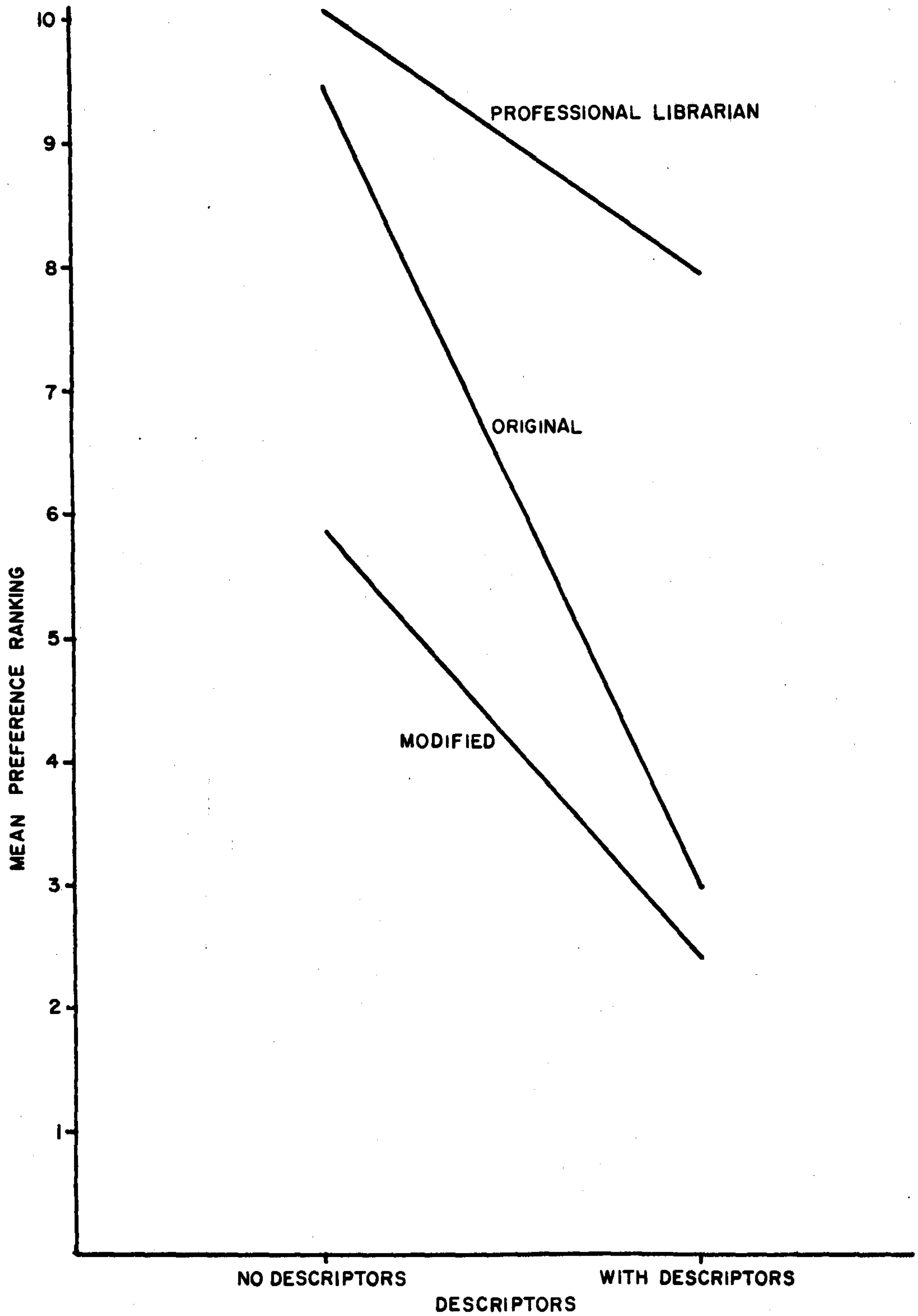

FIG. 3 COMPARISON OF MEAN PREFERENCE RANKING FOR INDEX PAGES WITH AND WITHOUT DESCRIPTORS, EXPERIMENT 2

APPENDIX A

THE 12 "FIRST" INDEX PAGES EMPLOYED IN EXPERIMENTS 1 AND 2

ORIGINAL:ND

1. GENERAL INTEREST GUIDE
2. BUSINESS GUIDE
3. CANADIAN GOVERNMENT
4. EMERGENCY
5. USER'S GUIDE
6. TELIDON EXPLANATION
7. TELEPHONE NUMBERS

MODIFIED:ND

1. TELIDON EXPLANATION AND INSTRUCTIONS
2. EMERGENCY
3. BUSINESS, FINANCE, AND GOVERNMENT INFORMATION
4. ENTERTAINMENT AND TRAVEL
5. GENERAL INFORMATION
6. BUY, SELL, RENT AND SERVICES
7. EDUCATION
8. EMPLOYMENT
9. TELEPHONE DIRECTORY

LIBRARIAN:ND

1. GENERAL INTEREST GUIDE
2. BUSINESS
3. MISCELLANY
4. EMERGENCY
5. TELIDON
6. NOTICE BOARD
7. TELEPHONE NUMBERS

EXPERT #1:ND

1. NEWS, WEATHER AND SPORTS, NOTICES
2. ENTERTAINMENT AND RECREATION
3. CONSUMER AND BUSINESS INFORMATION
4. CANADIAN GOVERNMENT
5. HEALTH, EDUCATION, WELFARE, EMPLOYMENT
6. EMERGENCIES
7. HOW TO USE TELIDON
8. HOW TELIDON WORKS (TECHNICAL)
9. TELEPHONE DIRECTORY
10. TRAVEL

EXPERT #2:D

1. EMERGENCY: Telephone numbers, Basic first aid

2. INFORMATION FOR TODAY: News, Weather, Sports, Neighbourhood news

3. GENERAL INFORMATION: Consumer reports, Metric conversion, Education, Health guide

4. BUSINESS SECTION: Buy, Sell, Rent, Commercial services, Financial matters, Business advice

5. NEED A JOB?

6. GOING ANYWHERE?: Transportation, Travel, Tourist guide

7. RELAX: Going out, Home leisure, Special events, Sports

8. CANADIAN GOVERNMENT: House of Commons, Members of Parliament, Federal reports

9. TELIDON INFORMATION: Technical descriptions and How to use it

10. TELEPHONE NUMBERS

EXPERT #3:1D

1. NEWS, WEATHER AND SPORTS

2. ENTERTAINMENT AND TRAVEL

3. EDUCATION AND REFERENCE INFORMATION

4. THE MARKETPLACE
   a. Local shopping/services
   b. Consumer guide
   c. Employment
   d. Real estate (housing)

5. BUSINESS AND GOVERNMENT

6. EMERGENCY AND MEDICAL AID

7. FIELD TRIAL DATABASES

8. TELIDON EXPLANATION AND INSTRUCTION

EXPERT #4:1D

1. GENERAL INTEREST GUIDE
   News, Weather and Sports
   Entertainment (local)
   Shopping/housing
   Education
   Daily noticeboard
   Travel and tourism
   Reference information
   Employment

2. BUSINESS AND FINANCIAL POST

3. THE CANADIAN GOVERNMENT

4. FIELD TRIAL DATABASES

5. EMERGENCY AND MEDICAL AID

6. TELIDON INSTRUCTIONS AND EXPLANATION

EXPERT #5:ND

1. PERSONAL ACTIVITIES

2. BUSINESS

3. GOVERNMENT

4. NEWS, WEATHER, SPORTS

5. EMERGENCY

6. UNDERSTANDING TELIDON

EXPERT #6:D

1. GOING OUT: Local entertainment, Travel services, Shopping, etc.

2. STAYING AT HOME: Educational games, Stories, Home improvements, etc.

3. RECENT EVENTS: News stories, Sports scores, Weather forecast, etc.

4. REFERENCE MATERIAL: Telephone numbers, Encyclopedia, etc.

ORIGINAL:D

1. GENERAL INTEREST GUIDE: News, Weather, Sport, Entertainment, Market Place, Employment, Travel, Leisure, Advice, Education, Notice board

2. BUSINESS GUIDE: Nation's business, List of companies, Investments, Careers, Opportunities, Real Estate Market, Advertising, Marketing

3. CANADIAN GOVERNMENT: House of Commons, Senate, Federal reports, Provincial Governments, Municipal Governments, Prime Ministers

4. EMERGENCY: Phone numbers, Agencies, Basic first aid

5. USER'S GUIDE: How to use TELIDON

6. TELIDON EXPLANATION: Technical explanation

7. TELEPHONE NUMBERS

MODIFIED:D

1. TELIDON EXPLANATION AND INSTRUCTIONS

2. EMERGENCY: Telephone numbers and Basic first aid procedures

3. BUSINESS, FINANCE AND GOVERNMENT INFORMATION: Financial matters, Investment, Economy, Taxes

4. ENTERTAINMENT AND TRAVEL: Entertainment outside or inside house, Sports, Transportation, Tourist guide, Travel Agencies

5. GENERAL INFORMATION: Consumer information, Metric conversion, Horoscope, News, Weather, Sports, Notices

6. BUY, SELL, RENT AND SERVICES: Buy merchandise, property, rent equipment or apartments, Commercial services such as home renovation, plumbing, electricity

7. EDUCATION

8. EMPLOYMENT

9. TELEPHONE DIRECTORY

LIBRARIAN:D

1. TABLE OF CONTENTS for the General Interest Guide

2. BUSINESS Reports, Business opportunities, Commercial real estate, Commercial services

3. MISCELLANY: Metric conversion tables, Horoscope...

4. EMERGENCY: Telephone numbers, First aid procedures

5. TELIDON: How to use it? How it works?

6. NOTICE BOARD: Births, Deaths, Personal advertisements

7. TELEPHONE NUMBERS

APPENDIX B

QUESTIONS

1. Find a list of televisions for sale.
2. Find out what is available for individual transportation.
3. Find out what the weather is for this area today.
4. Read about the good qualities of a stove.
5. Find out where the restaurant 'CHINESE VILLAGE' is situated.
6. Find the mathematical game 'What is the missing number?'
7. Find out how to complain when you know you have been cheated by a company.
8. Find information about the afghanistan restaurant 'KHYBER PASS'.
9. Read about the good qualities of a freezer.
10. Find what the world grain exports are.

APPENDIX C

INDIVIDUAL PREFERENCE RANKINGS

EXPERIMENT 1

Page

		1	2	3	4	5	6	7	8	9	10	11	12
Subject	#1	2	8	5	6	9	10	4	7	12	11	3	1
	2	8	4	2	3	12	10	7	1	9	11	6	5
	3	8	6	3	1	9.5	9.5	7	2	4	5	11.5	11.5
	4	12	3	6	1	8	7	9	2	4	5	11	10
	5	5	11	1	10	6	3	2	12	4	9	7	8
	6	4	6	1	9	3	11	5	10	8	7	2	12
	7	9	3	6	1	8	10	5	2	7	4	11	12
	8	4	12	3	9	6	8	7	2	11	10	5	1
	9	5	11	1	9	4	8	6	12	2	3	7	10
	10	5	11	1	2	7	9	12	8	3	4	6	10

Note. Page 1 = original:ND, 2 = original:D, 3 = modified:ND, 4 = modified:D, 5 = librarian:ND, 6 = librarian:D, 7 = expert#1:ND, 8 = expert#2:D, 9 = expert#3:1D, 10 = expert#4:1D, 11 = expert#5:ND, 12 = expert#6:D.

EXPERIMENT 2

	Page											
	1	2	3	4	5	6	7	8	9	10	11	12
Subject #1	10	1	8	2	12	9	7	3	5	4	11	6
2	8	3	4	1	10	7	5	2	12	6	11	9
3	10	1	6	5	11	9	7	3	4	2	12	8
4	10	2	8	3	11	7	9	1	4	5	12	6
5	8	3	5	2	11	7	6	1	4	12	10	9
6	11	3	1	2	12	7	4	6	5	10	9	8
7	9	5	6	4	12	8	2	1	3	7	11	10
8	11	1	6	3	9	12	4	2	5	7	8	10
9	8	2	11	3	10	6	7	1	5	4	9	12
10	11	3	8	1	10	9	7	2	5	4	12	6
11	10	3	8	1	11	6	9	2	4	5	12	7
12	10	2	4	1	11	9	5	6	3	7	8	12
13	11	2	9	3	10	4	6	1	5	8	12	7
14	11	2	8	3	10	7	6	1	5	4	12	9
15	10	8	9	2	11	7	3	1	5	4	12	6
16	7	5	1	3	8	9	2	11	12	10	4	6
17	7	5	1	2	9	10	3	4	6	8	11	12
18	11	2	7	3	10	9	6	1	4	5	12	8
19	8	3	6	2	10	9	5	1	4	7	11	12
20	7	4	1	2	3	8	5	11	12	6	9	10

Note. Page 1 = original:ND, 2 = original:D, 3 = modified:ND, 4 = modified:D, 5 = librarian:ND, 6 = librarian:D, 7 = expert#1:ND, 8 = expert#2:D, 9 = expert#3:1D, 10 = expert#4:1D, 11 = expert#5:ND, 12 = expert#6:D.

Canada

Government of Canada
Department of Communications

Gouvernement du Canada
Ministère des Communications