AN EXPERIMENT IN TEACHING NEPHIS, A NESTED-PHRASE INDEXING SYSTEM (UNE EXPERIENCE AVEC L'ENSEIGNEMENT DU SYSTEME D'INDEXATION NEPHIS)

Timothy C. Craven
School of Library and Information Science
The University of Western Ontario
London, Ontario
N6A 5B9

ABSTRACT

NEPHIS is a system of computer-assisted permuted subject indexing designed to be easy for the indexer, the programmer, and the user, and to be economical. An experiment is being performed, using a small group of paid subjects, with the aim of seeing how readily the NEPHIS system can be learned and what approach to teaching it might be the most suitable. (NEPHIS, c'est un système qui employe un ordinateur dans la création des index-matières permutés, et qui vise à être facile pour l'indexateur, pour le programmeur, et pour l'utilisateur, aussi qu'économique. On fait une expérience, avec un petit groupe de sujets payés, dans l'espoir de voir avec quelle facilité l'on puisse apprendre le système et quelle méthode d'enseignement soit la meilleure.)

INTRODUCTION

NEPHIS is a computer-assisted indexing system developed by the author at The University of Western Ontario's School of Library and Information Science (SLIS). Like D. Austin's PRECIS and G. Bhattacharyya's POPSI, it is a system of permuted subject indexing. Its core is the NEPHIS program (which has been implemented on the DECsystem10 at The University of Western Ontario's Computing Centre). This program generates, from a file of input strings produced by a human indexer, a file of output strings or permutations. When sorted and printed out, this file of permutations forms a printed index which is elegant and browsable. (Figure 1 gives a sample page out of such an index.)

Figure 1: Sample Page from a NEPHIS Index

INFORMATION SERVICES AT MECHANIZED INFORMATION CENTER, OHIO STATE UNIVERSITY
PROMOTION, METHODS: OPINION LEADERSHIP & BLITZ & TELEPHONE SOLICITATION, COMPARISON, JASIS 24.171
INFORMATION STORAGE & RETRIEVAL IN SOCIAL SCIENCES: EDUCATIONAL RESOURCES INFORMATION CENTERS (ERIC)
CHARACTERISTICS & PROBLEMS, JASIS 24.193
INFORMATION STORAGE & RETRIEVAL SYSTEMS
DESIGN, SYSTEMS APPROACH, JASIS 24.205
KEYWORD -: UN ECONOMIC COMMISSION FOR EUROPE & UN CONFERENCE ON TRADE & DEVELOPMENT, PROGRAMMING, JASIS 25.374
INFORMATION SYSTEMS
DIALECTIC -, APPLICATIONS: EXPERIMENTS, JASIS 25.252
RESOURCES, ALLOCATION: APPLICATIONS OF QUEUING THEORY & DYNAMIC PROGRAMMING, JASIS 25.52
INFORMATION TECHNOLOGY & INFORMATION SCIENCE
DEVELOPMENT, EFFECT OF CUTTING OF BUDGET OF NSF-OSIS, 1975, JASIS 25.77
INFORMATION THEORY
MEASURES OF INFORMATION CONTENT OF DOCUMENT SURROGATES, JASIS 24.300
INTERACTIVE RETRIEVAL SYSTEMS USING BIBLIOGRAPHICAL & BUSINESS-ORIENTED & SCIENTIFIC DATA BASES
PROGRAMMING LANGUAGES; DIRAC, DESIGN, JASIS 24.287
INTERACTIVE SEARCHING
ON-LINE - OF DATA BASES IN INDUSTRIAL RESEARCH ENVIRONMENTS: EXXON RESEARCH & ENGINEERING COMPANY, JASIS 25.364
INTERACTIVE SYSTEMS
COMPUTER-BASED -: NEGOTIATED SEARCH FACILITY, DEVELOPED AT IBM, SAN JOSE, CALIFORNIA, SUBJECT INDEXING, EFFECTIVENESS, EXPERIMENTS, JASIS 24.9
INTERFACE BETWEEN HUMAN BEINGS & MACHINES IN BIBLIOGRAPHICAL RETRIEVAL SYSTEMS
DESIGN, OPINIONS, SURVEYS, JASIS 24.142
INTERFACE BETWEEN MACHINES & USERS IN RETRIEVAL SYSTEMS
USING BIOSIS
DESIGN, JASIS 25.3
INTERNATIONAL DOCUMENTATION IN CHEMISTRY
TOSAR SYSTEM: REPRESENTATION OF CONCEPTS & RELATIONS BETWEEN CONCEPTS: APPLICATIONS OF GRAPH THEORY, JASIS 25.287
INTERNATIONAL INVISIBLE COLLEGES IN HIGH ENERGY PHYSICS
IDENTIFICATION: APPLICATIONS OF SOCIOMETRICS, JASIS 25.113
INVISIBLE COLLEGES
INTERNATIONAL - IN HIGH ENERGY PHYSICS, IDENTIFICATION: APPLICATIONS OF SOCIOMETRICS, JASIS 25.113
JOURNALS
BACK ISSUES, BINDING & DISCARDING & MICROCOPYING, DECISION, ALGORITHMS, JASIS 25.213

Like the index itself, the input strings are in many ways similar to natural language. Usually an input string takes the basic form of a noun phrase with other noun phrases "nested" within it. The indexer needs to know only four special characters ("<", ">", "?", and "@") and the four commands that these define. The symbols "<" and ">" are used to set off a nested phrase; the symbol "?" is used to flag a connective (which may be forward-reading or backward-reading, depending on whether it is terminated by "<" or by ">"; the symbol "@" is used to suppress an unwanted permutation. For example, a typical input string produced by a NEPHIS indexer is

> @Case Studies? of <Psychosocial Aspects? of <Adaptability? of <Adults? in <United States>>? to <Transitions>>>.
> #123

This instructs the NEPHIS program to generate the permutations

> Psychosocial Aspects of Adaptability of Adults in United States to Transitions. Case Studies. #123
>
> Adaptability of Adults in United States to Transitions. Psychosocial Aspects. Case Studies. #123
>
> Adults in United States. Adaptability to Transitions. Psychosocial Aspects. Case Studies. #123
>
> United States. Adults. Adaptability to Transitions. Psychosocial Aspects. Case Studies. #123
>
> Transitions. Adaptability of Adults in United States. Psychosocial Aspects. Case Studies. #123

The NEPHIS system was designed with four objectives in mind: 1. it should be easy for the indexer; 2. the program should be easy to write; 3. running the program should be economical; 4. the index produced should satisfy the users. It is with the first of these objectives that this paper is chiefly concerned.

From the example given above, it is clear that the NEPHIS system requires a minimal amount of typing-in on the part of the indexer, and so helps to eliminate random typographical errors, as well as speeding up the indexing process. Moreover, on the face of it, a system which requires the indexer to know only four commands should be easy to learn and to remember. Yet there is somewhat more to indexing using the NEPHIS system than merely knowing the commands that can be given to produce various permutations. There is also the matter of mastering the techniques of the system.

It was decided to do a little exploratory experimentation in the teaching of NEPHIS to naive prospective indexers. This experimentation aims at getting answers to the following sort of questions:

1. How quickly can the techniques of NEPHIS indexing be acquired?

2. How may the system and the techniques associated with it best be taught?

3. What information should be included in an indexer's manual?

4. What policy decisions are likely to have to be made in a specific environment?

5. Are there any improvements that can be made in the basic permutation algorithm?

METHODOLOGY

With these diverse aims in mind, it cannot be a question of performing a classic experiment, but must rather be one of collecting and analysing a variety of data in a semi-controlled situation. Methodology must be determined by circumstances.

A group of ten (later seven) paid subjects was obtained from among the student population of SLIS. As a preliminary step, each subject was given a copy of the questionnaire illustrated in Figure 2, together with a copy of a paper describing the general features of NEPHIS and instruction sheets for the use of the on-line demonstration program. This demonstration program (called NEPHEX) was written especially for the purpose of the experiment. Instead of reading input strings from an input file and writing the permutations on an output file, the demonstration program accepts input strings as they are typed in by the person running the program and returns the resulting permutations on the terminal when requested to do so. It also allows the person running the program to edit a faulty input string on line, instead of having to retype the whole thing. (Figure 3 shows a sample of how the demonstration program works.)

Five steps were planned after the preliminary step of reading the background material and answering the questionnaire. Each of these steps entails the subject's using the demonstration program to complete an exercise in the use of the NEPHIS system, each exercise consisting of ten items. All work on the exercises is to be done on line, the connect time being used as the basis of payment. As a subject completes an exercise, he or she submits all printouts produced. The printouts can then be analysed

Figure 2: Questionnaire

STEP # 0

(Please take note of the time it takes you to complete this step, and enter it in the space marked "Time" below).

Please answer the following questions by checking the appropriate positions.

1. Do you have any practical experience with
   a. indexing? . . . . yes / / no / /
   b. assigning subject headings? . . . . yes / / no / /
   c. classification . . . . yes / / no / /

2. Are you familiar with
   a. PRECIS? . . . . yes / / somewhat / / no / /
   b. Relational Indexing? . . . . yes / / somewhat / / no / /

3. Have you used an on-line computer system before? yes / / no / /

4. How well can you type? very well / / quite well / / adequately / /
   not very well / / not at all / /

Now read the paper entitled "NEPHIS: a Nested-Phrase Indexing System". Make sure you understand it reasonable well, but do not spend too much time on fine points.

Read also the sheets on the on-line system NEPHEX. Concentrate on how to initiate and how to terminate a run.

Record your time in the space below.

Time: ____ hrs. ____ mins.

YOUR INDEXER NUMBER IS:

Figure 3: Sample Run of the Demonstration Program
(Part of the Typed Instructions Given to Subjects)

The following example illustrates how NEPHEX works in practice (underlined parts are those typed in by the user):--

```
.EX NEPHEX
LOADING
NEPHEX 1K CORE
EXECUTION
*$FEEDING? OF <AARDVARKS. PAGE 123
*%
? UNMATCHED BRACKET IN INPUT STRING
*-ARKS   ARKS>
FEEDING? OF <AARDVARKS>. PAGE 123
*%
FEEDING OF AARDVARKS
  PAGE 123
AARDVARKS
  FEEDING. PAGE 123
*-VARKS  WOLVES
FEEDING? OF <AARDWOLVES>. PAGE 123
*-3      4
FEEDING? OF <AARDWOLVES>. PAGE 124
*%
FEEDING OF AARDWOLVES
  PAGE 124
AARDWOLVES
  FEEDING. PAGE 124
*$ZEBRAS? & <ZEBUS? & >. PAGE 567
*%
ZEBRAS & ZEBUS
  PAGE 567
ZEBUS & ZEBRAS
  PAGE 567
*^
EXIT
..K/F
```

in various ways (to see, for example, how well the subject is doing or what problems are arising). At the same time, the subject has an opportunity to discuss any particular difficulties encountered.

Each exercise has both teaching purposes and data-gathering purposes, as indicated in the following table:

Exercise Number	Information Provided for Each Item	Task to be Performed for each Item	Main Teaching Purpose	Sample Data-Gathering Purpose
1	Set of NEPHIS permutations	Find an input string which generates the set	Familiarization with the mechanics of the system	What "mechanical" errors are likely to occur?
2	One NEPHIS permutation plus the initial words of the other permutations of the set	Find an input string which generates the set	"	"
3	Title of a book, plus the initial words of a set of NEPHIS permutations	Find an input string which generates the set	How to construct an input string	What connectives are chosen between terms?
4	Subject heading for use in a classified catalogue or bibliography	Find an input string which generates a set of permutations to index the subject	Translation from one system to another	How well are conventional subject headings understood?
5	Abstract of an article	"	How to index in a "real" situation	How much agreement is there on the subject of a document?

For exercises 1 through 4, examples were taken from the September 24, 1975 issue of BNB. The examples provided by this source supply instances for many, though not all, of the techniques that might be used by a NEPHIS indexer.

RESULTS TO DATE

Response to the questionnaire showed that subjects were not familiar with PRECIS or with Relational Indexing, and were thus unlikely to be influenced by them in their approach to NEPHIS. Most said they could type"adequately". On other questions, there was a fairly even split between the "yes"'s and the "no"'s. Time spent on the preliminary step of answering the questionnaire and reading the background material varied from 30 minutes to 2 1/4 hours, suggesting likely differences in ability or approach among subjects. No strong correlations were observed.

Initial results from subsequent steps (the experiment is still proceeding at the time of writing) suggest that instruction merely by giving prospective indexers a general description of the system and then having them learn to use it by trial and error is likely to be inadequate (or at least extremely inefficient) in most cases. That is, for practical purposes, the system cannot be considered to be self-teaching.

In general, it may apparently be said that the prospective NEPHIS indexer needs help with analysing his or her mistakes. Some such help can be provided by refining the error messages of the demonstration program, but it will likely be a good plan to include examples of common errors in the indexer's manual.

It has been found that the instructions for online editing of input strings are inadequate. As a result, subjects have tended simply to ignore the editing capability and so have wasted a good deal of time re-typing the entire input string whenever a mistake is made.

Another problem that will have to be overcome when the NEPHIS system is taught in future is an excessive tendency on the part of subjects to be satisfied with approximations. For example, subjects have tended to be quite careless in their use of spaces. It should be pointed out in a future indexer's manual that spaces have significance in sorting, and should therefore be checked carefully. Moreover, the instructor (or, if there is no instructor, the student indexer him- or herself) should see that each exercise is completed correctly before the next exercise is begun.

CONCLUSION

It is the author's hope that, by making use of the mine of information provided by this experiment, it will be possible to produce a well-thought-

out indexer's manual and other documentation that will enable the NEPHIS system to be implemented quickly in a great variety of environments.