INTERACTIVE COMPUTER GRAPHICS IN MEDICAL INFORMATION RETRIEVAL AND CLASSIFICATION

David A. Thompson
Industrial Engineering
Stanford University

The size of the body of medical literature is vast and growing at a rapid rate. The National Library of Medicine collection contains about two million items and at the present rate of growth will double in about 7 years.

Attempts to improve the access to medical literature by professional practitioners, para-medical personnel, and life science researchers has principally included the MEDLARS service, a telephone/mail query service operating in batch mode; and interactive retrieval services operating on-line over long distance telephone lines such as AIM-TWX, developed by Systems Development Corporation.

While this latter service and the recently announced MEDLINE bring interactive search to the medical profession, they operate essentially in a typewriter-type mode. Keywords describing the problem area to be searched are entered into the terminal using a certain format and sequence, and if documents exist in the subset bounded by the keywords entered, they will be listed out by author and title.

About 5 years ago, in an attempt to rethink the nature of the interaction between man and the computer, we started developing a computer graphics procedure whereby an index tree, in the manner of a decision tree, would be displayed to the user describing the structure and contents of the literature file. The user could then search the displayed tree selecting alternative branches with a light pen. Normally, a hierarchically structured index branches from the general to the specific, and at any one point the user would be presented with the alternative choices he may make to define his problem more specifically as well as the list of decisions he has made so far in the event he may wish to back up and make alternate decisions. Figure 1 shows a typical portion of a search sequence.

At any point during the search the user may stop his selection among appropriate keywords and start paging through document surrogates listed under the last selected keyword. Figure 2 shows two typical document surrogates. Each surrogate lists all the keywords assigned to that document, in addition to the usual citation data, so that the user can get a reasonably good idea of its contents.

The complete philosophy behind this approach has been explained in detail elsewhere (Thompson, 1971B), but it should be noted here that the displayed index approach shifts much of the memory load for search query formulation to the computer rather than requiring the man to articulate a logical search request.

Experiments with medical student subjects have resulted in rather rapid response times for completing a search to the fifth level of the tree. After 3 initial searches during which added learning time is required, search times settle down to 17-18 seconds per completed search to a typical keyword target five levels deep in the tree. Paging through a document list appended to this keyword required additional time proportional to the list length, but typically takes only 2-3 seconds per document (Thompson, 1971B and Thompson & Di Giulio, 1972).

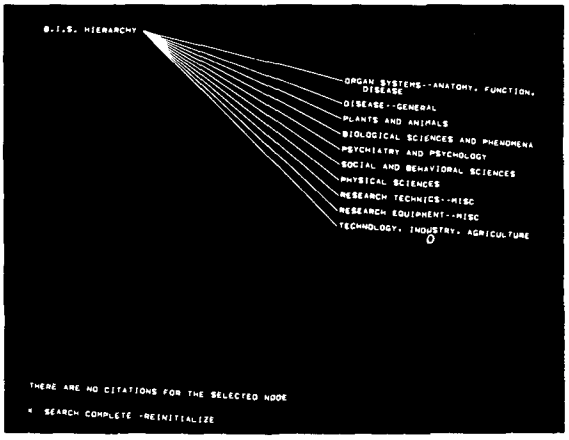
An in-depth study of a user's visual behavior while searching the index tree and making his decisions (Kirschner, 1971) is summarized on Figure 3 and shows, for each displayed branch, the frequency distributions for observing the various alternatives prior to selection of the "correct" one. It should be noted that this activity differs from what we hypothesized would occur, namely, a uniform frequency distribution.

A further analysis of the information content of the displayed branches indicated that the optimal branch width (number of displayed alternatives) was nine in order to maximize the displayed information (measured in hits/second). This compares with the optimal width of 4-5 found in an earlier paper based on a discrete model of the searching process and including more simplistic assumptions (Thompson, 1968). It thus appears that the optimal branch width falls somewhere in the range 7 ± 2 as suggested by Miller (1954).

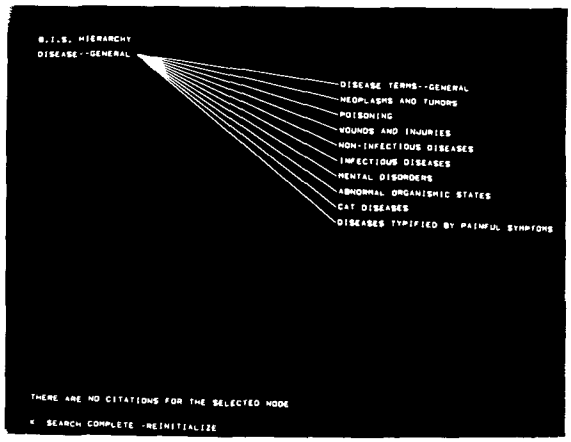
Further modification of this system will permit the direct entry of a keyword into the system as well as

Research support by the National Library of Medicine, U.S. Public Health Service, is gratefully acknowledged.

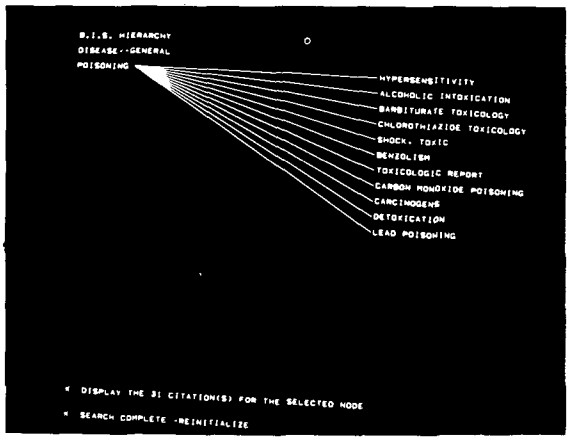
searching the displayed tree branches for it. The latter is more appropriate for the naive user while the former is more appropriate for the sophisticated user. This modification will consequently allow a greater range of users to find the system a comfortable one with which to work (Thompson, 1971A).



(a)



(b)



(c)

Figure 1. Illustration of CRT displays of successive decision tree branches selected during the search of a hierarchically structured indes. When (a) is displayed, DISEASE--GENERAL is selected and becomes the second line on the left and its sub-alternatives are in turn displayed in (b). When POISONING is selected in (b), it joins the left-hand memory trail in (c) and in turn is sub-divided into its sub-alternatives. This process continues until the term selected is sufficiently specific and germane to the query to warrent requesting related documents.

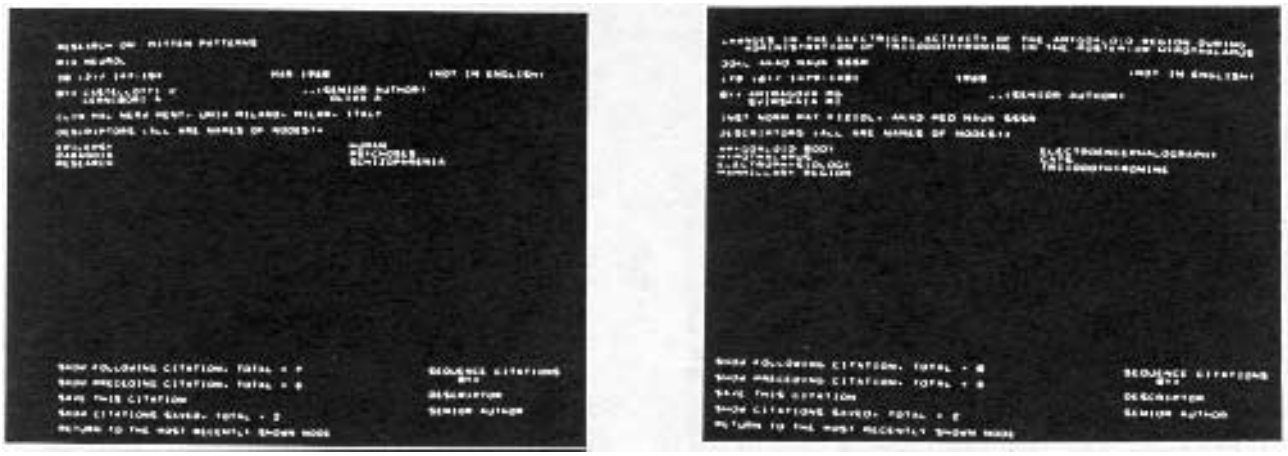


Figure 2. Two examples of document surrogates that would be displayed on the CRT in response to selecting any of the keyword descriptors shown for the document. The instructions at the bottom of each display permit the user to page through other documents indexed by this same keyword, save any for future reference, or return to the most recently selected node in the classification structure (as shown in Figure 2). Provisions is also made in the lower right of each display to sequence other than the one initially chosen, or by the senior author in the event he may have written similar papers.

REFERENCES

Kirschner, Paulo - "A Study of Vision as a Measurement of Decision-Making Behavior," Stanford University, Dept. of Industrial Engineering, Technical Report 71-2, June 1971, 105 pp.

Miller, G.A., "The Magic Number 7 Plus or Minus 2," Psychological Review, (1956).

Thompson, David A. - "A Proposed Structure for Displayed Information to Minimize Search Time Through a Data Base," American Documentation, Vol. 19, No. 1, January 1968.

Thompson, David A. - "The Design of an Interactive Information Storage and Retrieval System for Human Factors Data and Literature: A Program Plan", Dept. of Industrial Engineering, Technical Report TR 71-5, Stanford University, 37 pp., July 1971 (1971A).

Thompson, David A. - "Interface Design for an Interactive Information Retrieval System: A Literature Survey and a Research System Description," Journal of American Society for Information Science, Vol. 22, No. 6, November-December 1971.

Thompson, David A. and H.A. Di Giulio, "A System for Automatic Manipulation and Display of Hierarchical Bibliographic Indexes," accepted by Information Storage and Retrieval for publication in 1972.

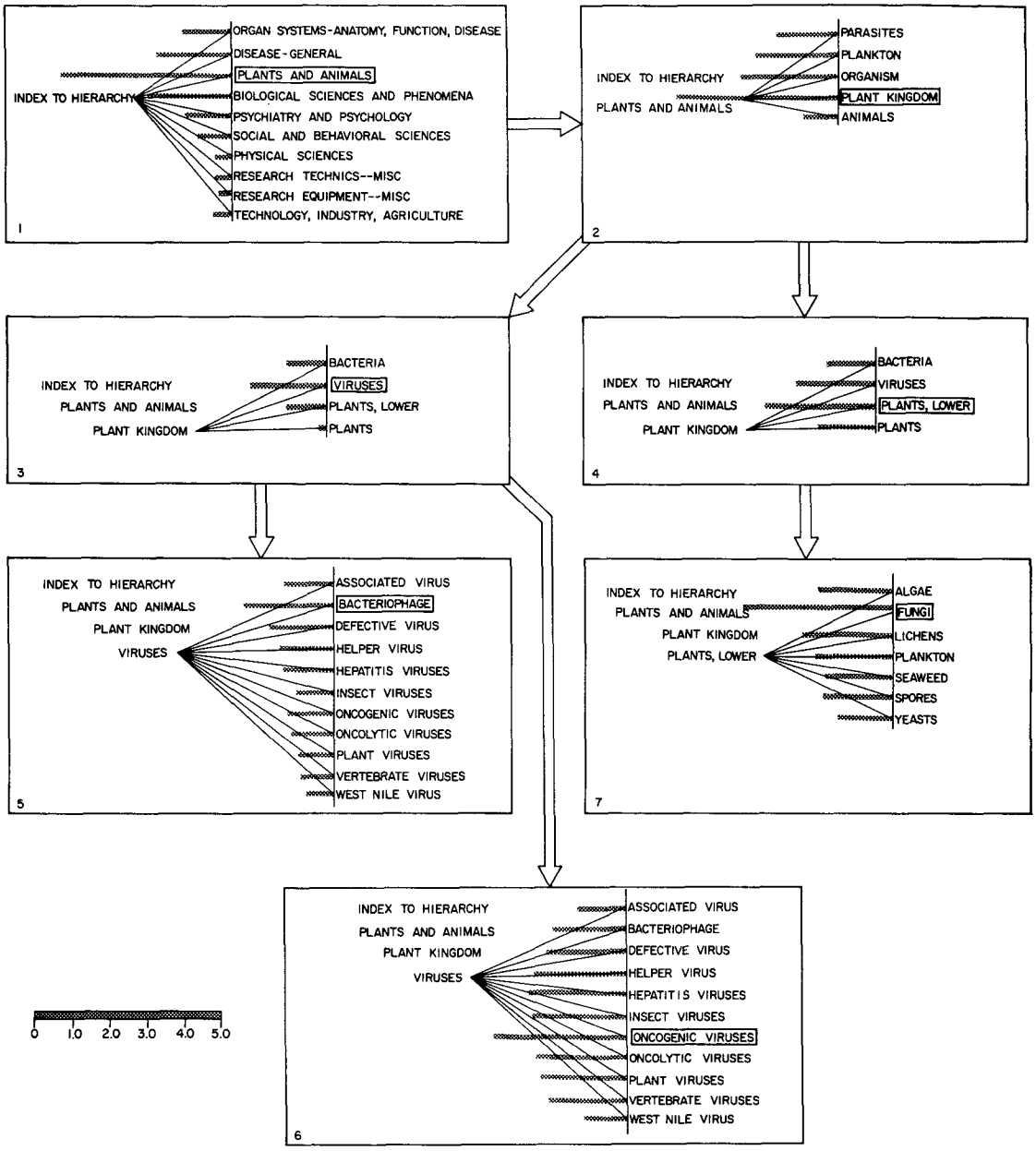

INDEX TO HIERARCHY
ORGAN SYSTEMS-ANATOMY, FUNCTION, DISEASE
DISEASE-GENERAL
PLANTS AND ANIMALS
BIOLOGICAL SCIENCES AND PHENOMENA
PSYCHIATRY AND PSYCHOLOGY
SOCIAL AND BEHAVIORAL SCIENCES
PHYSICAL SCIENCES
RESEARCH TECHNICS--MISC
RESEARCH EQUIPMENT--MISC
TECHNOLOGY, INDUSTRY, AGRICULTURE
1
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PARASITES
PLANKTON
ORGANISM
PLANT KINGDOM
ANIMALS
2
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PLANT KINGDOM
BACTERIA
VIRUSES
PLANTS, LOWER
PLANTS
3
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PLANT KINGDOM
BACTERIA
VIRUSES
PLANTS, LOWER
PLANTS
4
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PLANT KINGDOM
VIRUSES
ASSOCIATED VIRUS
BACTERIOPHAGE
DEFECTIVE VIRUS
HELPER VIRUS
HEPATITIS VIRUSES
INSECT VIRUSES
ONCOGENIC VIRUSES
ONCOLYTIC VIRUSES
PLANT VIRUSES
VERTEBRATE VIRUSES
WEST NILE VIRUS
5
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PLANT KINGDOM
PLANTS, LOWER
ALGAE
FUNGI
LICHENS
PLANKTON
SEAWEED
SPORES
YEASTS
7
INDEX TO HIERARCHY
PLANTS AND ANIMALS
PLANT KINGDOM
VIRUSES
ASSOCIATED VIRUS
BACTERIOPHAGE
DEFECTIVE VIRUS
HELPER VIRUS
HEPATITIS VIRUSES
INSECT VIRUSES
ONCOGENIC VIRUSES
ONCOLYTIC VIRUSES
PLANT VIRUSES
VERTEBRATE VIRUSES
WEST NILE VIRUS
6
0 1.0 2.0 3.0 4.0 5.0

Figure 3