

8. SEARCH STRATEGIES AND SYSTEM DESIGN

In the design of an information system that supports the user's choice of search strategies a database is required with representation of information that matches the user's need and the requirements to his/her performance of the task. Secondly, a computer-aided user-system dialogue is needed with the proper sharing of tasks between the user and the computer, depending on their different resource characteristics in relation to different strategies. (Figure 13).

The bibliographical strategy.

In the bibliographical search strategy with instrumental assistance, the user approaches the information system to search for a known item. The choice of information is made before the information system is entered. The task is to verify the user's information about the known item and locate the item physically, and inform the user about its availability. Assistance in physical search through maps of shelf arrangements in libraries, and in their reservation policies, pricing policies, waiting queues, acquisition policies etc. is part of this task.

The bibliographical strategy with verification routine differs from this strategy, when the user initiates a check of the assumed value of the known item. Verification of its relevance is determined through additional information about its contents. Our empirical analysis of bibliographical strategies show, however, that this information is not sufficient. Users have difficulty in remembering names of authors, titles etc. In some of these cases they remember colours and pictures of the front page of documents, the size or weight of an item, or their memory is related to the source for their discovery of its existence (for instance a review, advertisement etc.). A bibliographical problem domain should therefore be extended with new kinds of information in order to be successful.

The Analytical Strategy.

In the analytical strategy, a systematic flow of information between user and intermediary checks the attributes of the user's need. The search for matching documents is a search in variables of document contents. The success of the

REVIEW OF STRATEGIES

1. Bibliographical Strategy

User: selects and decides author/title, explores retrieved items, compares bibliographical data with need and compares content with bibliographical data.

Computer: assists with identification, verification and location of retrieved items. Communicates information contents for verification of identified items.

2. Analytical Strategy

User: communicates information on need. Accepts and rejects proposed documents.

Computer: explores user's need. Compares need with representation of information contents of documents in the database and selects documents. Displays document contents for user's consideration.

3. Empirical Strategy

User: communicates characteristic demographic, personal characteristics. Accepts and rejects proposed documents.

Computer: explores user characteristics by questions and proposals. Classifies user according to user characteristics and associates to stereotypical sets of items.

4. Search by Analogy

User: communicates information on need through a model document. Accepts and rejects proposed documents.

Computer: explores attributes of contents of model document, compares to document representations in database and selects, suggests documents for user's consideration.

OR:

classifies document and associates to stereotypical set of items in the database.

5. The Browsing Strategy

User: selects and skims subsets of database information. Communicates field of interest and explores potentials of the database.

Computer: scans through information content of subsets of database and compare match with profile of user interest.

6. The Check Routine

User: requires selective information on documents. Compares match with need. Accepts and rejects retrieved documents.

Computer: displays representations of retrieved documents in appropriate formats matching the requirements of strategy.

7. Shift of Strategies

User: accepts or rejects retrieved documents, or postpone decision about match with need. Communicates new information on need or selects and decides known item.

Computer: checks the user's choice of strategy, probes the user's need and compare choice of strategy with need, suggests switch of strategy.

Figure 13

Review of the allocation of roles between user and computer.

analytical strategy will depend greatly on a computerized bibliographical database including documents indexed in terms explicitly characterizing their attributes in relation to the various dimensions and categories of user needs as identified in the AMP classification scheme. A user-system dialogue designed for retrieval of documents according to requests expressed within this framework is also needed. The user formulates his/her needs from experience with a narrow selection of documents. Questions to the intermediary are often not precise, when seen in relation to the real needs. The study of user-intermediary negotiations illustrates how identical formulations can cover different needs due to differences in choice of hierarchical levels for query formulation, in personal experiences, social backgrounds and semantic conceptions etc. A demand for "books with much action" can cover a wish for stories as different as family chronicles and detective stories. A request for a science fiction novel may cover a need for a book about futural technology (subject-matter dimension); an exciting book (emotional experience dimension); a social criticism (cognition dimension); and an easy book (accessibility dimension). The science-fiction class in this four-dimensional example is the class which is limited by the sub-classes along the four dimensions- and by offering suggestions along these four axes it is possible to identify which of the dimensions really represents the user's need - and the intermediary may then be able to satisfy the user with a ghost-story (emotional experience dimension). The efficiency of the classification scheme proposed in this paper as a support to the searchers' analytical strategy is also due to the fact that the classification is not exclusive. This redundancy in the classification can be of great help to the searcher determining what the user desires, since the title of the document may become some sort of cross reference, initiating a search in fields which might not be obvious from the user's original request. If a user asks for a war novel, the intermediary or the computer-system, by asking other questions, will have to make sure whether the user means a book about war, or a polemic book about war crime etc. The supplementary questions must be asked. When the intermediary has placed the information "war novel" in his reference system, he can ask relevant questions by searching dimensions and classes - possibly in the form of suggestions of documents identified by the system. In the same way, the user's information about what he has read with interest can only be useful if it can be placed in the same frame of reference. The intermediary is able to conclude what the user's real needs are only after several questions and answers centered around a number of documents or subjects with some common features. This technique of inquiry is

necessary in order to ascertain whether the user formulates his need in a broader sense than it actually is. We have seen in recorded user-intermediary negotiations that fiction requests often contain clues to the categorial context of the requested features. To search for keyword terms only in a relevant category might therefore speed up retrieval and increase precision. If, for instance, a user wants a novel about war with an emphasis on an analytical attitude towards war, the search for the term war could be restricted to the author's intention category of cognition/information, and retrieval of novels which deal with war from a less relevant viewpoint could be avoided.

Input to the retrieval system is then the specification of user needs in terms of the dimensions and classes of the AMP scheme. Such a retrieval system includes database and intermediary, the latter being a professional intermediary or the computer itself. The two key processes in the user-system dialogue are then :

1. Translation of user needs to the AMP classification search terms. This translation can be shared between user and system in various ways.
2. Comparing user needs with a retrieved set of documents to verify match and to derive new guides for search planning and need translation.

One user-system interface is relevant to users who are familiar with the AMP classification scheme, and who know their needs, and are able to express their needs explicitly. The dialogue in this case is directly performed in search terms along dimensions and categories applied to access points in the retrieval system. An intermediary is not necessary, except for guidance to users in search planning. This guidance is treated separately from the information search process itself, and can be presented to users by a human intermediary or a computer through displays of a "menu" structured in the AMP classification scheme.

A high mental load is involved in users' introspection of their own needs in order to explain the need in terms of attributes related to the organization and contents of documents in the database. The dialogue between searchers and the computer which determines a user's need thus becomes the central field for searchers' resources. It should therefore be considered how the relief of demands in the determination of a user's need can be achieved through computer support (Figure 14).

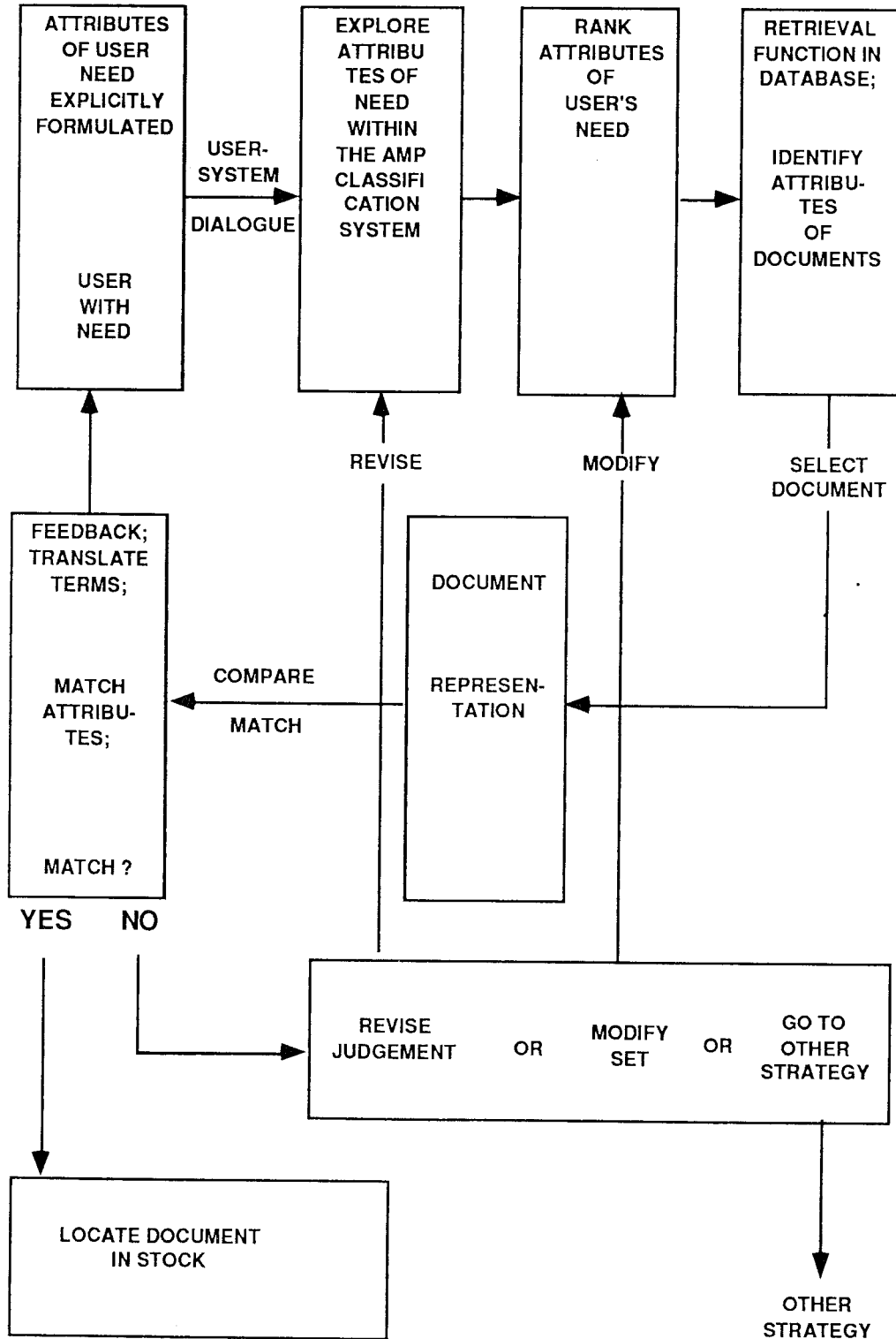


Figure 14: Analytical search strategy

Search by Analogy.

A certain amount of users' inquiries take the form: "I like the book about Robin Hood, have you got something similar"? Or, the user is asked by the intermediary to give an example of a document, which she/he has read, and liked, and which is then used as a model for a subsequent search for documents with analogical features. A user-system interface involving search by analogy is relevant for users unable themselves to analyze and express their needs explicitly, either because of conceptual and linguistic incompatibility, or because of the vague and unspecific nature of the need, which cannot be specified more precisely than, for example, "something similar" to previous reading. This search by analogy usually occurs in conjunction with the empirical search strategy, where the intermediary associates from the model book to a repertoire of titles. It rarely occurs with the analytical strategy as suggested here. In this situation, the risk that a search may switch to empirical strategies based on stereotypes should be considered in the design phase. To determine the user's need, the intermediary must discuss the model book with the user and try to identify specific features which the user found interesting. If the novel that serves as a "model" is already in the database, its details can be displayed on the screen and the terms assigned to it can then be discussed with the user. Keywords for a subsequent boolean search can then be selected directly from the record of the novel already known by the user. Documents retrieved in this way will share some features with the "model", but they will also differ from it and from each other in possessing additional features which may enable the user to select some and reject others, or they may suggest ideas for new combinations of features. One solution may be the use of numerical weights which the system automatically add to the indexed features of the user's example, and then calculate the value of documents in the rest of the database in comparison with the value of the example. Documents similar to the example are then displayed automatically to the user, highest value first, lowest last. The user's accept/reject to displayed documents can then be used by the system to adapt its automatical weighting to the user's preferences, and a new search can be performed. The system must calculate the value of the individual feature of each dimension of each document in comparison with the attributes of the model document in order to make inferences about the user's need from a comparison between a model document and the user's rejectance or acceptance of other documents with a combination of

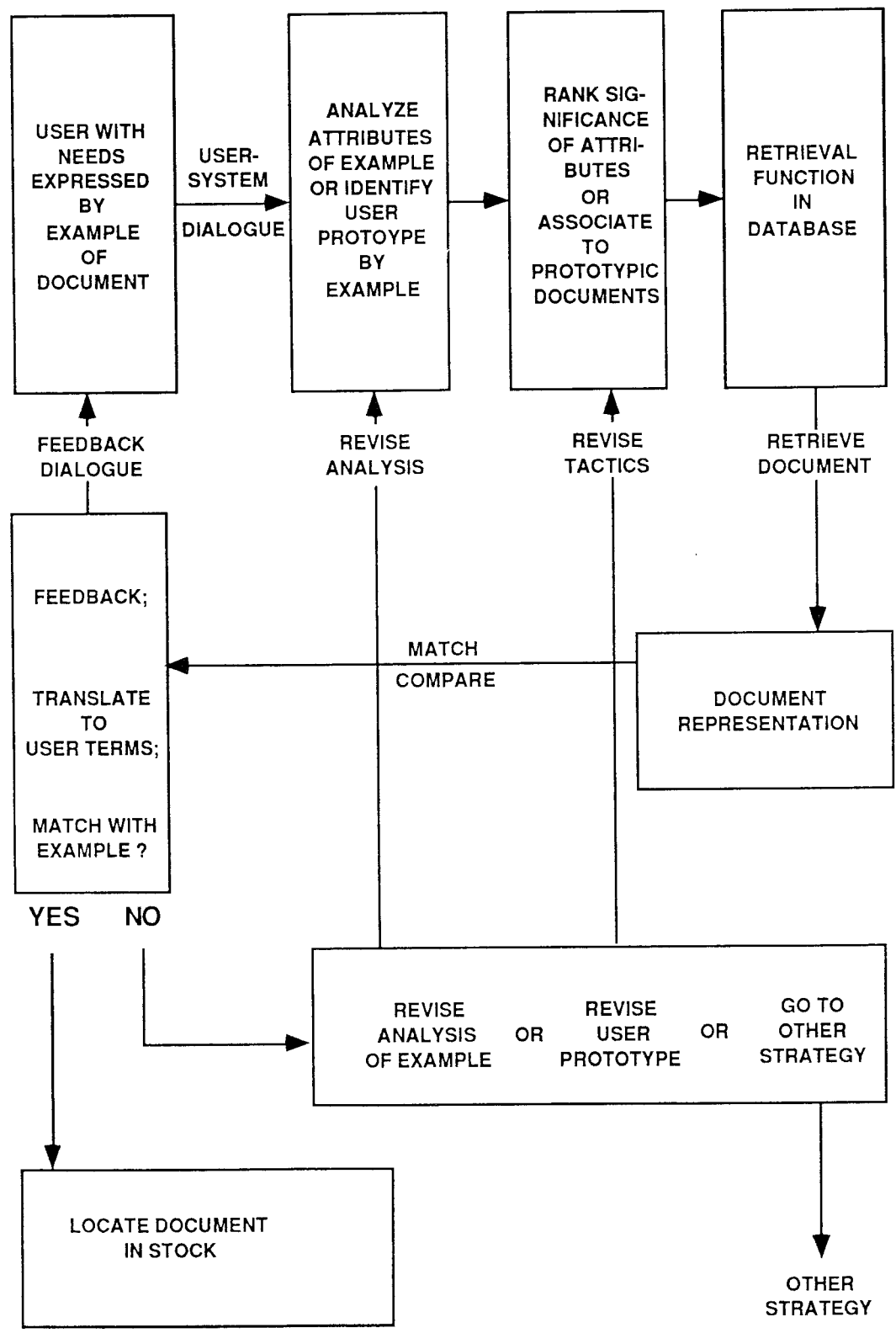


Figure 15. Search by analogy performed by means of analytical or empirical strategy.

similar and dissimilar features. The success of these ways of automatizing the analytical user-system dialogue for determination of a user's need will depend on the number of document examples that are needed by the system to reach an acceptable formulation of the need seen from the user's point of view. A high number of rules are necessary, not only in relation to criteria for display of document, but also in relation to the boolean retrieval function of the database. Figure 15.

The Empirical Strategy.

Prototypical models are derived from experience with characteristic features of the needs of groups of users that share one or more common characteristics; these models are used as a basis for classification of users, their questions and their needs. The search for documents corresponding to a model of prototypical needs consists of associations from one category of user needs to a matching category of documents that has one or more prototypical features in common. Where the analytical strategy is multi-dimensional, the prototypical search is a one-dimensional search in an association network. Generalized concepts of sets of needs and sets of documents that synthesize collective similarities instead of individual differences are used in this search approach. The following is a discussion of this particular strategy as one design aspect of a retrieval system that matches and aids users' choice of strategies. Computer support is especially relevant for users with a vague and unspecific nature of need, who have difficulty in an explicit formulation of needs and with the identification of document aspects that might satisfy their needs.

A computer can take over this task by identification of their needs within a model of standardized categories of need-document matches, established before the user enters the computer. During the retrieval process, the computer should be able to adapt these standardized concepts to the individual user's need.

The implementation of the empirical strategy in the design of bibliographic databases involves a representation of contents by attributes related to well known categories of user needs. Two key processes are needed.

1. Identify user prototype. Input to the retrieval system are then specifications of the users' characteristics. This identification can be shared between user and system in various ways.

a) Users access the computer directly and are requested by the computer to give information about themselves. This user-system interface may only request users to describe themselves to the system. Based on tested evidence of demographic features built into the computer, documents associated with these input characteristics will be retrieved. The computer should possess a reasonable number of person characteristics along different dimensions, such as sex, age, profession, etc. The judgement of the relevance of personal features for the determination of need is left to the user.

b) Users access the computer directly and are requested to answer questions posed by the computer with regard to information about user characteristics. This can be done by interrogating users along dimensions of user characteristics, or by displaying options of features that the computer would be able to recognize and among which users can choose. This restricts the user to the vocabulary and prototype known by the computer. Therefore, the computer should be able to revise and widen the prototype according to experience with users of the system.

c) Identification of user prototypes can be carried out in a quick way by displaying the categories of combined user-characteristics and matching document categories known by the computer to end users and let the user recognize his/her membership of groups through a pattern recognition approach.

It is necessary to employ uniform criteria in the characterization of both categories of users' needs and descriptions of contents of documents. Hence, the system can evaluate users and document aspects on the same level and in the same terms and dimensions. The analysis of documents will be based on uniform principles, and each search will, dependent on the activated user prototype, involve one or more dimensions of document contents in order to retrieve the associated set of document. In this way, the computer with time can build up the empirical link between already empirically recognized categories of needs, and the characteristics of the users of the database. After a certain number of searches the system should automatically deduce patterns from the individual data and constitute prototypes by the use of straight-forward pattern classification techniques. Special issues arise with the empirical strategy, such as ethical questions of privacy. Documents must be retrievable in relation to

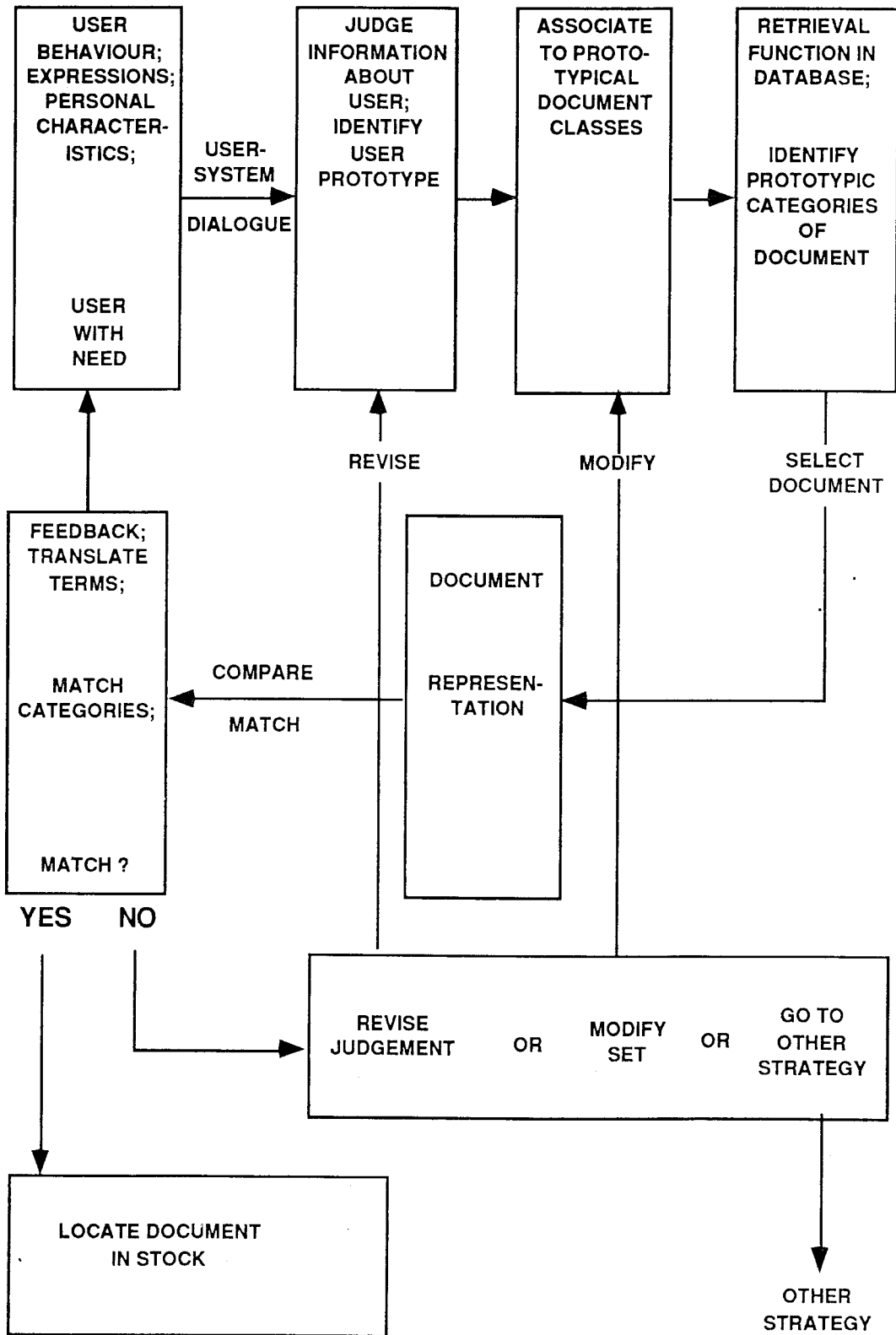


Figure 16. Empirical search strategy

information about users, which they will accept to provide to the information system. The needed information must be easily available in users' memory and it must only be accessible through the user's password.

2. Comparison between user prototype and representation of document contents. A check of match between category of need and representation of contents in the selected category of documents involves a process of feedback to the user, where a translation may be appropriate between a) the criteria used for choice and representation of documents and b) characteristics that constitute the prototypical category of need. This may result in a confirmation of the retrieval system's inferences about the user's need, or in a revision of judgement of a user's membership of the assumed category of users.

A re-evaluation of the user prototype can lead to a shift in weight of criterion of the actual information about the user. Or a search for more information about the user can be initiated. If a user is an atypical member of a category, a switch to another strategy may be advisable. Figure 16.

The Browsing Strategy.

'When the user chooses a browsing strategy, an information scanning method is used where little is formally planned. Many different intentions may lead to this choice, which usually demands few mental resources, but is a time consuming strategy. If the need is already known, a choice of a browsing strategy may be the result of vaguely formulated ideas of what features a document should possess that could satisfy a need. It may be a result of a certain openness to revision of anticipated needs. When a need is known, it would be expected that users would pursue the most effective methods of utilizing retrieval tools; most effective in terms of getting fast document output corresponding to terms of the search question. When a less effective strategy is chosen, other values than a fast solution must have a high priority. Values such as enjoying the scanning process itself or its "side effects", like the enlightenment gained from a variety of features of literature, an improved comprehension of how the literature is organized, and how a need can most appropriately be transformed into a search question or another search strategy. Values of the browsing strategy are its similarity with a learning process, its easiness and its few constraints. The browsing approach can be a separate strategy, but is

also likely to occur as a sequence before and during other search strategies. A bibliographical database can be used effectively for browsing if the documents are classified according to the AMP scheme. In order to browse systematically and avoid a "blind" search, a kind of hierarchical structure of document/database collections is necessary. The following information processes are relevant for any browsing purpose:

1. In the user-system dialogue the system displays systematically selections of database content like text passages, representations of documents, classification schemes, authors, titles, thesaurus terms etc. The aim of the dialogue is to provoke associations and ideas, something of immediate interest that catches the user's attention and rouses his/her memory. Scanning through information can also be a means to get a quick view of what is available within a given category of information.
2. Comparison between contents and the need, implicitly or explicitly stated. In the above mentioned instances, a feed-back mechanism can be in function during the scanning of information. In spite of the various values of the browsing strategy, the overall goal of the search is a match of need-document.

The AMP scheme provides the structure for a "menu", from which users can select browsing options at any level, such as: dimensions, bibliographical data, list of subject terms, etc. The users' selection of where to browse may be aided by a review/birds-eye-view of the system at various levels, and in terms of descriptions compatible with possible interests. The structure and organization of the actual information in the database must be visible to the user .

Feedback Routine.

A feed-back routine may involve a translation of the user's need and the document representation into a set of terms that can be compared for match between the user's need and the retrieved documents. However, the route of the feed-back routine when using a specific strategy depends on the availability of need relevant information in the database to carry out the process of matching variables. It also depends on the various kinds of input to the retrieval system about the user's need. As described

previously, this input may vary from need formulations in bibliographical terms, in terms of the AMP classification, in terms of the user's yes/no answers to displayed examples of documents and input in terms of user characteristics. The information flow of the feed back routine will have to vary to the same extent as does the user's input to the system. But two main activities will occur independently of choice of strategy.

Dialogue case A: The intermediary can compare match, when the user is unable to make the comparison. The intermediary assists in the translation of the user's need into terms of comparison, and if necessary, the translation of document representation into terms of comparison. This dialogue has features similar with the probation of the user's need in the user-system dialogue of the analytical strategy.

Dialogue case B: The user performs the comparison of match between need and document representation. This is generally the case. The level of comparison within users may not be known. The user will select features from the representation of retrieved documents as a basis for the judgement.

If the feed-back dialogue does not succeed or if a match is not achieved, a reformulation of the user's need is needed, eventually through a dialogue similar to the initial dialogue of a chosen strategy with the system, or shift of strategy.

Shift of strategy and computer aid

When designing an intelligent interface based on users' preference of search strategies, their shifts between strategies become an important issue. A system should be flexible enough to identify the appropriate situations for suggestions of shifts to the user as well as the user's own attempt to switch to another strategy. Shifts between strategies take place for two main reasons: Either because searchers associate to other needs and get new ideas, or when problems occur because of lack of match between resources and the requirements of the chosen strategy. This lack of resources refer to:

1. need formulation,
2. retrieval of documents for consideration,
3. determination of match between need and retrieved documents.

The following cues for shifts of strategies may be useful as guidelines to the user about the strategy to be pursued. They may be used as "triggers" to use, when the search within a given strategy does not advance any longer. Humans make intuitive judgements of cues based on social and professional skills, which indicate to them when search strategies are no longer sufficient. Such cues are therefore generally difficult to define and even more difficult to give the formal description needed for a computer design task. Nevertheless, some ideas of cues for shift of strategies are given.

Shift from Bibliographical Strategy.

When a title selected by users is identified, but is inaccessible (out of print, on loan, acquisition policy, etc.) or a selected title is unidentifiable (not included in retrieval tools, wrong data, etc.). Switch to the analytical strategy, where the need is explored. It is checked whether this title is a unique wish, or is symptomatic of a need, and therefore can be substituted by other documents. Or, when a title selected by a user does not represent a sufficient number of documents, and the user is short of other titles. Switch to the analytical search strategy, where the need is explored as a basis for retrieval of more documents. When a title selected by a user is retrieved and has satisfied one of the user's needs. Check whether the user has other needs. Switch to the analytical strategy, where other needs are explored, and matching documents are retrieved, or switch to empirical strategy to identify category of user and need.

Shift from Empirical Strategy.

When rejections or ambiguous reactions to retrieved documents occur and a discrepancy between a user's need and documents retrieved is possible. Switch to the analytical search strategy, where variables of users' needs are probed or switch to the bibliographical search strategy, where the user takes over the task of choosing

documents. Or switch to the check-routine where the correctness of the assessment of need category match with document category is confirmed or disapproved of, and the search criterion is maintained or sustained. Or when the user has another kind of need besides the need satisfied with the retrieved documents. Switch to the bibliographical or the analytical strategy.

Shift from Analytical Strategy.

When the information flow between user and information system is unsuccessful or cannot be activated, neither through an interview initiated by the computer nor through a need formulation in terms known by the system. Switch to the empirical strategy, where prototypical models of users' needs substitute analysis of variables of need. Or when retrieved documents are rejected. Uncertainty of match between variables of needs and variables of document contents. Switch to check routine to search for an approval of search criterion.

Shift from Check Routine.

When the user's need has been checked and the search criterion has proved unsuccessful, or if the user confirms the search criterion used by the system. In both cases switch from the check routine, either back to the previous strategy, if it has proved successful, or to a new strategy, browsing, analytical, empirical or bibliographical dependent among other factors of the specificity of the user's need.

Conclusion

Empirical investigations of the current work in a situation without computer support cannot be reflected automatically in the design of a computer based support system. But the results can give important principles for the designer.

The above discussion illustrates a major finding, namely how users effectively develop a number of strategies and fluently shift among these, when appropriate for a quick solution of a problem, and thus adapt with great skill to the nature of their surroundings. This experience was used in the design of the interface to the SPRING Database. Books should not only be accessible (and displayed) from many different dimensions, but users should also be able to take different routes/strategies to the

books in the database, and, importantly, to shift among these routes at any time during a search. The flexibility in the functionality of the BOOK HOUSE was defined and designed along these principles.

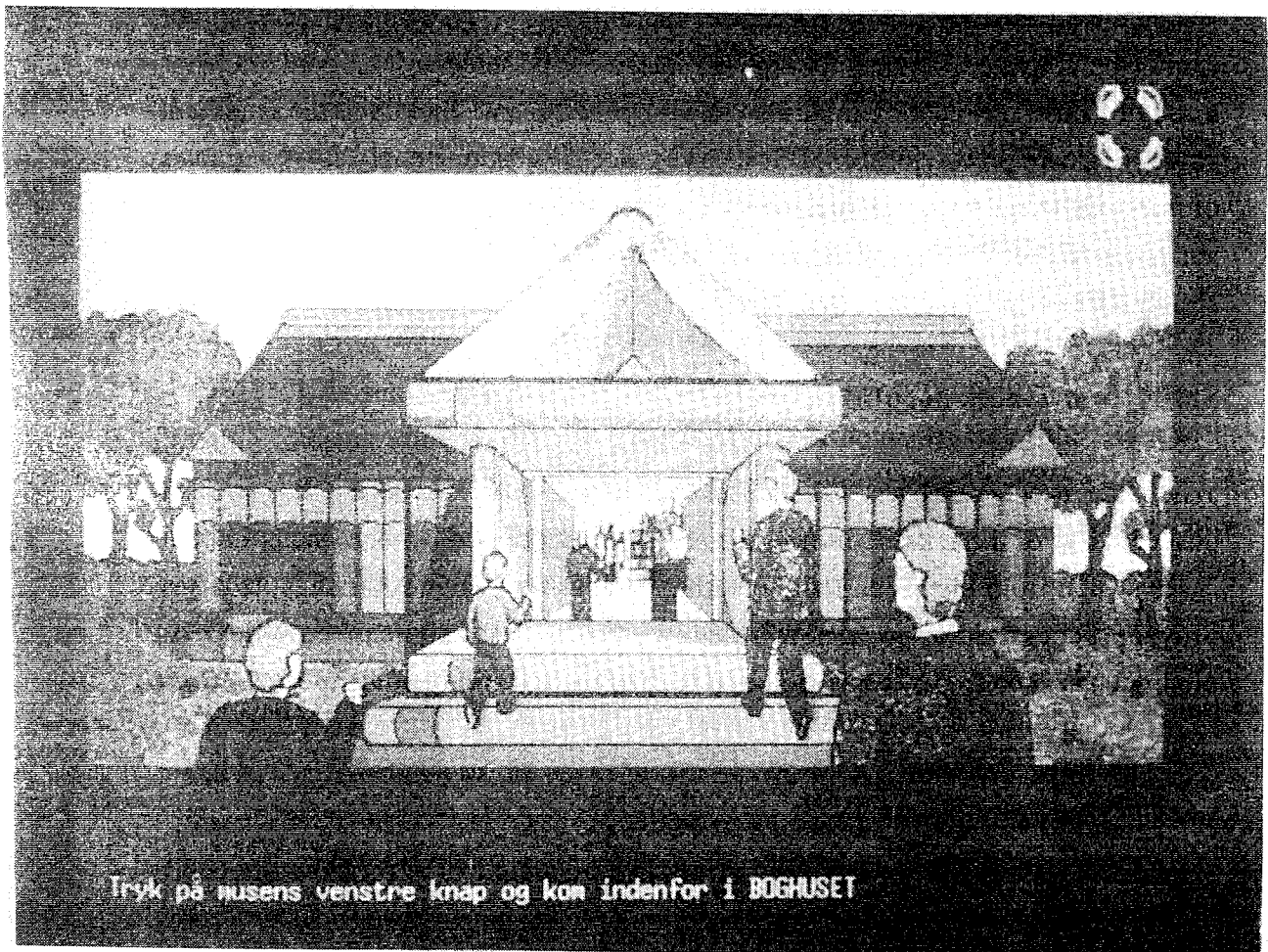


FIGURE 17

THE BOOKHOUSE: ACCES THROUGH THE OPEN DOORS WITH A CLICK ON THE MOUSE.

9. DESIGN OF THE BOOK HOUSE

The above described investigations and the derived models of searchers' problem domain have been the conceptual foundation for the development of a prototype system for retrieval of novels in public libraries. The AMP classification system was used to design the SPRING Database. The structure of this database was, in turn, used in the interface to support the user's navigation in the system. Specification for design of the interface will be described shortly in the following. For more detailed information see Goodstein and Pejtersen, 1989, Agger and Jensen, 1989. Below is a short description of the BOOK HOUSE followed by a summary of the evaluation results of relevance for the investigations described in this report.

The BOOK HOUSE metaphor

A primary postulate behind the BOOK HOUSE project is that efficient navigation in the SPRING data base by a casual user will be greatly enhanced if the data base structure can be embedded in a context which can be represented directly on the interface display surface. This will make it possible for the user to quickly develop efficient skills in communicating through the interface without the need for complex mental juggling to convert one representation to another. One way to attempt this is to consistently recode the abstract attributes of the items in the data base to positions in various spatial arrangements on the display surface. This will directly reinforce the development of manual skills involving the mouse to cope with the resulting spatial-temporal working space in what one today calls a direct manipulation mode - i.e. an efficient automated interaction with the entities on the visible surface.

This approach gives rise to an integrated spatial metaphoric representation for the BOOK HOUSE which includes (a) The SPRING data base contents and structure as well as (b) support for the associated information processing and retrieval strategies. It can be seen that this indeed is a powerful and potentially universal metaphor because it can accommodate changing to other databases having a different structure and/or contents through appropriate alterations in the spatial conversion while the basic facilities for supporting the search process itself remain essentially the same. This stands in contrast to the widely used *Desk Top Metaphor* which is nothing more than an arrangement on the display screen of a variety of tools (files,

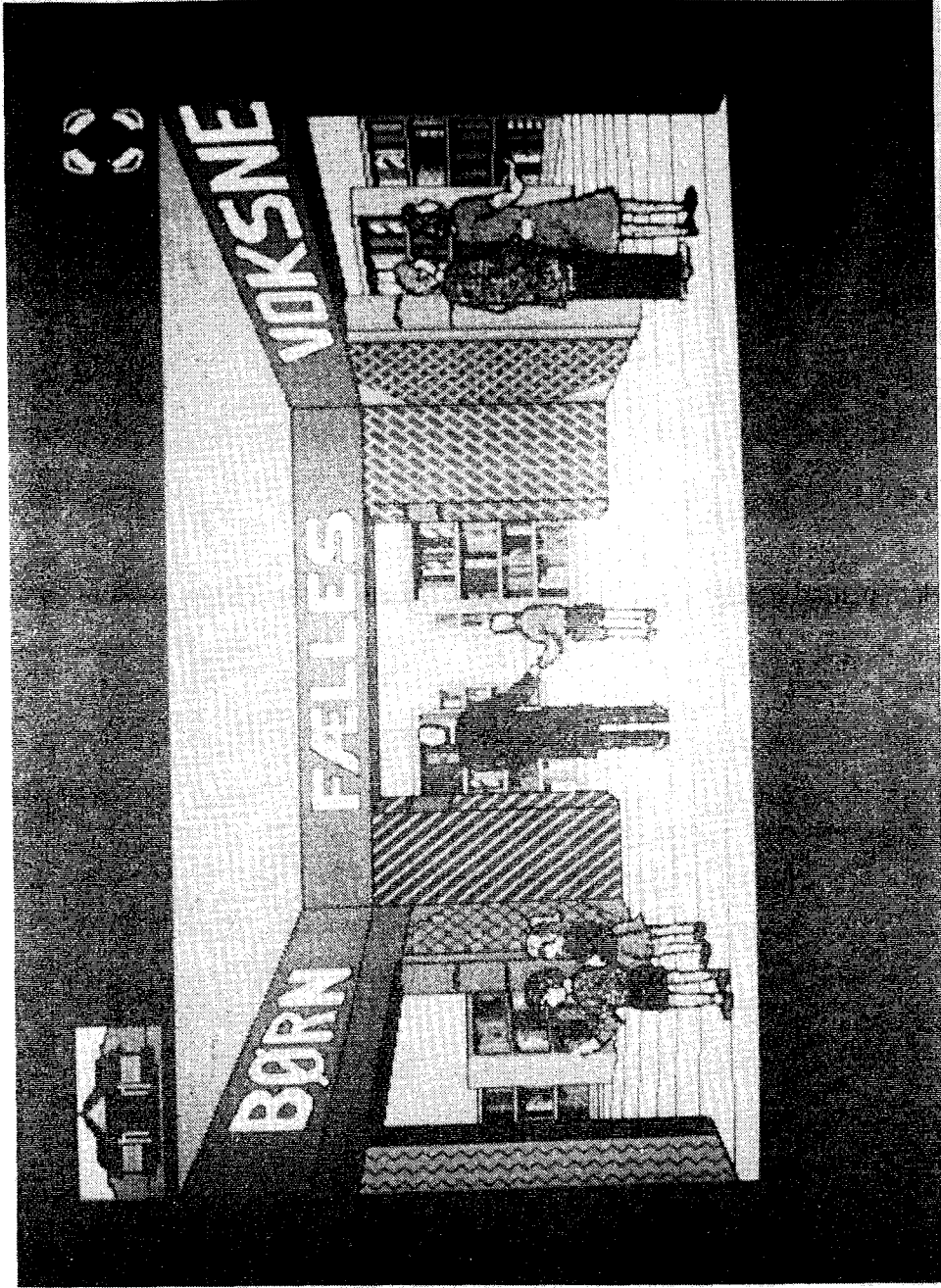


FIGURE 18

CHOISE OF DATABASE: BOOKS FOR CHILDREN, FOR ADULTS AND CHILDREN AND FOR ADULTS.

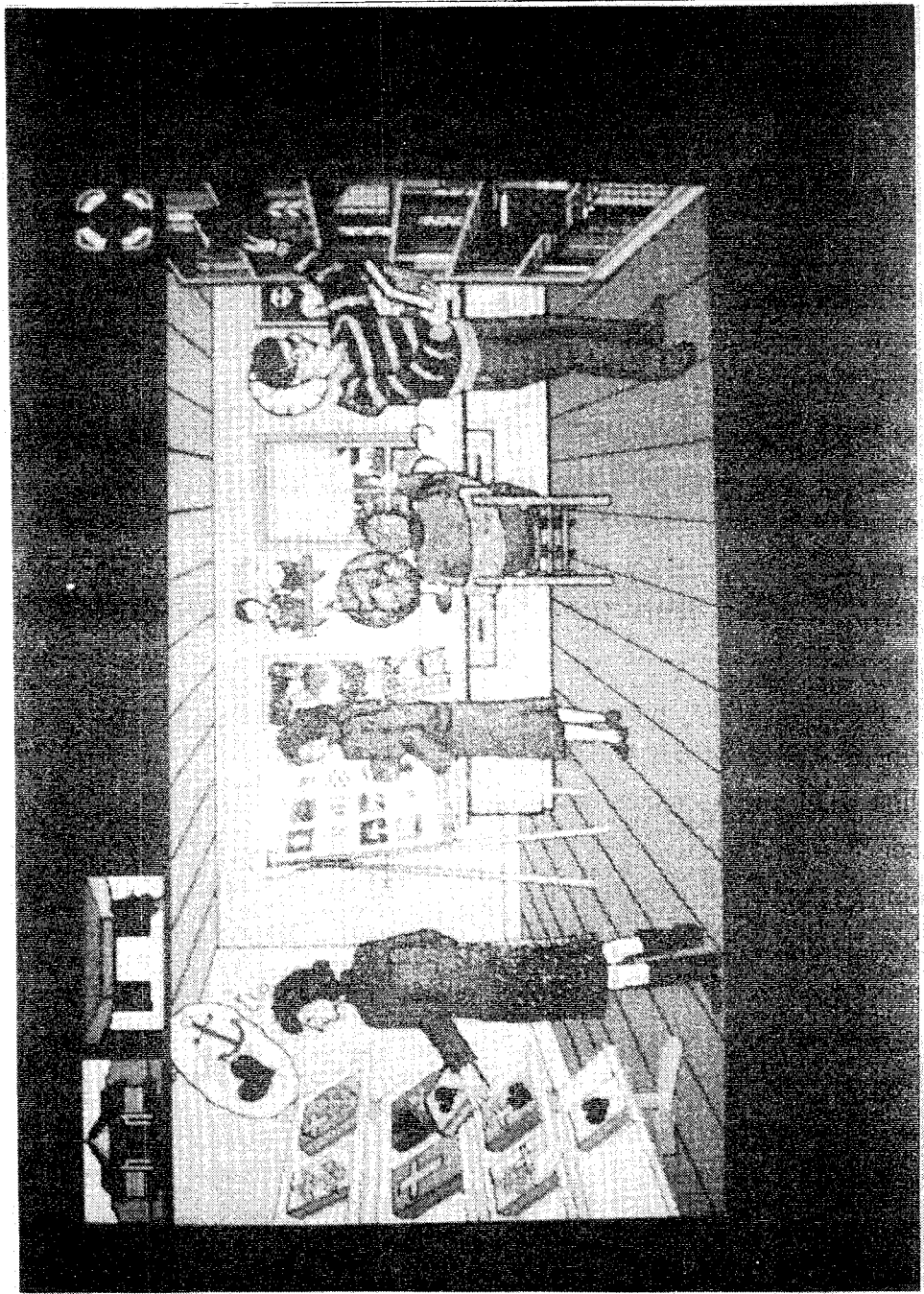


FIGURE 19

DISPLAY OF THE STRATEGY SELECTION ROOM FOR ADULTS.

documents, folders, trash baskets,..) mainly depicted in terms of conventional technology without any direct support of the work situations the user could be confronted with.

Navigation in the SPRING database

Of course, in a real-life library - as for most data bases - the number of dimensions in the base can be numerous, the number of stored items is high and the search attributes are not always well known to the user. In the BOOK HOUSE intermediary, an attempt was made to exploit the flexible display capabilities of computers to relate both information in and about the data base as well as the various means for communicating with the data base to a location in a virtual space - i.e. by creating a kind of storehouse (or book house). The many dimensions were allocated locations in appropriate rooms or sections of rooms within the storehouse. This type of metaphor was felt to be easily understood and remembered - by both children and adults. This is exemplified by the poster introducing the BOOK HOUSE (Figure 17) which is a house made of books (like a pancake restaurant). The book is the reason for, the object of and the target for the acts of users during a search. The house is the symbol of the database system in which the user has to navigate.

Thus when the user has entered the BOOK HOUSE, he/she first enters a hall connecting to three different rooms having books on shelves (Figure 18). The left room has books for children, the right room has books for adults while the center room has books for both. The difference is reinforced by the size of the people seen entering the three rooms.

After the user has chosen a database, a new room will appear showing people (children, adults or both - depending on the previous choice) busy searching for books in different ways (Figure 19). The user can thus select one of four different search strategies based on the results of the empirical studies mentioned previously. The direct manipulation feature allows the user to click directly on the figure executing the same strategy that he/she is interested in. A selection here leads the user to a new area where the required set of "tools" for carrying out the chosen search strategy are made available. Each of the four strategies will be described briefly below in terms of the metaphor.

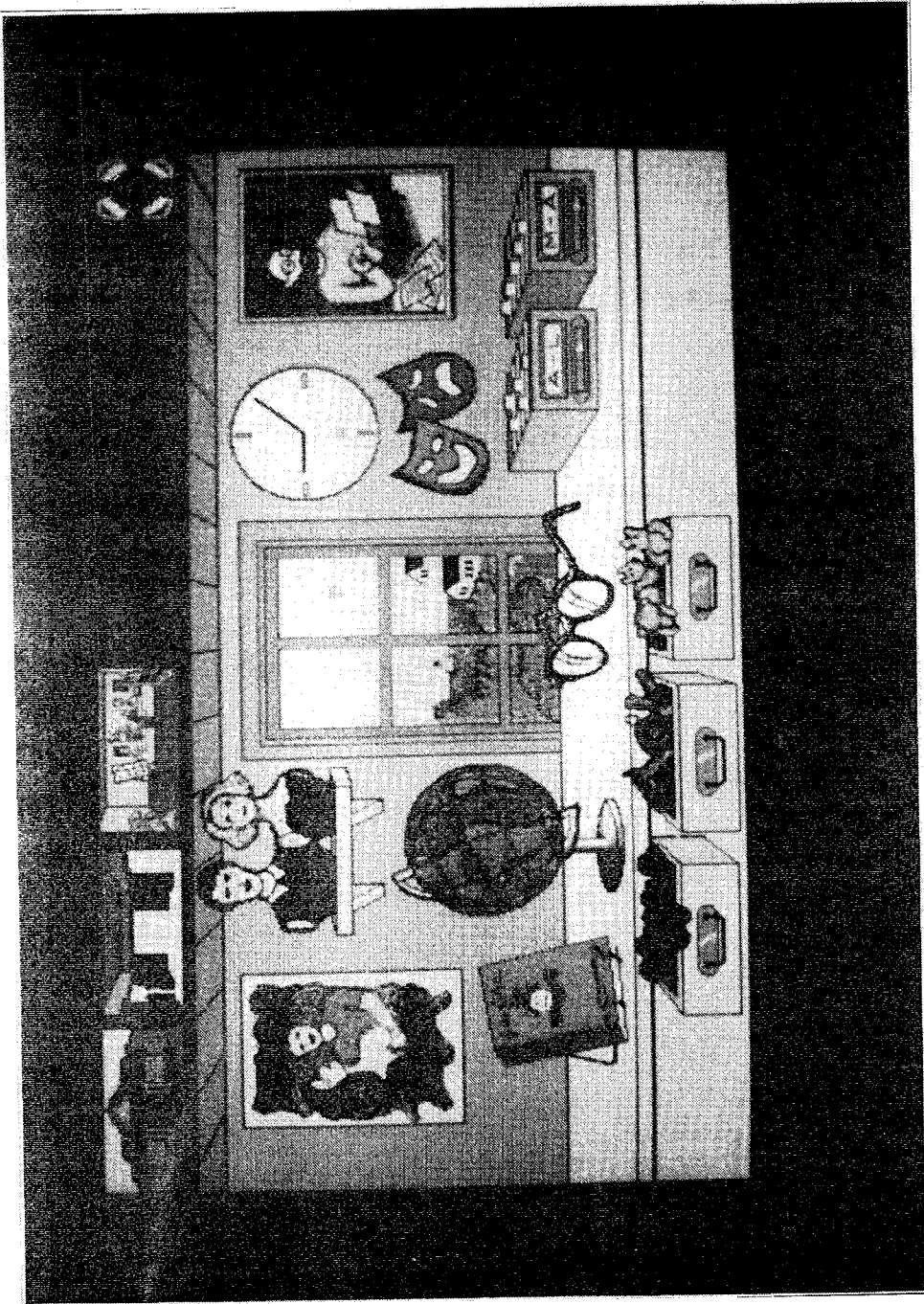


FIGURE 20
THE WORKING ROOM FOR SEARCH FOR SUBJECT TERMS (ANALYTICAL SEARCH).

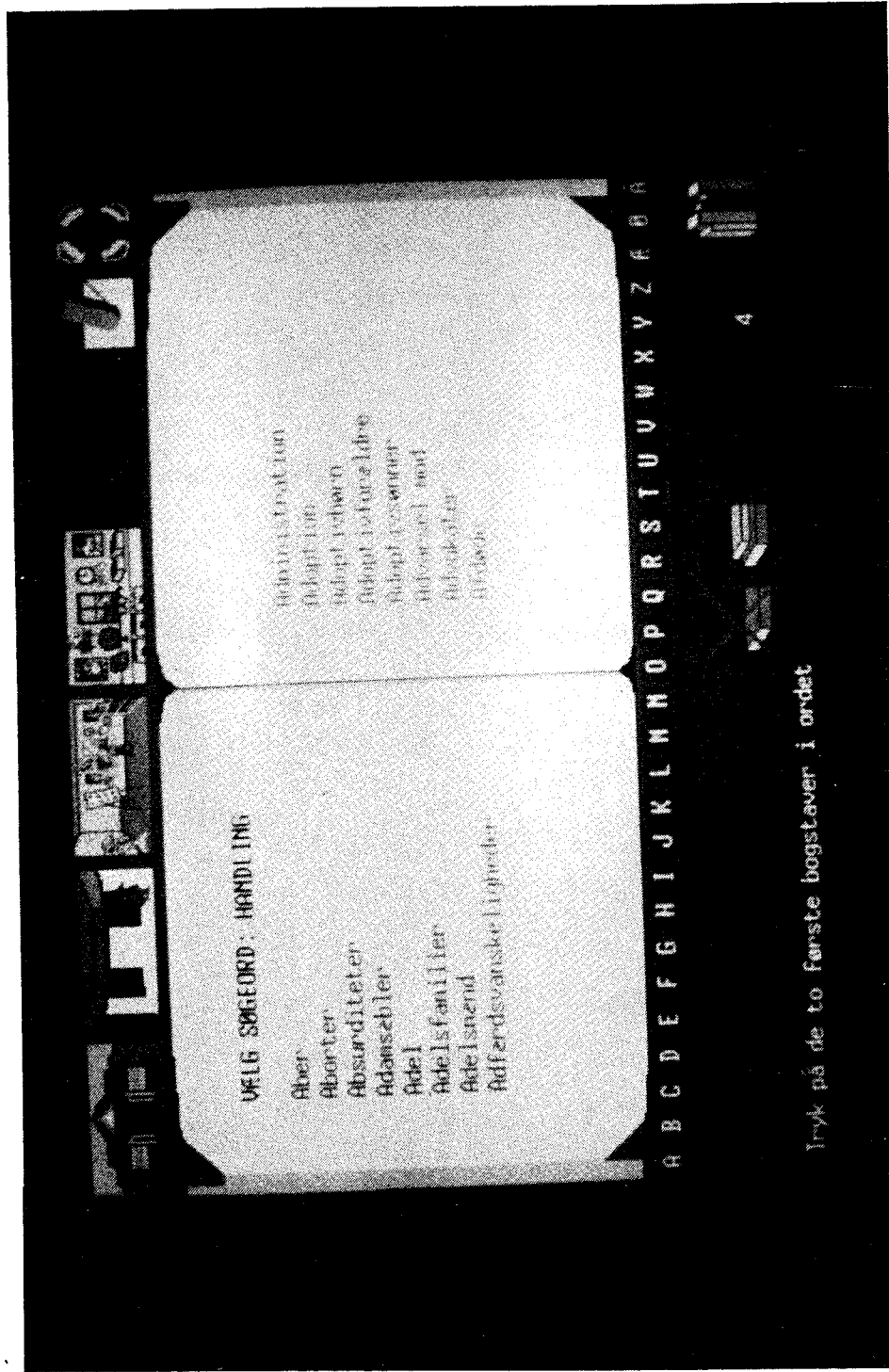


FIGURE 21
AN OPEN BOOK WITH A LIST OF KEYWORDS

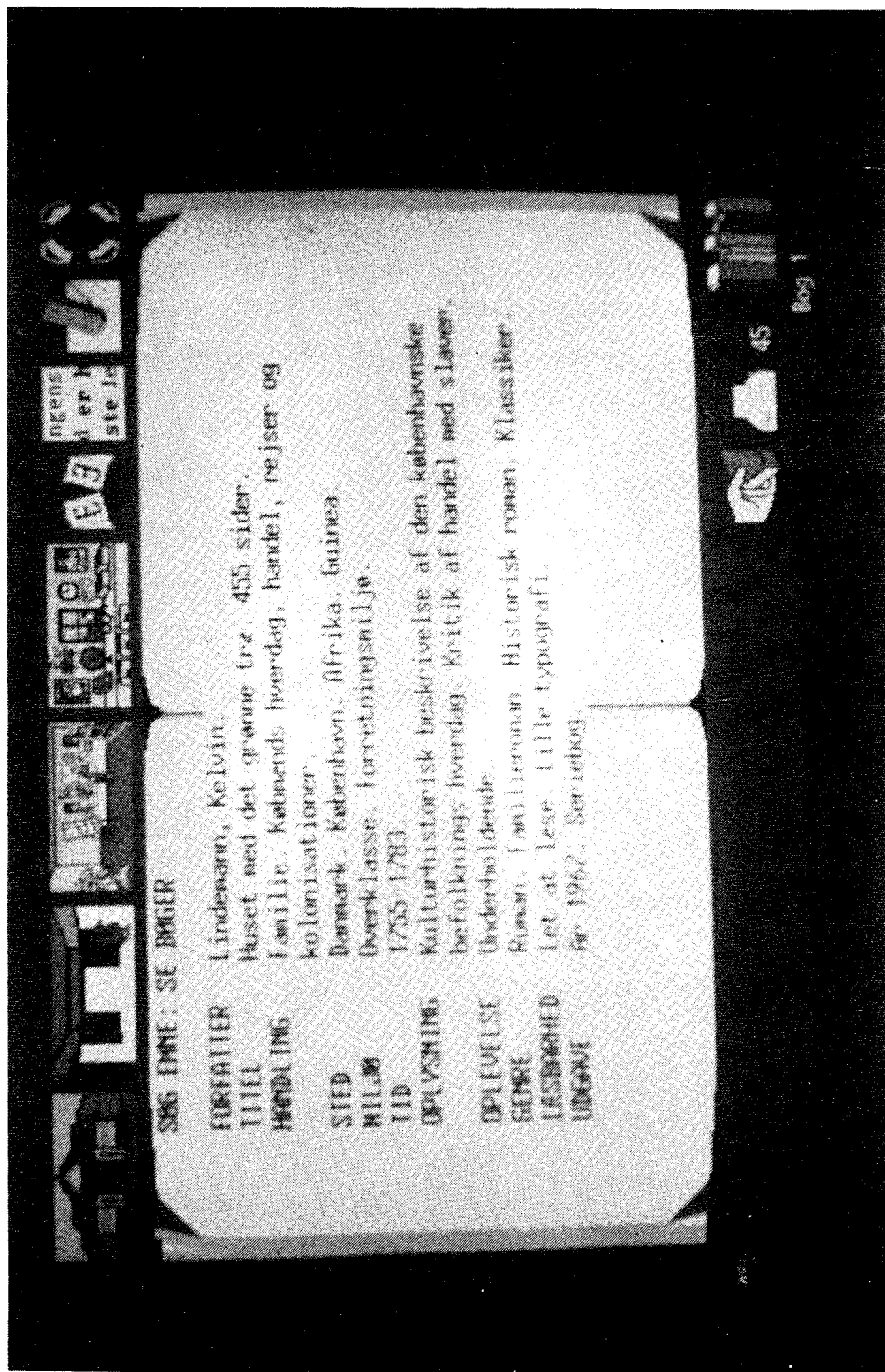


FIGURE 22
AN OPEN BOOK WITH A DESCRIPTION OF A RETRIEVED DOCUMENT

Analytical strategy

Figure 20 is one of four work rooms and is designed to support an analytic search. The user enters this room and takes the place of the figure sitting at the table in the previous picture. On and around the table are icons representing the different dimensions of the AMP classification system. There are thirteen of them. For example, the world globe represents the geographical setting of the book, the clock icon denotes the time dimension and the theatre mask refers to the emotional experience provided by the books. The user can select one of these at a time so as to get access to an open book display with the textual listing of the particular set of search terms which belong to the selected dimension (figure 21). The user can also combine search terms from the same or different dimensions. The Boolean operators are automatically inserted - generally with AND between dimensions and OR and AND within dimensions. This set of selected terms is the user's current search profile which he/she can revise by deleting terms, adding terms or make a NOT operation. Thereafter the user can see descriptions of the books which the BOOK HOUSE has found. These appear in an open book format (Figure 22) and are structured according to the classification system so that the user can directly see the correspondance between his/her selected terms and the actual contents of each book.

Search by analogy

If find a similar book is selected, another room is entered which shows a book containing a title or author index for the database (Figure 23). This can be opened to help the user identify his/her reference book. After selection of this reference book, the system will automatically attempt to find other books in the collection which include as many of the same indexed attributes of the reference book as possible. The descriptions of the ten most similar books will be presented one at a time in an open book format in decreasing order of relevance. The multifaceted classification scheme is used as the basis for a weighting of all books in the collection with respect to the reference book. The algorithm is quite simple and can be described as follows: If, for a given dimension of the classification system (e.g. plot or time), c is the number of index terms common to a given book i and the reference book which itself has a index terms within this dimension, the given book's weight in this dimension is

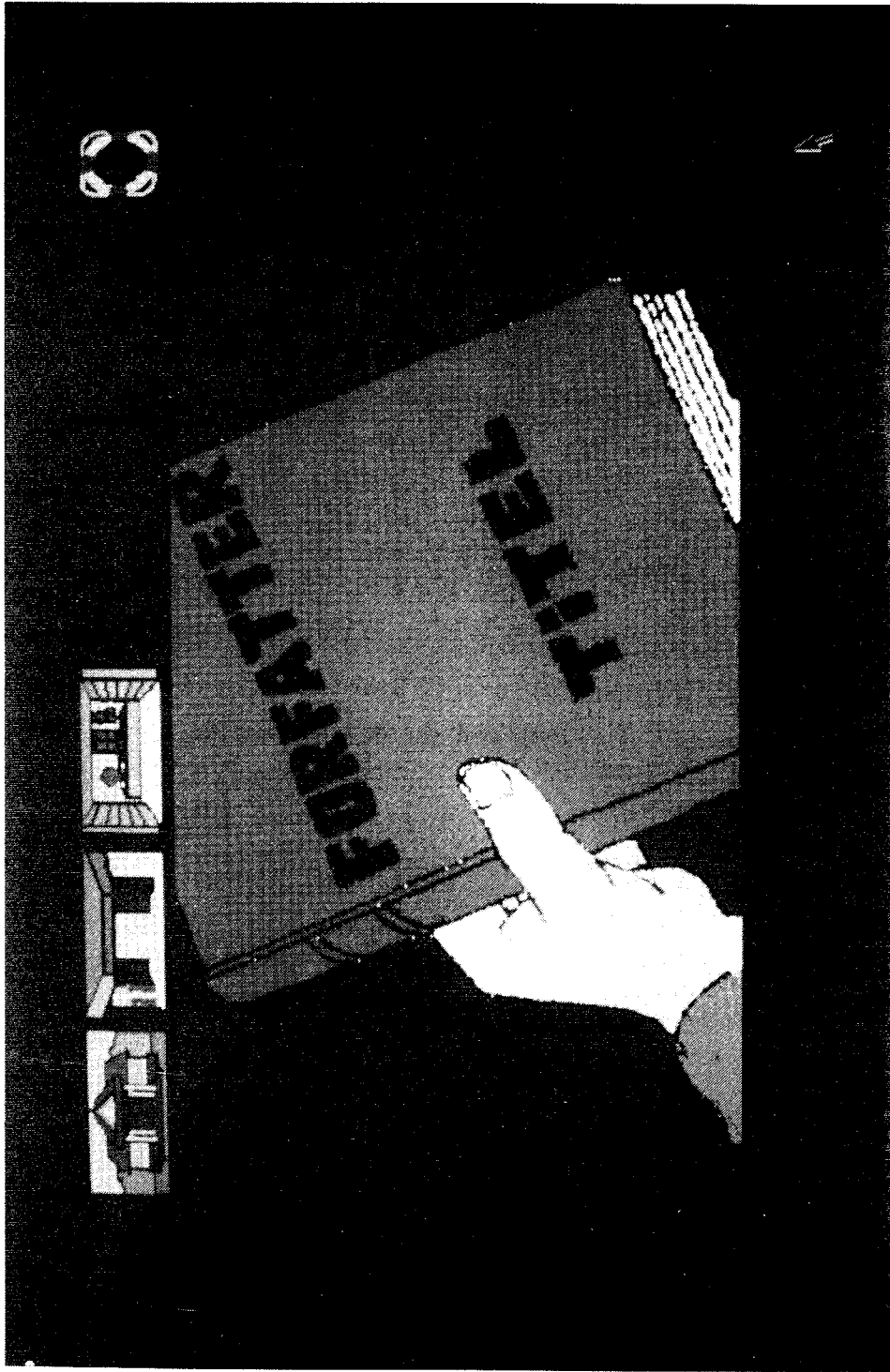


FIGURE 23
THE REFERENCE BOOK WITH AUTHOR/TITLE INDEX FOR SEARCH BY ANALOGY

$$w_{di} = w_d \times c/a$$

where w_{di} is the weight of dimension d for book i and w_d is the collection weight for dimension d . A book's total weight is then the sum of its dimension weights. In the BOOK HOUSE, seven of the thirteen dimensions (plot, setting, place, time and impression, cognition, genre) were utilized in this calculation of similarity.

The system calculates collection similarity on-line after each request for "find similar book".

A choice of a browsing strategy indicates that the user doesn't know the "specific address" of a good book but would prefer to "wander around town" until a good/familiar/interesting item is discovered. Thus after "clicking" on this strategy, the system shifts immediately to an open book representation with a randomly chosen book description. The user can thereafter continue to step through other descriptions.

Browsing strategy

The browsing strategy also includes an iconic version where the user can browse through small pictures describing book content (Figure 24). This could be appropriate when the user does not have a well formulated need but would like a quick bird's eye view of what the books in the database are all about. Of course, skimming through several pages of lists of indexed terms - one for each dimension in the classification scheme - is feasible but tedious. Therefore, icons are an economic way of supporting intuitive searches since a single picture can communicate many different facets of meaning and thus can include terms from more than one dimension. In this way, apart from being effective, a skimming through icons can provide an unexpected aesthetic and emotional experience and give rise to potentially new perspectives on a topic.

This emphasis on storehouses and work rooms facilitates the utilization by the user of the old (Simonides 500 BC) mnemonic trick of locating items to be remembered in given physical locations which one can then retrace in order to retrieve a given

item and/or freely roam (browse) through in order to gain an overview. In addition, users can always keep track of where in the BOOK HOUSE they are so that they can trace their way back if they so desire. Users are able at any time to change strategy, shift database, choose new search terms, print out book descriptions for later use, etc. Shift of strategy can be performed at any level of the search.

Use of icons in the BOOK HOUSE

The most important issue regarding the use of pictures or icons is of course their comprehension by the user. This is not only a function of their content and form but, more importantly, the context within which they are viewed - e.g. where in the retrieval task one is - as well as the experience level of the user. Since the BOOK HOUSE is for novices from 7 to 70 years of age, this is a critical issue. In addition, the use of icons does not preclude the use of other more traditional forms for interface communication and the BOOK HOUSE includes both icons and text.

The BOOK HOUSE uses icons for several purposes. The design has been aimed particularly at the novice user where, when appropriate, the displayed icons preferably should be multi-purpose and permit a direct mapping onto effective cues/signs for action at the manipulative level and onto the semantic organization/content of the database for use in planning and decision making. In addition, as mentioned earlier, they are used to support the user's memory of where in the BOOK HOUSE the various functions/information are located. For example, figure 25 contain a set of small icons along the upper left; these show the rooms which the user has been in earlier and which he/she can return to.

More specifically, icons are often used as signs to the user regarding the actions/choices which can be taken during a search. There are icons available to seek help, see book descriptions, change search terms, etc. The problem is that the user should be able - at least when gaining experience with the system - to connect icon attributes with the particular action. This connection can be with an action state, an object, the act, the agent or even a relation (Pejtersen and Rasmussen 1987). In addition, the pictorial form can be based on a pictogram (a realistic reproduction), an analogy or an arbitrary representation (Lodding 1983). The possibilities are

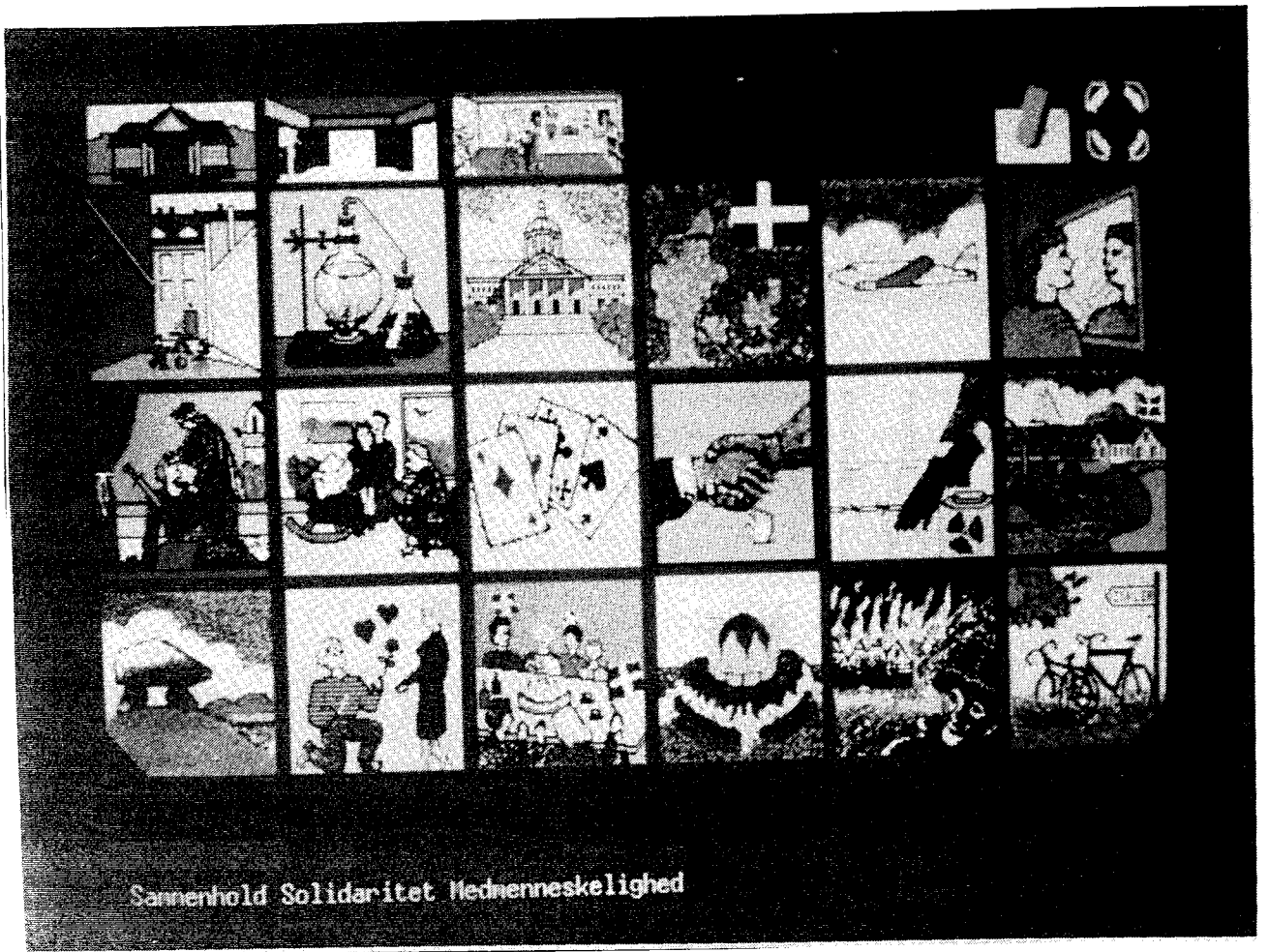


FIGURE 24
ICONS FOR BROWSING IN PICTURES

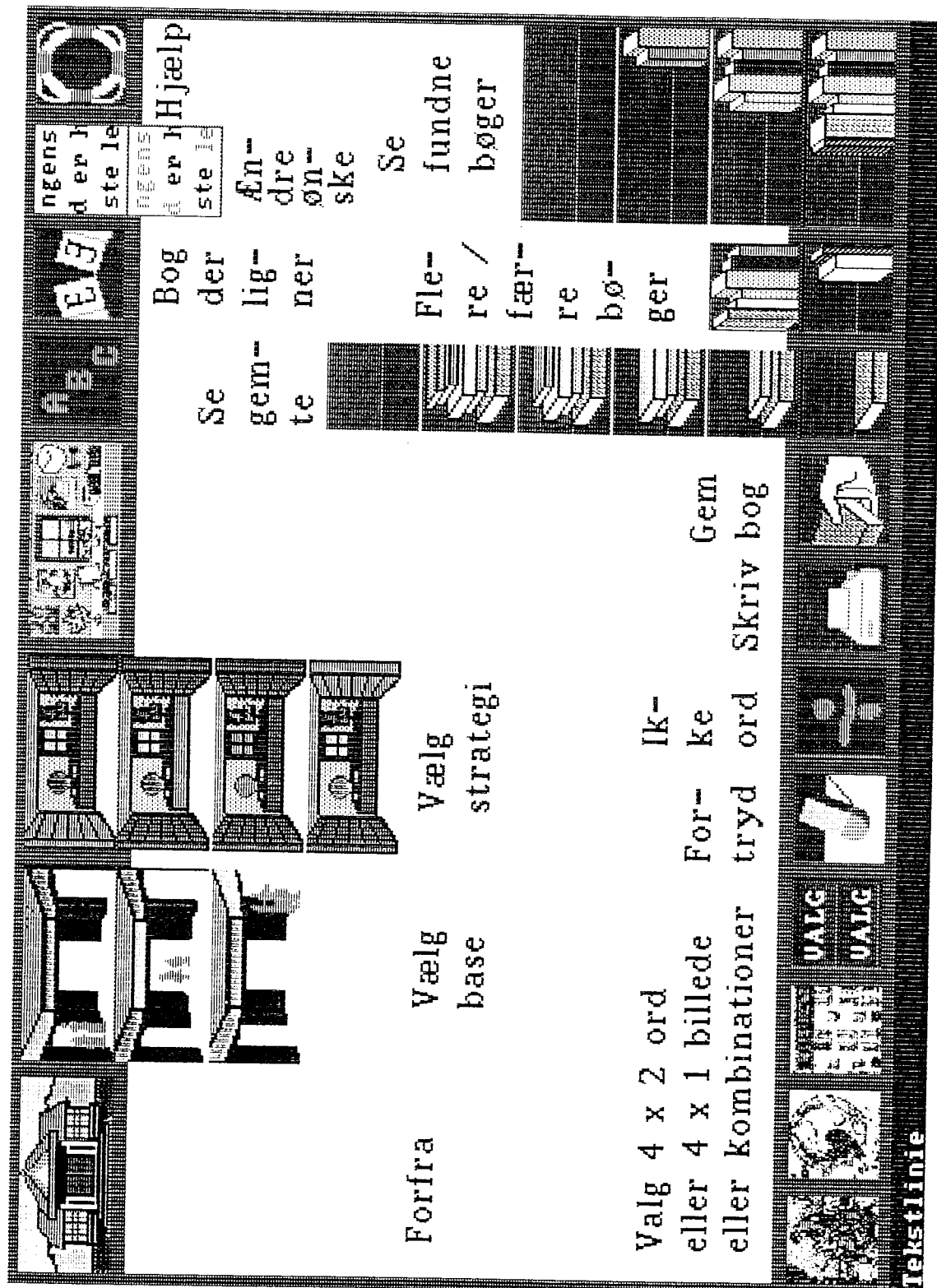


FIGURE 25

ICONS FOR SIGN OF ACTION

enormous and there are no rules or guidelines for making the best (or avoiding the worst) selection.

The most important issue regarding the use of icons is of course their comprehension by the user. This is not only a function of their form and content but, more importantly, the context within which they are viewed - e.g. where in the retrieval task one is - as well as the experience level of the user. Since the BOOK HOUSE is for novices from 7 to 70 years of age, this is a critical issue.

Icons - indeed all forms for pictorial expression - can be perceived in three different ways (Rasmussen 1986, Panofsky 1983):

- as a world of *forms* - lines, colors, use of formed material - as expressions about and/or depictions of "objects" and their relations. Rasmussen calls this the world of signals.

- as *links* between pictorial themes with given worldly ideas, artifacts - more specifically as indicators of an object, act, event, state,... according to conventions building on social, professional, historical traditions. Corresponds to the use of icons as *signs*.

- as *symbols* relating information readily interpretable by the observer on the basis of experience, education, feelings, background.

Thus how a user interprets an icon in a given situation depends on his/her intentions/experience at the given moment. The challenge for the designer is to provide a match between the context (of the information retrieval task) and the icons so that the user perceives the latter in the intended fashion. In the BOOK HOUSE, the design has been focused on the use of icons as:

- *signs* indicating action alternatives during a search.

and/or

- *symbols* providing a symbolic representation of the semantic content of the database - i.e. the substance of book contents and features.

Indeed a trained expert might conceivably perceive most icons as signals in a temporal-spatial metaphorical world where direct skilled manipulation is possible at all steps in the retrieval process. However, as stated, the design has been aimed particularly at the novice user where, when appropriate, the displayed icons preferably should be multi-purpose and permit a direct mapping onto effective cues/signs for action at the manipulative level and onto the semantic organization/content of the database for use in planning and decision making.

The above remarks hold for the content of icons. The communication mode or icon form (called functionality by Arnheim as cited by Lodding (1983)) can be classified as one of the following:

- pictograph (P) = resemblance, captures important qualities
- analogy (ANAL) = a concept, a representation at a higher level of abstraction and therefore often abstract in appearance.
- arbitrary (ARB) = invented, usually using simple geometric shapes to ease the learning process and provide a rapidly recognizable image

All of these alternatives have been utilized.

Special mention should be made of the content icons for representing book contents when users select the browse through pictures strategy. These pictures are implemented as six pages of small icons, with eighteen icons on each page (see figure 24). The approach to constructing these icons was to use the most frequently used index terms as the basis for imagining and drawing an appropriate pictorial analogy. It was found of course that the same icon could serve as the pictorial counterpart of more than one index term. Thus the 108 icons cover more than 1000 index terms. The evaluation of the BOOK HOUSE included a second, but limited, check of match between icons and books in the database (Goodstein and Pejtersen, 1989).

In connection with the creation of these content icons, an pilot experiment was conducted in advance of the final implementation of the BOOK HOUSE to determine whether there was a sufficient consensus between the designers' conceptions of match between content icons and index terms and those of different groups of potential users of the system. The results indicated a good match. ?

Description of the user-BOOK HOUSE dialogue

The dialogue flow is shown on Figure 26 which attempts to compress the many options provided by the system onto a single illustrative representation. There are four main interaction phases:-

A) Select one of two dialogue forms.

The BOOK HOUSE system allows the user to choose between conducting book searches through use of the traditional keyboard-based command dialogue which all Danish public libraries have access to and experience with OR by means of the mouse-based BOOK HOUSE system. This paper covers the latter approach only.

B) Select one of three book bases.

After choosing the mouse-based dialogue, the user is asked to select the appropriate data base. There are three possibilities - the children's bookbase, covering literature for children up to about sixteen years of age, the adults' bookbase or the total bookbase covering both children's and adults' books. A single mouse selection is required.

C) Select one of four search strategies and specify need.

At this point in the dialogue, the user must decide how the search for books will be carried out. As stated earlier, the BOOK HOUSE provides four alternatives:-

1. - an analytical search
2. - a search by analogy
3. - browsing in pictures
4. - browsing in book descriptions

These have already been briefly described earlier. Further details are given in Goodstein and Pejtersen (1989).

BOOK HOUSE DIALOGUE

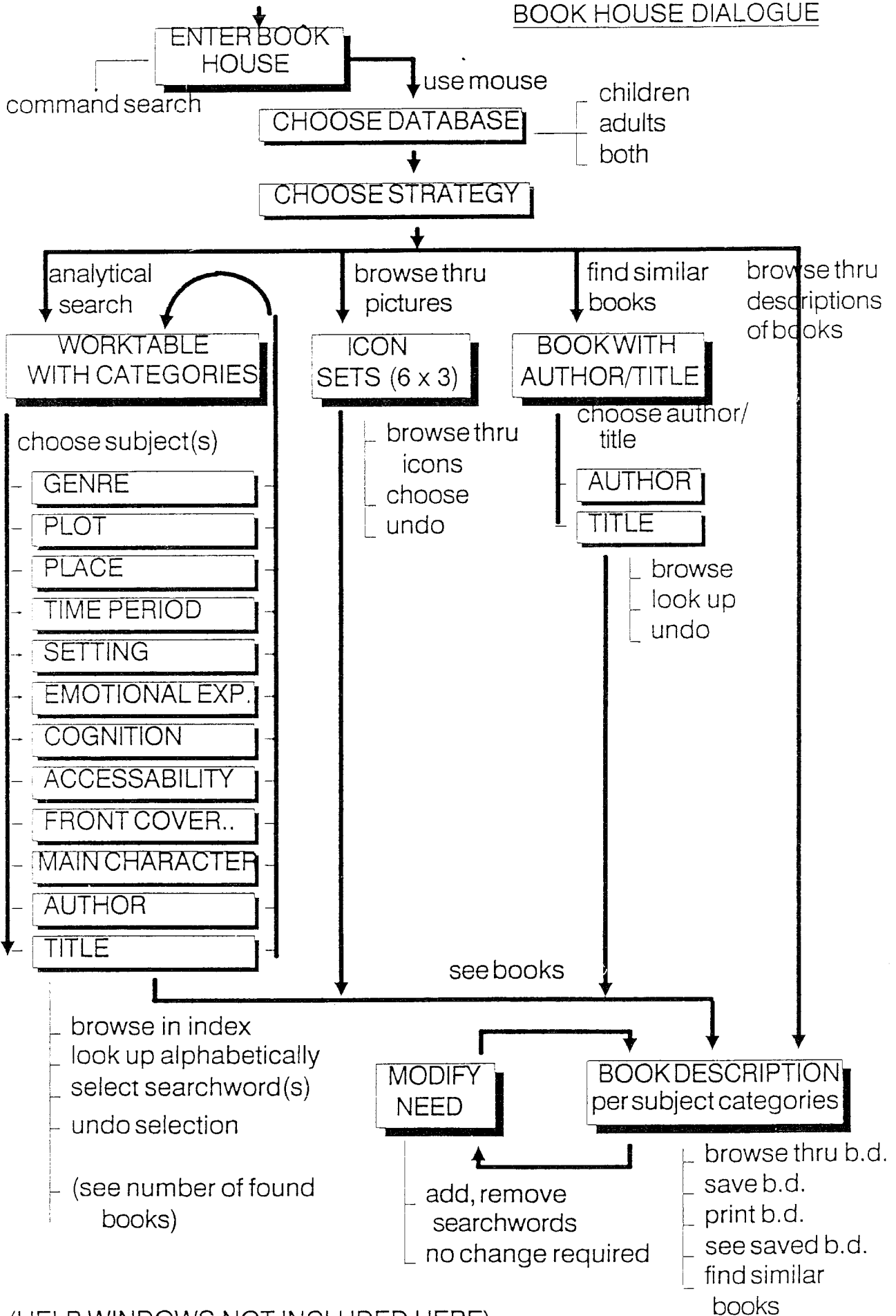


FIGURE 26

Display retrieved books

In a normal dialogue, this phase of the dialogue is entered from each of the above four searches: after either the analytical specification or the browse in pictures selection has been made and see books has been chosen with the mouse; after a reference book title has been specified in a find similar books dialogue or after the user has desired to browse in book descriptions.

A host of options now becomes available independent of the search strategy selected. Firstly, a description of the first book of the current set is displayed on the screen organized in accordance with the classification system. The user can browse through the set - one after the other in both directions or larger leaps can be taken. The number of the book in the current set is displayed in the corner of the screen. The user can now repetitively save interesting candidates in the BOOK HOUSE for later browsing and/or request that hard copies of interesting descriptions be printed - e.g. for use later in finding the books on the shelves. One can also see previously saved book descriptions.

Shift of strategy and revise search profile

After choice of search strategy and examination of resulting book descriptions and process of eventual candidates, the user can revise search criteria and reexamine the search results. This can be done by shifts to either the analytical search strategy or the search by analogy by using the display of the current book on the screen. The user can:

- find similar books which, if the analogical input was selected, will as a start use the reference book as the basis for finding (at least) ten other similar books in the selected book base. However the option can always be used at this point in the dialogue in connection with any current book description on the screen.
- modify the current need which gives the user the option of adding (one or more) or removing (only one) search terms from the current search profile. This is done by displaying in red the searchable index words in the current book description which

the user has access to with the mouse. Each change generates an automatic search, the number of books found is displayed and the first book description in the new set is displayed. At the present time, only the current book set is retained by the system.

This description has only dealt with a "normal" dialogue trace. Of course, the BOOK HOUSE makes it possible for users to make strategy shifts, abandon current searches, select a new database, etc. at any time.

Help texts

There are also help facilities available which attempt to give the user both a context-dependent explanation of where the user is at the moment as well as more general information about the system facilities. These will not be dealt with here.

Implementation

The BOOK HOUSE is built as an application of the GEM (Graphics Environment Manager) system in a DOS environment - however with much original C code for functions not covered by GEM. A laptop PC with an external color display is normally used to demonstrate the system.