4. DEVELOPMENT OF THE SPRING DATABASE

The SPRING Database is the name of the experimental database with 3.500 novels for children and adults, which was used in the BOOK HOUSE. The database was created to investigate how end users in public libraries would react to a new device in the library such as an online subject retrieval facility for novels. It was also created to validate the classification system for fiction in a real life setting. Prior to the SPRING database a number of experimental data bases based on the AMP classification scheme had been developed and evaluated in a laboratory setting at the library school in Copenhagen. In the BOOK HOUSE project the indexing rules needed for a large, realistic database had to be explored in order to achieve an optimal recall and precision, when searching the database. A typical small size public library in the surburb of Copenhagen, The Hjortespring Library, was chosen for the experiment. Half of the stock of novels for children and adults at that time were indexed. The whole stock of the library, 7000 novels, were registered by Faust number. A programme chose every second book as a candidate for the database to ensure that the sample of the database was representive of the fiction stock in the library, i.e. that any kind of novel (classics, modern, popular, complicated etc.) available in the library could be classified. Four librarians skimmed and classified the 3.500 books with an average of one book per hour. Along with the classification a number of rules were developed to achieve a consistent analysis and representation of subject concepts of the books; For the same purpose rules for vocabulary control were devised. These rules are described in Pejtersen and Austin (1990).

Each record in the database has information according to the AMP classification scheme. Information consists of an annotation within each dimension, which tells the reader what the book is about. Each annotation contains a number of controlled keywords. The *exhaustivity* of this annotation is a balance between a trade off: Sufficient information about the content should be provided for the user's decision about relevance -without revealing too much about the course of events in a book- and its ending. Special rules were used for information about the ending of a book. The indexing included the dominant and characteristic features of a novel, not minor aspects. The annotation and its controlled keywords have a *specificity* corresponding to the level of specificity chosen by the author of the book.

Special syntactic rules were developed for the dimensions of subject matter and the author's intention of cognition and information.

LISA   ALTHER: KINFLICKS

Front Page Colours:	White, red and black
Front Page Pictures:	Faces
Subject Matter:	A woman´s visit to her mother´s sickbed and her revival of her youth, student days and marriage. Her experience of her mother´s death.
Place:	USA. Tennessee.
Setting:	Southern States. Middle-class. High school, Feminism.
Cognition/Information :	Realistic description of the American society and of a woman´s love affairs, her development and identity problems. The relationship between mothers and daughters.
Emotional experience:	Humoristic.
Literary form:	Novel. Related in first and third person. Feminist novel. Developmental novel.
Readability:	Average.
Typography:	Normal.

Figure 6

A record from the fiction database of the BOOK HOUSE. It illustrates the multi-dimensional input to an on-line retrieval system. All information in a record can be retrieved by a Boolean search of keywords. The record can also be retrieved by the selection of an icon that shows arguing females or an icon of the USA.

Rules have been developed for the analysis and description of books in all details to get traditionel bibliographical information as well as new content features. The list below is not exhaustive, but the main new features are:

Bibliographical data supplied with:

colour illustrations
types of series,
front page colours,
front page picture,
names of principal characters,
institutions etc.,
age of principal characters in childrens's books.

Controlled keywords supplied with an annotation of:

action and course of events,
place (country, province, town),
time (numerical, periods),
setting (social class, profession, activities),
cognition/information (paradigms, attitudes, literary model),
emotional experience,
literary form (genre, classics, composition, language, literary period/school),
happy end,
readability (read aloud, read alone, age, school class, typography) place on the shelf.

Figure 6 and 8 show two examples from the SPRING database with the information contents of novels analyzed, represented and made accessible according to the dimensions of the AMP indexing scheme. Figure 7 shows examples of pictures representing the contents of classified books.
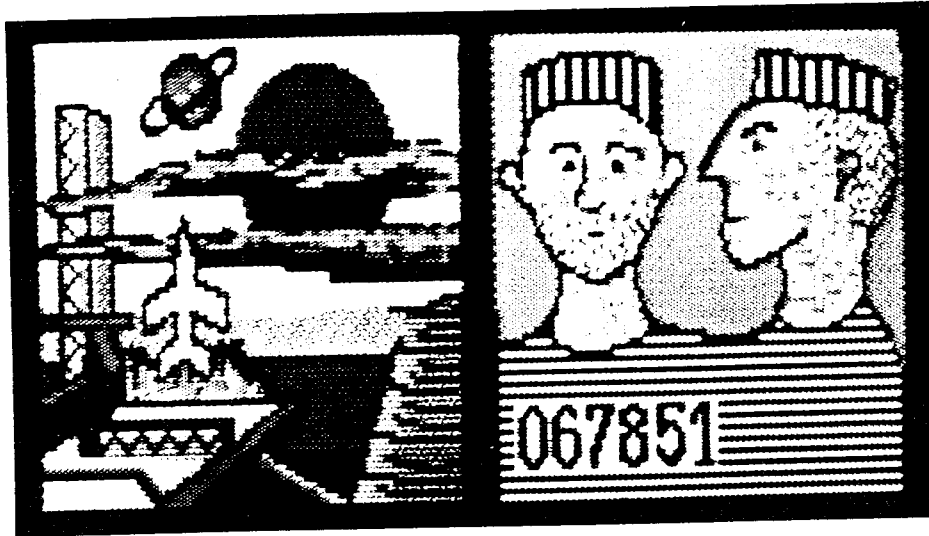
Figure 7

In addition to the verbal description of books in the SPRING database as seen in figure 6 and 8, icons are used to represent book content. Each icon represents all the books in the database related to the topic of the icon. Icons are designed on the basis of keywords in the SPRING database - and its meaning is then tested by end users. Thus each icon represents an associatively related semantic network. The left icon represents the content of books that deal with *future, future society, science fiction, fantastic setting, space* etc. The right icon represents all the books in the database about *prison, prisoners, staff in prison, punishment, criminals* etc.

When a user selects for instance the left icon, the system will retrieve all the books about future *or* future society *or* science fiction *or* fantastic setting *or* space etc. Thus in one step the system performs a multidimensional keyword search across the dimensions of the AMP classification system.

BENT HALLER: THE SONG OF SPERM WHALES

Author:	Haller, Bent
Title:	Kaskelotternes sang, 1983, 137 pages.
Front page:	Blue, sea, whales, icebergs.
Names:	Gamle Dick, Gylte, Tangøje.
Subject Matter:	A calf of a sperm whale's life in the sea. Its struggle to survive in spite of pollution, hunger, and man's killing of the whales. The sperm whales' sticking together in their struggle against the dangers of the sea.
Setting:	Sea environment.
Time:	1980's
Cognition/Information:	Criticism of man's pollutin of the seas and killing of animals on the point of extinction.
Emotional Experience:	Exciting, sad.
Literary form:	Novel, animal story.
Readability:	Age of 11, reading aloud from the age of 7, (happy ending).
Typography:	Big sized letters

Figure 8

A record from the fiction database of the BOOK HOUSE. It illustrates the multi-dimensional input to an on-line retrieval system. All information in a record can be retrieved by a Boolean search of keywords. The record can also be retrieved by the selection of icons.

A1: Bibliographical Search with Instrumental Assistance

The User:

- explores book stock
- compares book contents and need
- selects and decides

The Librarian:

- assists with physical search, with use of tools, explains arrangements and equipments

---

A2: Bibliographical Search with Verification Assistance

The User:

- as above

The Librarian:

- communicates information on content of books selected by user

---

B: Analytical Search

The User:

- communicates information on needs
- accepts/rejects proposed books

The Librarian:

- explores user's needs
- compares to contents of bookstock and selects
- suggests books for user's consideration

C: Empirical Search

The user:

- transmits explicit and implicit signals
- accepts/rejects proposed books

The Librarian:

- initiates signals by questions and proposals
- classifies user according to verbal/visual signals and associates to stereotyped set of authors/titles

---

D: Check Routine

The user:

- requires information on book contents
- accepts/rejects

The Librarian:

- communicates information on content of selected books

Figure 9.

The Strategies and Roles in Search for Fiction

5. SEARCH STRATEGIES

Search strategies developed and employed by searchers in different search situations are good indicators of searchers' preferences in their search environments. These are not necessarily relevant in future search environments, since users' search behaviour is influenced by the retrieval tools available to support the search. Nevertheless, information about search strategies in the present systems is very valuable as a background for the formulation of search strategies in a new information system. The strategies that searchers use to plan the search and their criteria for choice among the different strategies in the present information environment were studied in 134 examples from a set of user-intermediary negotiations collected in 1976 in Danish public libraries under every day library conditions. This analysis owes much to a similar analysis for process control (Rasmussen, 1986). The study involved analysis of strategies employed by end users and intermediaries. The aim of the investigation was to reveal the various tasks and search strategies which the user-intermediary encounter in everyday work in public libraries. When the intermediaries had finished each conversation with a user, they were interviewed about their knowledge of the retrieved documents, their reasons for recommending the individual book, and their aquaintance with the user. The same pattern as described below was also found in negotiations between intermediaries and children as end users in an investigation conducted in 1979.

The users visit the library to select documents which have a reasonable chance of satisfying their needs. With the intermediary's assistance this document selection can be made in principally two different ways. Since the choice of documents consists of a search for a fit between book-contents and user-need, it can only be made by a single person. Either the user or the intermediary can search for this fit. The user knows his need and if he searches by himself, he must procure information about the library system, the character of the literature and its contents. The intermediary on the other hand possesses knowledge about the library system, the character of the literature and its contents, but must obtain information about the user's needs if he is to undertake the search for a match. Several different search strategies have been identified within these frames and are described below and in figure 9 and 9 a. (Pejtersen, 1979).
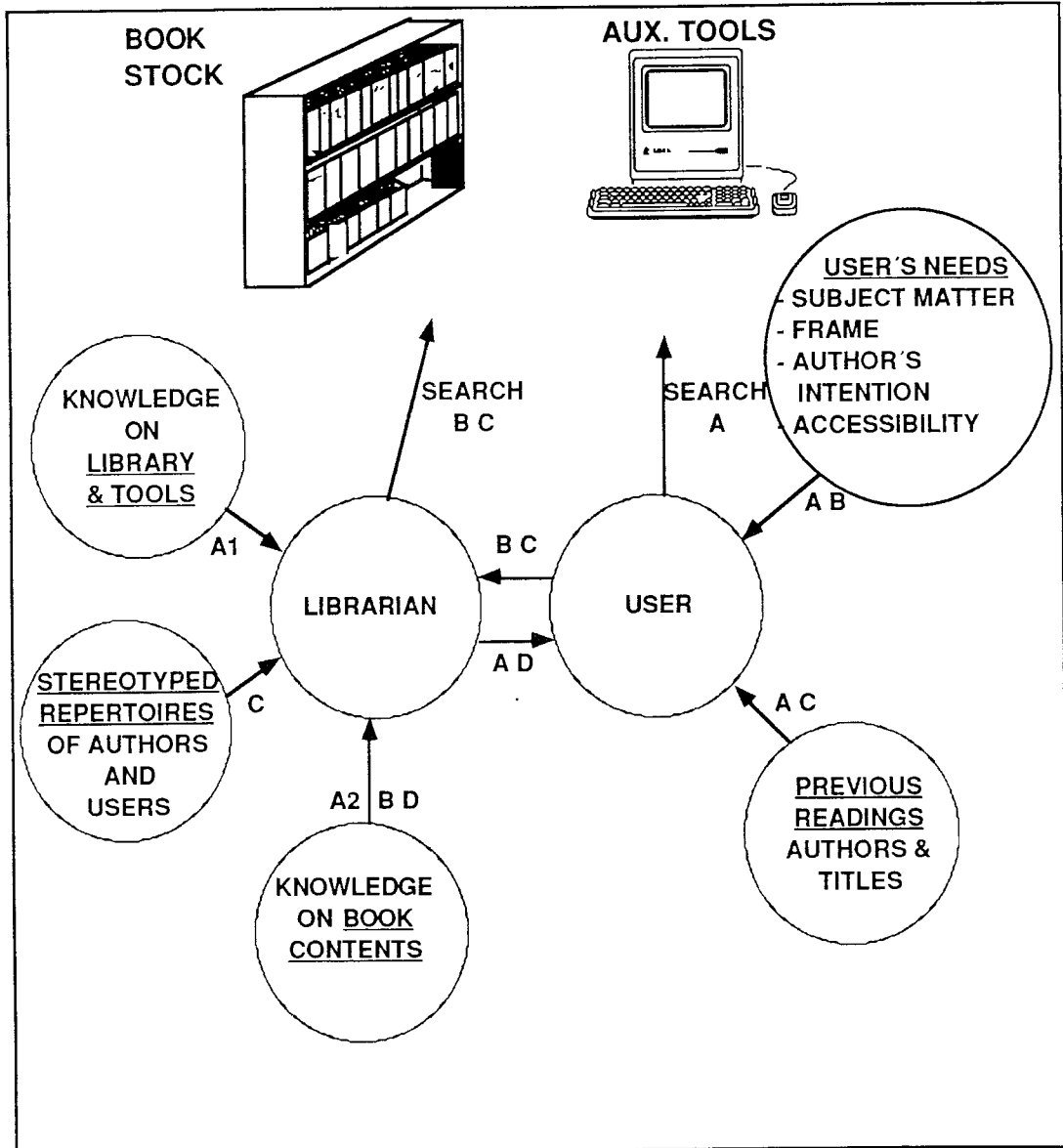
Figure 9 a.

Information flows in search strategies:

A1 Bibliographical search, instrumental assistance

A2 Bibliographical search, verification assistance

B Analytical search

C Empirical search

D Check routine

Bibliographical Strategy with Instrumental Assistance.

In this strategy, the user probes the book-stock and selects books himself/herself and typically asks for assistance from the librarian with questions which refer to author/title. The intermediary typically plays two different roles which activate two variations of the bibliographical strategy. In the bibliographical search, both user and intermediary communicate in library terms and they are occupied with the user's need in terms of authors/titles and their location in the library. The intermediary is the passive helper, assisting in the user's physical search in the library or verifying his/her incomplete references and instructing in library conditions and management. The present auxiliary tools represent the logical, rational and effective solutions to bibliographical and instrumental problems in fiction which the intermediary generally solve in a short time through fixed routine work with few variations. The closing act in this encounter is either the finding of the wanted book or the intermediary's proposal for a book reservation. It is thus characteristic that the intermediary rarely changes to another strategy; he/she is first and foremost defining the task as the manipulation of tools within the framework of the library system. The bibliographical search strategy is the most frequent with an occurrence of 125 times totally, of which only two are initiated by the librarian. It is the most monotonous and least complicated of all search methods identified in this study. This may possibly explain the intermediary's low interest in shifting to other strategies, even when the user seems to set the stage for it.

Generally speaking, the intermediary transfers his knowledge of the library system and its tools to the user, whose need on the contrary is not communicated to the librarian, but exists only implicitly as the user's own criteria of book selection. This strategy is a highly skilled routine task, where the user's and the intermediary's mutual understanding of the task is straightforward. The retrieval tools match the task, and hence the information flow between user and intermediary is not complex.

Bibliographical Strategy with Verification Routine.

This is another variety of the bibliographical strategy with which it shares all the previously mentioned features. The distinction occurs when the used has selected authors/titles that might satisfy his need, but is uncertain about the contents of the

selected books and hence about his choice. The user requests the librarian to assist with the provision of information about book contents by answering questions such as "How does he write?" Although the user's need is only expressed implicitly through his criteria for book selection and therefore is not communicated explicitly, the intermediary should be in a position to characterize book contents in terms compatible with users' needs. The intermediary is dependent on an information system with access to document attributes that match the users' formulations of needs. The information flow is both instrumentally directed as well as directed towards attributes of contents of selected documents.

Analytical Search Strategy.

This strategy is employed when explicitly formulated information about the user's need is transferred to the librarian, who then can compare the need with the book-stock and suggest a selection. In this communication, user and intermediary can play different roles as to initiative. The user can probe his own need and introduce it to the librarian by formulations such as: "It must not take place in the countryside; if it is too boring I'll fall asleep". Or "I would like something from the 1600 - 1700 century, that is what I think is exciting". The intermediary can examine the user's need actively, either through questions about his criteria of value and by suggesting various book qualities in terms of needs, or by repeatedly inviting the user to explicate his reading tastes. For example, this intermediary's questions about book aspects in terms of needs, when the user asks for historical novels: "Do you want books from old ages? From the antiquity?" "What countries should it be about?" "Do you want something exciting - or may it be humorous?" Since the analytical search is based on an analysis of the user's expressed needs and a match between document attributes and formulations of user needs, it is a search in a network of relations between document attributes and parameters of user needs in order to achieve a match between these variables. One would therefore consider this strategy to be the most frequently used. Nevertheless, this strategy is only used by the librarian in few conversations. The user frequently tries to initiate this strategy to a various extent. The explanation of these facts may be found in the circumstances in the library. The analytical search method demands a complicated treatment of data because the librarian in communicating with the user must search simultaneously at four different levels: Analysis of user's need, translation of needs into book contents, identification

of a relevant selection of authors/titles and physical search for them. The success of the librarian's application of the analytical strategy depends on a thorough knowledge of contents of the book-stock systematized according to users' criteria of value as well as an adequate interview technique. The lack of an online retrieval system with representations of document contents as a support to the intermediary's memory which can simplify the treatment of complex data appears to be an important reason for the infrequent use of this strategy. This strategy is the rational way to reach a match between the user's need and document contents. However, skilled intermediaries typically replace this strategy by purely empirical routines based on associations from typical user categories to a repertoire of book-titles, as described in the following.

Empirical Search Strategy.

This strategy is based upon the intermediary's purely empirical classification of users into typical categories which are associated with a repertoire of typical sets of genres and book titles, the contents of which the librarian usually does now know in depth. The user can express his need in many different ways - in library terms as well as in terms of needs, - but the statements are not conceived as a starting point for an analysis of the need. Together with other features in the situation such as sex, age, language etc. they are perceived as signals which characterize the user in relation to the intermediary's typical categories. The negotiations demonstrate an evident correlation between sex, age, and specific authors/titles. For example, middle-aged women are repeatedly offered the same set of authors.

Example of conversation

User, woman, about 50 years old.
User: "Where do you have fiction? - I mean love stories and such things?"
Intermediary: (follows the user to the fiction shelf and asks): "What about Morten Korch?"
User: "I know him by heart".
Intermediary: "Do you know Guldbrandsen?"
User: "I have read them".
Intermediary: "Do you know "Det blæser fra Dødningefjeld?""
and so on.

The intermediary "recognizes" the user's need by a combination of sex, age, question formulation and probably other features in the situation which the material does not

inform us about. In some negotiations the user puts a question to the intermediary after getting about ten book suggestions from the librarian and receives this answer:

User: "I wonder whether that is something for me?"
Intermediary: "Well, I can't tell, as I don't know what you like to read".

In the interview of the intermediary that follows this conversation, the intermediary motivates his book suggestions with two factors: The authors stand on the same shelf and the intermediary has read both authors and thinks they fit middle aged ladies. This is symptomatic and confirms the hypothesis of the application of an empirically based search method. The librarian "recognizes" the need and uses the shelf to support his memory of title/author associations. There are indications from the way the intermediary uses and comments the individual books that the stereotypical set of titles are classified unconsciously in a manner which is not exclusive but in accordance with various "user signals". Consequently, the intermediary's questions to the user, when searching empirically, call for confirmation of signals on the basis of which the intermediary's book sets are classified in order to check that he operates within the right set. A hypothesis worthwhile testing relates to what extent the librarian inherits this book classification from the user. Years of experience with different user signals and acceptance of specific authors/titles may supply the librarian with the empirical foundation. To judge from the user's reactions to book proposals, the empirical strategy often seems to represent a sufficient search, and it is a very efficient short cut to results for both users and intermediaries.

Problems arise when the intermediary's set of titles is too limited, either because the user already knows the books or because the stereotyped set does not correspond with the individual user's actual need, even though the user situation has common features with the intermediary's typical criteria of categorizing. The most common reaction to problems of this kind is associative leaps to other categories of books or simply free associations, the origin of which cannot be traced in the conversations. Paradoxically, it seems to be an advantage when using this strategy to serve unknown users, who are not too familiar with the librarian's repertoire. This is intimated by expressions to users, such as: "You have read practically all the good books" or "Certainly, I would like to help you, but you have already read so much". In his/her own way and without any auxiliary tool, the intermediary has intuitively reduced the great number of unknown factors when searching for book qualities corresponding to users' needs. He/she does this partly by creating his own tool in the

form of sets of user-classified titles/authors, and partly from conducting searches on the basis of vague user formulations, which are not fully explored but instead are conceived of as signals. This is an effective compensation for tools, and it is most likely the best solution to the search problems caused by the present situation in the library. It is effected in 65 cases and is thus the second most frequent strategy employed. It seems to be a significant observation that the intermediary may also substitute an analytical search for the intuitive empirical search method, when an unusual situation occurs. One of the analytical searches occur in a situation which is explicitly classified as especially difficult by the intermediary because the user has read so much. Another example is exceptional because the user has come to the library on behalf of his father. In this conversation, the intermediary's search is based on an interview of the father's attitude to various book aspects in terms of users' needs. Obviously, the empirical search based on signals cannot function, when the user himself is not present in the library, and consequently the intermediary must utilize the exploration of the user's need as a source for book selection.

Search by Analogy.

In the search by analogy, a model book mentioned by the user is the basis for the retrieval of new documents with features identical with those of the model document. "I want something similar to book X" is the usual question. This search is therefore a search in a network of relations between attributes of documents in order to achieve a match between the features of documents in the information system and the features of the model book. Or it is a search for documents with a specified, selective number of features in common with the model book. In both cases, the search by analogy is a version of the analytical search strategy. The model book can be the focus for the analysis of users' needs, checking document features along different dimensions in order to identify which of the features are representative of the user's need. The search by analogy is often combined with the empirical strategy. From this follows that the intermediary associates from the model to its membership of a stereotypical category of documents and other items belonging to the same category are retrieved. In the analytical search by analogy the information flow is the probing of attributes of the model document in order to determine the selectivity of the need. In the empirical search by analogy the information flow is identical with that of the empirical strategy, where the user's acceptance/rejection of the intermediary's suggestions of

authors/titles proves or disproves the intermediary's association from model book to a similar category of books.

The Browsing Strategy.

The browsing strategy is characterized by a lack of requirements to document as well as information specificity. The search can be part of a process of need recognition, of search question formulation and of a learning process, where the strategy is a means to get new knowledge or new associations and ideas within a familiar or a new subject domain. The need is usually vaguely defined and when browsing, the user and intermediary intuitively explore the book-stock for good ideas. An approach which often is manifested by sequences of author suggestions in alphabetical order, following the shelving arrangement of fiction.

Check Routine.

This is one of the important subroutines, initiated by the user or the intermediary during the analytical and empirical search strategy, when the intermediary selects books on behalf of the user. The user requests the intermediary to communicate book contents as a means of checking the agreement between his criteria of value and the qualities of the document suggested by the intermediary. The check routine thus functions as feed-back to the user telling about intermediary's concept of his need. This is also the case when the intermediary takes the initiative to a feed-back by giving supplementary remarks about book contents regarding his book suggestion. This information enables the user to evaluate the identity between book contents and need as well as the intermediary's criteria of book selection. The number of check routines is 40, carried out in several variations. The full realization of the check routine depends entirely on the intermediary's capability for expressing book contents in user-relevant terms. The intermediary often has to deduce information from casual, secondary sources, such as bookflaps, or even decline to enter the check routine. In most cases neither the user nor the intermediary has information available to check the correspondence between need and book qualities. The front page of the bookflap is often commented on by the user, and his/her rejection/acceptance of the intermediary's "title/author only" proposals indicate that the user compensates for missing information about book qualities with value

judgements relating to the quality appearance of books. On the other hand, the intermediary has developed a number of routines for participation in check routines as well as verification routines and carries out what might be called a pseudo check routine, consisting of a stereotyped social interaction. A number of examples can illustrate how the intermediary substitutes secondary information for information about book contents in terms of needs:

1. The intermediary generalizes and refers to the authority of other users.
2. The intermediary refers to his own authority by saying: "It is a very good book, indeed, I like it myself".
3. The intermediary characterizes the book by traditional user signals and genre classifications according to which he has unconsciously classified the book.
 4. The intermediary repeats the user's formulations of need telling indirectly that the book he suggests will satisfy his need.
5. The intermediary quotes aloud from the book description on the backflap before handling it to the user.
6. The intermediary reports formal book information such as time of edition, numbers of volumes, series etc.
7. The intermediary explains that he/she has not read the book.

Frequently combinations of these different ways of controlling the check routine are found.

Shift of strategies

The main features of characteristic shifts from one strategy to another on the initiative of the user or the intermediary have been identified. Generally speaking, changes in strategies occur primarily when problems arise, because the use of one strategy does not lead to a result that is satisfactory to either the user or the intermediary. Changes occur also when the user expresses several needs subsequently, which may demand different strategies in order to be satisfied. The previously mentioned search strategies are regarded as different forms of dialogues between user and intermediary during a search process. Both can activate this communication, and whenever an initiative occurs in the negotiation, whether it is taken by user or intermediary, it is counted as part of the analysis, even if the opening of a strategy does not result in the

participation of both parts in a complete progress of the strategy. Especially, the analytical strategy and the routines of verification and check often exist as rudimentary pieces which obviously belong to the categories but do not always appear as a consistent communication. This tendency is clearly connected to strategies that call for the intermediary explanation of book contents and for the user's examination of his own need, both of which seem to cause problems. But an attempt to change and carry out a different strategy is considered more important than the skill with which it has been undertaken.

Shifts from Bibliographical to Analytical Strategies.

Users' initiatives in changing from a bibliographical to an analytical search occur when the user himself has selected relevant books that for some reason or other are inaccessible. They occur also when the user has insufficient knowledge about the book-stock and wants more books with the intermediary's help to satisfy the same need, or when he/she has an alternative need. This is the most frequent change of strategy made by the user. This may lead to the assumption that a user's enquiry for authors/titles in many cases represents a way of opening a conversation rather than a wish for specific titles. Therefore, the intermediary should always check whether the user's bibliographical search could be replaced by an analytical search. When the user changes to the analytical strategy with an explicit demand for help, it is normally met by the use of the empirical strategy to which the intermediary changes. In some cases, the user changes to the analytical search by explaining his reading taste to the librarian in support of his query for a specific title. The intermediary does not follow the user in this situation, but continues with the bibliographical strategy. When books selected by the user are not available, it happens that the intermediary takes the initiative to change from bibliographical to analytical search. Though the intermediary may be succesful with changes from the bibliographical search to the analytical search, it does not happen frequently. On the contrary, authors/titles requests are often considered as goals rather than as symptoms/exponents of needs.

Shift from Analytical to Empirical Strategies.

The alternation of the analytical strategy with the empirical strategy can only be activated by the intermediary. He/she alternates an analytical search initiated by the user with an empirical search for sets of book suggestions in 44 cases. When the intermediary makes a fully systematic exploration of the user's needs he/she changes to the empirical strategy after a while, probably because the intermediary soon runs short of books with the combination of all features that would fit to the user's need. By changing to the empirical strategy authors from his/her standard repertoire are retrieved, and the user may be asked to rank the attributes of his requirements to document contents.

Shifts from Empirical to Analytical Strategies.

The change to the analytical search is usually prepared for by the user when the librarian operates within the empirical search, either as a reaction to the intermediary's book suggestions or as feed-back comments to retrieved document. Usually, the intermediary continues the empirical strategy. From a realistic point of view, the change to the analytical strategy when he/she is met with reactions from the user could be more risky than the continuation of the empirical strategy. The exploration of the user's need might underline the probable discrepancy between need and repertoire. There is little probability that the intermediary through an improvised long term memory search outside his repertoire and without a retrieval tool would be able to supply the user with books of more relevant contents than those which empirically are known to be popular with certain groups of users. Another way of indicating a misfit between need and book proposals is the user's change from the empirical strategy to the bibliographical search, as seen in the following.

Shift from Empirical to Bibliographical Strategies.

When the set of titles presented to the user during the intermediary's empirical search does not fit the user's actual need, he interrupts the search process by taking over the selection of authors/titles as alternatives to the intermediary's book selections. Regression to items known by the user caused by the incompatibility of retrieved

documents with the user's need is one reason for this shift. Another function of this shift is its opening to a different need.

Shift to Check Routines.

During the empirical search, the user compensates for the lack of exploration of his need by inserting check routines in between the intermediary's book suggestions. This happens, when retrieved documents prompt further elucidation of their content and its potential match with the user's need. The intermediary initiates the check routine during empirical searches as a feed back to his/her choice of document category. This feed-back may lead to a change of category of documents, but it does not lead to a shift to the analytical strategy.

Resource Demands	Bibliographical	Analytical	Empirical
Short-term memory	Low	High	Low
Time	Low	High	Low
User motivation		High	Low
Book knowledge	Low	High	Low
Retrieval tools		High	Low
Personal capability		High	Low
Experience with users	Low	Low	High

Figure 10

Relation of source demands to different search strategies

6. RESOURCE REQUIREMENTS OF STRATEGIES

The information retrieval situation is complex. To a large extent, the choice of a given strategy will depend on the match between the requirements of the strategy in terms of information about users and documents, information processing capacity, etc., and the human resources that are available in the search situation. These demands will change from one question situation to another, from one search environment to another, from one searcher to another as well as during an individual user's search. Resources that cannot fulfil the requirement of one strategy may be adequate if another strategy is chosen. It is therefore necessary to characterize the type of demands on searchers' resources that the different strategies require. In considering the demands on mental resources, the complexity of the information processing task is important. The amount of information to be treated varies with strategy, but is highest in the analytical strategy. The demands on short-term memory and long-term memory are higher within the analytical strategy than within other strategies, and there is a greater need for depth of knowledge about document contents and subject fields. The time needed for the analysis of the user's world, the probing of the need, etc. is high in the analytical search strategy. When going through a list of demands in relation to the various strategies, the analytical strategy seems to put the heaviest demands on the searcher's mental resources (see figure 10, 11 and 12) where a number of different demands for resources illustrates the characteristics of the search strategies). This is significant for the fiction retrieval situation, where no retrieval tools with subject access were available in the libraries where the searches took place.

INFORMATION BASIS	SEARCH STRATEGY		
	BIBLIOGRAPHICAL	ANALYTICAL	EMPIRICAL
FRAME OF REFERENCE	ORGANISATION OF INFORMATION SYSTEM	USER NEEDS; CONTENTS OF DOCUMENTS	TYPES OF USER BEHAVIOUR; CATEGORIES OF DOCUMENTS
INFORMATION FLOW	IDENTIFICATION, VERIFICATION, AND LOCATION OF ITEMS	ASPECTS OF USER NEEDS; ASPECTS OF DOCUMENTS FOUND	USERS' APPEARANCE, VISUAL AND VERBAL EXPRES-SION; DOCUMENTS READ

Figure 11

Knowledge needed for search strategies.

TIME SPENT
IN SEARCH

EMPIRICAL STRATEGY   ANALYTICAL STRATEGY

KNOWLEDGE   SHORT TERM

ABOUT USER   MEMORY LOAD
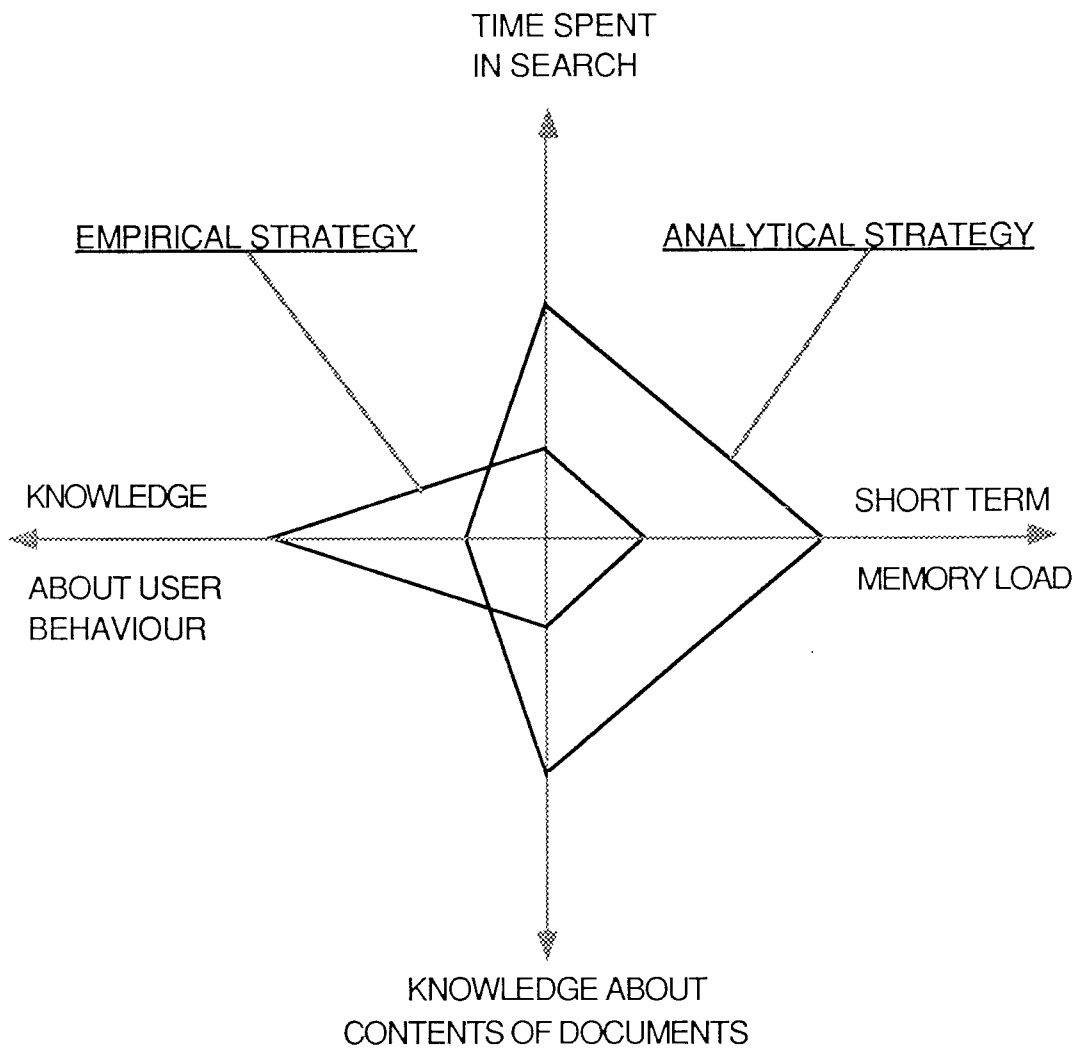BEHAVIOUR

KNOWLEDGE ABOUT
CONTENTS OF DOCUMENTS

Figure 12

Comparison of resource requirements of the analytical and the empirical strategies to illustrate how a switch of strategy can resolve a resource-demand conflict.

7. TYPES OF CATEGORIZATION

Natural classification

Choice and development of search strategies are necessarily context dependent, but since the need formulation and the search for document identification and match are in essence a process of categorization, certain similarities with results from context-free experiments are likely to exist. Studies of natural classification, i.e., intuitive classifications that are made in daily life situations prove that approaches other than the employment of formal defining attributes and logical classification rules are common. Traditional classification theory is used in classification schemes in bibliographic retrieval systems of non-fiction works which typically follow such rules. However, a questioning of the long tradition in classification that has its origin in Aristotelian logic has been initiated by arguing that a natural classification of situations, peoples and objects is not based on a single set of criterial attributes shared by every member of a given category. On the contrary, it is claimed that natural classification is based on a resemblance between similar families of classes and patterns of similarities. This categorization is probabilistic; it has members that vary in degree of membership of different classes, with unclear limits between classes. These classes are therefore better described by clear and unambiguous examples, prototypes, than by classification attributes. In other words, the criterion of representativeness is always involved in natural categorization. (Rosch, 1980).

Prototypes in system design

In fiction searching, the phrase "something similar" is frequently used to formulate a user's need and to communicate desired features of books. The search is then accomplished through an evaluation of similarity of retrieved documents with a title, an author's name, or a genre term perceived as a typical member of a class of documents. Or a search by similarity is initiated by the intermediary asking the user to give an example of a previously read document that is evaluated positively by the user. Documents are then retrieved through an evaluation of their similarity with the previously classified examples, the class membership of which are therefore well known. The perception of the user's need is then in both cases modified through his/her positive or negative reactions to suggested titles. Studies of the development

of rules for membership of a class show that categories are better learned through prototypes and well-known examples than by a number of abstract rules for a class membership. This may explain fiction users' preference for retrieval and negotiation of need through similarity with prototypical examples of document categories, and their infrequent use of definitions of the desired features of documents on a higher level of abstraction. Thus there seems to be an indication that information retrieval systems should allow the user multiple access to information - not only analytical strategies, but also strategies for searching by the most representative features such as the empirical strategy, and as well searching through similarities between individual features of a document (e.g., the analytical search by analogy). This means that the indexing language must express document contents on different levels, such as specific terms, and most representative features of contents, which could then be associated with similar features in other documents. This policy was followed in the indexing of the books in the SPRING Database: Specific terms were used as well as broad, generally accepted labels for each book. The problem with the associations between terms of documents, whether these are individual attributes or features representing prototypical categories, is that very little is known about how humans in an information-seeking environment actually make associations between documents.

Categorization by stereotypes

When the prototypes used for categorization degenerate to include only a few outstanding features, they turn into stereotypes. Most research about stereotypes deals with stereotypes as a set of inferences and expectations towards a person based on the person's group affiliation. Usually physically outstanding features will be used in the classification of persons, and stereotypical models are therefore often based on appearance (e.g. sex, age, race) as is also seen in the empirical strategy. The criteria attributes seem to be based on perceptual categorization, when stereotypes form the basis of a classification. Stereotypes can also more generally be described as a result of the use of prototypes that are not representative of a category. The categorization by a prototype in a natural classification is an attempt to seek attributes common to all members, and thus avoid the individual attributes of the members of a category. This is a sufficient means of categorization as long as it is not

the individual features which have importance in a particular instance. In an information-seeking environment, where users are likely to use natural categorization of information by similarity with prototypes, stereotypical models are likely to play an important role. Several experiments show that the effect of stereotypes may be normalization of new information in agreement with a stereotype, addition of attributes in order to supply lacking information, and neglect of information that is obviously present. Studies of the function of memory in relation to stereotypes indicate that search and recall of information in memory are controlled by stereotypical models in the same way as the perception of new information (Hamilton, 1979). If the strategy required by the information retrieval system is too complex in terms of demands on memory and the need for the use of conceptual analysis, stereotypical models have a high probability of being used in a searcher's inferences and expectancies towards needs and information of documents. This happens in real life searches, when stereotypical models of users based on their sex, age, question formulation, etc. are used to infer their information need, as for instance when elderly women are correlated with family chronicles, young men with suspense novels, etc. It makes the retrieval of new information easier, because it will enable a searcher to predict the appearance of information and to plan the search strategy accordingly. A searcher will often be in a situation characterized by uncertainty about need, type of information that might satisfy the need, search options, etc. In such situations, many intuitive judgements are to be expected, together with the use of a number of simple rules that reduce a complex evaluation problem to more simple judgements. The conclusion of this review is that there seems to exist at least three different ways in which humans categorize information about their surroundings. There is a fundamental difference between these three types of classification codes and the way in which information is processed:

1. The classical method, where a list of necessary and sufficient conditions in terms of attributes is probed. A conceptual categorization of defining attributes is utilized. The analytical search strategy implies this identification of abstract attributes treated independently of empirical means.

2. The example method, where the similarity of an example with other examples from different categories is probed, based on common attributes. A perceptual

categorization of similar features is used. This information process is characteristic of the search by analogy.

3. The prototypical method, where the similarity with typical members of a class is probed, and empirical rather than formal information is used. The categorization is perceptual. This information processing is characteristic of the empirical search strategy, where the identification of prototypical features is involved in the determination of user category.

We may assume that in familiar routine tasks in manual searches, perceptual categorizations by analysis of attributes are likely to be used. Conceptual categorizations are probably used more in unfamiliar tasks, where intuitive jugements of similarity with past instances are insufficient as was the case with the analytical searches. Furthermore, a searcher's categorization of information will also be influenced by the structure of retrieval tools. With computer-based bibliographic databases, new search facilities will be available, which may support a conceptual categorization with analysis of attributes of user needs and attributes of document contents if properly designed.