Online Review

A cognitive view of three selected online search facilities

Peter Ingwersen

A cognitive view of three selected online search facilities

Peter Ingwersen

Abstract: This paper discusses the impact of three selected command language facilities on the man–system interface in relation to operational online information retrieval (IR). The concept of data and information is briefly considered in relation to the cognitive viewpoint and Brookes 'fundamental equation of information science.' A cognitive IR model is outlined, followed by a discussion of variously experienced searchers' knowledge structures. Recently developed online search facilities such as use of positional operators (free text operations), crossfile searching and term frequency analysis (zooming — as the ESA-Quest command language calls it) are discussed in relation to the IR process. The cognitive view is applied as the means to describe and emphasise the retrieval possibilities and the state of knowledge involved in the interface processes.

1. Introduction

At present the global information environment appears to be in a transition period. In particular, the online industry is beginning to realise that certain basic views and assumptions about IR will have to be revised. There are several reasons for this.

Most important is the challenge posed to the industry by personal computer, videotex and word processor manufacturers.

In addition, the demand for and production of data seems ever increasing. The commercial (or semi-commercial) online IR systems are becoming highly complex and interdependent with the surrounding society.

New user groups are evolving. To the traditional user community deriving from scientific, technological and business fields will be added the ordinary citizen with his home computer or videotex equipment. Unlike a decade ago, user groups now consist of persons covering a very broad range of learning and educational levels and knowledge of information retrieval.

New technological possibilities, the increasing complexity as well as the need for relevant data posed by a growing body of users to IR systems have a great impact on the man–machine interface. This interface is controlled by the command or search language possessed by the IR system. Two crucial factors on which the

The author's address is Royal School Librarianship, Birketinget, DK 2300, Copenhagen S, Denmark.

About the author

Peter Ingwersen

Peter Ingwersen, at present on leave from the Royal School of Librarianship, Copenhagen, joined the ESA-IRS Online Service staff as an ESA Research Fellow in autumn 1982. His R & D activities are concerned with user-system interface improvements and the development of a new family of online computer-aided self instruction tools. He graduated in Library and Information Studies at the Royal School of Librarianship in 1973, and since then he has lectured on information storage and retrieval in Copenhagen. From 1976 to 1980 he carried out experimental research on cognitive psychological aspects of processes dealing with user-searcher systems interaction. He is at present working on online tutorial improvements, multilingual database management and download aspects of knowledge-based IR systems.

success of today's online searching depends are the searcher's ability to cope with the functions and effects of the command langauge facilities and his understanding of data output, for example the record display layout — meaning of codes — confirming or error messages. Consequently, the searcher's background knowledge, experience and intellectual flexibility, as well as the capabilities of the IR system, become highly important.

The aim of this paper is thus to analyse the impact of some selected search language features on the user-system interface in online retrieval, seen from a cognitive point of view.

This means to distinguish between data and information and discuss which kind of knowledge structures (KS) and variously skilled user groups are involved and how the perception and information processing is carried out by the participant in online IR using such facilities.

The cognitive point of view is relevant because it deals with disciplines such as psychology, linguistics and artificial intelligence, (their intersection of ten defined as cognitive science [1]) — as well as processes such as perception, problem solving, learning and cognition — all closely related to and overlapping information science and IR.

Briefly, this view states that it is the individual model of the world (the actual knowledge structures) of the information processing device — whether human or

" . . . . the user is rarely able to specify the information required — and rarely in a suitable form that matches the system — at least not in one single statement. From IR studies it is known that initiating query statements often take the form of a 'label' containing few terms, conceptually far from the underlying need"

machine — that mediates the actual information processing [2]. This model of the world is determined by the individual and social collective experiences, education and developed skills.

Within this framework one must emphasise that present operational IR systems consist of passive or fixed KS, embedded in mechanisms like database structures, command language syntax and indexing rules — the system's setting — at the event of problem formulation and information searching. The only active and changeable cognitive structures in the interaction are those of the searcher.

The following features are related to the IR process and exemplified using the ESA-Quest (ESA-IRS) command language:

— free text searching using contextual logic
— crossfile searching (Questindex)
— term frequency analysis (Zoom).

The first two are common to search languages available on most operational systems. They are traditional system spin-off effects, designed to cope with the increasing number of accessible term relations, records and databases.

The Zoom facility is based on the realistic view that the user is rarely able to specify the information required — and rarely in a suitable form that matches the system — at least not in one single statement. From IR studies it is known that initiating query statements often take the form of a 'label' containing few terms, conceptually far from the underlying need [3].

By displaying ranked concepts from an entered set of references Zoom may support the user restructuring or specifying the information need expressions. Zoom allows alternative ideas to emerge and serves as a relevance feedback mechanism.

2. Data and information

The central point in any information processing, according to the cognitive view, is that 'the processing is mediated by a system of categories or concepts, which — for the processing device — are a model of its world.' This world model is called 'Image' in Figure 2 and consists of knowledge structures.

De Mey has defined 'information' as seen as "supplementary to a conceptual system that represents the information processing system's world knowledge" [2].

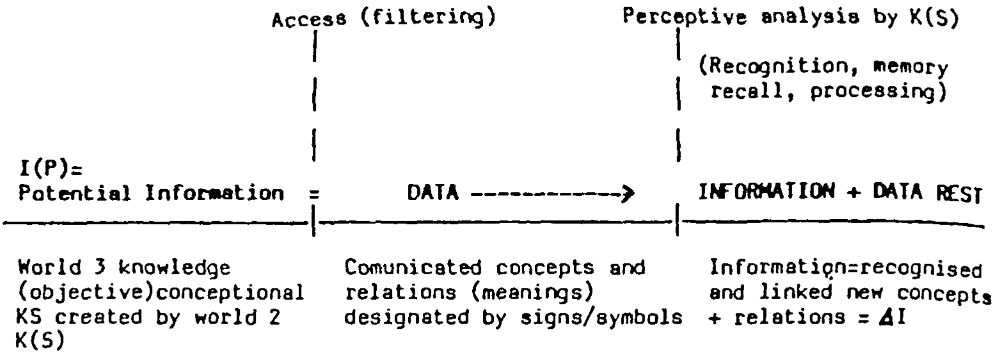

Figure 1a. *Information retrieval.*

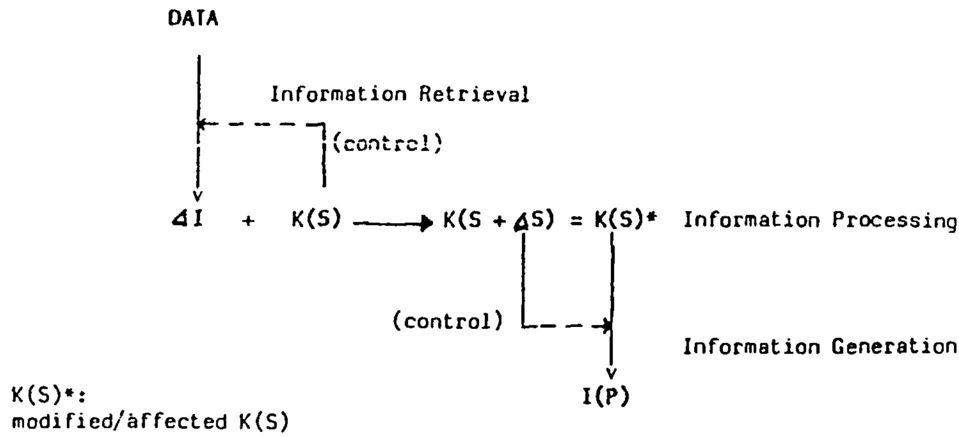

Figure 1b. *Information processing.*

The system's world knowledge — its knowledge structures — provides the basis for decisions on which ambiguities (or problems) should be eliminated.

This implies that 'information' is that knowledge — or conceptual context — which must be present as a part of the processing system's KS to solve ambiguities or problems. Consequently, 'information' is knowledge which has to be perceived and/or produced by the system. Otherwise, the actual problem cannot be solved completely. Thus, the system may try to find the information, either by obtaining it from outside itself, i.e. to perceive knowledge from other systems and/or by 'thinking,' i.e. modifying its KS.

This notion of 'information' is similar to that of Brookes, who introduced the 3-world model of Popper and its relevance to information science [4]. It was also reflected by Mackay in 1960 [5]. Other concepts of information are reviewed by Belkin [6].

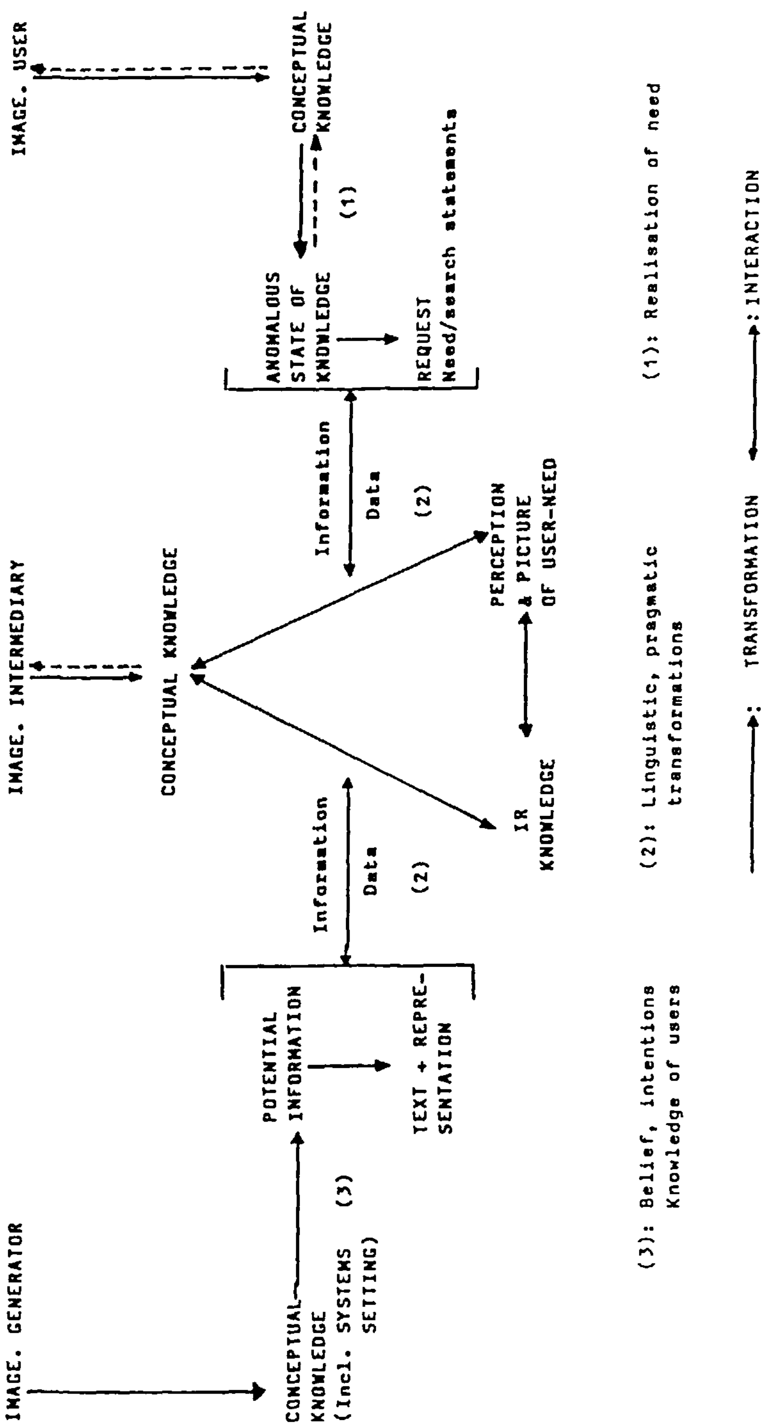

Figure 2. *A cognitive communication system for information retrieval (from Belkin [19] and Ingwersen [3]).*

Brooke's Equation,

$$K(S) + \Delta I = K(S) + \Delta S)$$

in which K(S) is an existing knowledge structure, affected by some increment of information, ΔI, and K(S + ΔS) is the adjusted or new, modified structure [7], is very useful since it stresses what happens if a processing system (e.g. a person) perceives and interprets 'something' which thus becomes information.

This 'something' is data. If not perceived (or mediated) it remains data. To perceive data the processing device requires KS — category or concept structures — that are able to interpret the data. This may happen via recognition. From studies of the memory [8, 9] and the learning process [10] it is known that we memorise, learn and think by means of concepts. To supply a KS with new concepts or concept relations implies that they somehow are linked conceptually to recognised or known structures.

Concepts are defined by Gowin as "perceived regularities in events or objects as designated by a sign or symbol" [11]. One may therefore suggest that Popper's World 3 — the world of objective knowledge — [12] is 'potential information' consisting of generated beliefs, intentions, ideas, theories . . . in the form of (objective) conceptual KS. See Figure 1a and 2. When World 3 is accessed the 'potential information' is data in the first place. The data is communicated designations, i.e. signs, symbols, words, text . . . that contain what Schank calls meaning and inference [8] and the French philosopher Merleau-Ponty called 'silence' or thought [13].

In the event of perception, the data is transformed by the actual KS into information omitting non-perceptable restdata. Some potential information, however, does not represent new knowledge. Concepts or concept relations may, as stated above, be recognised, being previously stored in the memory. It is the new information that may transform the KS or act as a clue to modification. How the information affects the KS depends on the state of knowledge of the individual and the complexity of the information perceived [ΔI, Figure 1].

In view of this, Brookes' formula may contain an arrow and not an equation sign to stress the dynamic aspects and active role of the Knowledge Structures:

$$\Delta I + K(S) \rightleftharpoons K(S + \Delta S)$$

in which ΔI is recognised and new information and ΔS is the impact of ΔI on the former K(S). Figure 1 outlines some key steps in this highly complex process. Figure 1a demonstrates Brookes' scheme embedded in (human, subjective) information processing, based on supplementary data input, which is a vital part of the cognition process.

In online IR the searcher operates on two kinds of data, communicated by the system:

— conceptual knowledge data
— system setting data

On the other hand the IR system receives similar data from the searcher. The system's information processing, however, follows a slightly different pattern than outlined in Figure 1. ΔI is reduced to solely recognised designations. Consequently, when rejecting non-recognised (= new, unknown designation = rest data) present IR systems' K(S) remain the same. However, they may look intelligent and perceptive, exactly because they displayed stored, objective knowledge structures containing ideas of searcher behaviour.

3. Knowledge structures in IR

A searcher is defined as an intermediary, e.g. an information specialist, or an end-user, i.e. the person or team that need to use information for example to solve a problem.

To be able to interrogate an IR system the searcher must possess sufficient 'IR knowledge' consisting of knowledge on:

— the system setting, i.e.
- — command language functions & effects
- — database structures
- — indexing rules

— the retrieval process

To be able to process the conceptual computer responses as information the searcher must possess a certain degree of 'conceptual knowledge'. When initiating a search the end-user has recognised an inadequacy in his/her conceptual state of knowledge, named ASK (Anomalous State of Knowledge) by Belkin and Ingwersen [see Figure 2] and the Conscious Need by Taylor [14].

The system's KS consists of the setting as well as the conceptual relations generated by authors. Thus, the notion 'generator' contains implicitly objective knowledge structures deriving from authors, systems designers and maintenance staff as well as file producers — all responsible for the total image of the IR system [15].

If an intermediary mechanism is involved, that person or machine must possess an ability to 'picture' what the ASK is concerned with.

The model emphasises the possible implications and complications to be considered in designing intelligent IR systems, eventually replacing the human intermediary;

(1) the dynamic interaction between conceptual knowledge, actual perception of needs and professional IR knowledge, (2) the limitations in the interface situation of the images involved with respect to transformation of data into valid information [3].

If a human intermediary participates, the crucial problems or limitation for them are to be able to:

(a) understand the user's ASK — forming a picture of it by means of interview
(b) analyse the current IR situation and the setting of the IR system(s)
(c) create representations of conceptual KS of large scale systems in relation to the subject — to form part of his own conceptual KS
(d) link interactively the picture and the representation, presenting potential information to the user.

The intermediary may not be in a position of sufficient conceptual KS to deal with the subject asked for. The optimal situation in such cases is when the end-user and the intermediary act together — each one operating his own speciality. Klaséns

"In case of automated intermediaries — whether forming part of a mainframe or being a personal computer — the problems concerned are basically similar to those of humans"

survey of the Swedish online population from 1982 emphasises explicitly this problem area [16].

In case of automated intermediaries — whether forming part of a mainframe or being a personal computer — the problems concerned are basically similar to those of humans. It is the automated interview, analysis and restructuring mechanisms to which one has to pay attention. Consequently, mainframe and micro software should be designed to enable the involved knowledge structures to adapt their model of the others through a series of interactions, i.e. instantaneous, complex learning processes are involved. This relational software flexibility requires considerable working and memory space if it is to be applied to large scale IR systems containing databases with up to four to six million records — e.g. Pascal and Chemabs on ESA-IRS. Relevance feedback features like e.g. Zoom point in this direction.

If an intermediary mechanism does not participate the end-user may lack IR knowledge to such an extent that IR becomes a kind of gamble. Besides, they may be unable to 'picture' their own information need — to define the unknown.

So far only a few attempts have been made to produce experimental, automated intermediaries. Meadow *et al.* has created the IIDA package which serves as an IR training device and acts as IR supervisor during real online searching on Dialog [17]. IIDA operates only on (b) — using diagnostics of the interface. Preece has presented an online associative query modification system operating on network representation of controlled vocabulary and thesaurus in very small databases [18]. This system works on parts of (c), not aiming at adapting to the user's ASK. The experimental ASK system by Belkin *et al.* relies mainly on conceptual mapping of data concerned with (c) and (d). It has so far not been applied to large scale systems and no automated device exists to obtain information about the user's ASK — (a). Instead, the ASK statements are regarded as identical to the need 'picture', deriving from a human interview [19].

Traditionally the intermediary systems operate within the framework of the 'best match' principle — as do almost all of today's online services.

The 'best match' principle implies two basic assumptions:

(a) that the individual user is able to specify the information required;
(b) that the information need expressions are functionally equivalent to document texts, i.e. equivalent to information or knowledge.

Both assumptions are basically wrong, as stated by Belkin [19]. Assumption (a) may be correct if an intermediary participates. The keyword is interviewing to obtain knowledge about the problem situation behind the initial request. Assumption (b) is correct in terms of expression-text equivalence. However, texts (or text representation) are not necessarily equal to information/knowledge in relation to the actual, individual search; and need expressions do not have to mirror the actual need [see (a)].

4. The searcher's knowledge structures

Several recent studies of online user communities distinguish between different kinds of intermediaries and end-users. Klasén's investigations demonstrate the importance of IR skills and experience [16]. Howard shows how the interface activity is dependent on knowledge of search language facilities and database structures, i.e. the setting in Figure 2 [20].

A searcher may be categorised as belonging to one of the following four groups, according to the kind of KS possessed:

1. — the 'elite'
   persons possessing both IR and highly specialised conceptual knowledge
2. — the 'intermediary'
   persons with IR knowledge but limited specialised conceptual knowledge
3. — the 'end-user'
   persons possessing highly specialised knowledge but scarce or no IR knowledge
4. — the 'layman'
   persons having neither specialised or IR knowledge in general.

The elite group is small — approximately 15% [16] — and consists mainly of information specialists who may act as intermediaries or end-users within their specialised subject field. As indicated by the survey of UK online users, a large amount of searches can be attributed to this group — perhaps up to 40% [21].

Group 2 may be regarded as the typical intermediary group of today. It consists of librarians, information specialists and others who have obtained IR skills via formal education and/or experience.

All online user studies estimate that a vast majority of today's searches are carried out by intermediaries. Klasén, for example states 75%. In his survey, the notion 'intermediary' refers to group 2 and parts of group 1 and 3.

The typical end-user belongs to group 3. No surveys as yet cover group 4. These persons are the potential future users on the online market; the ordinary citizen. At present it seems that the end-user group is growing faster than the others.

The retrieval result depends on whether there is or may be created accordance — via transfer of data into information — between the searcher's KS, i.e. IR structures and conceptual KS, and the generators' KS, i.e. the setting and the system's conceptual KS [see Figure 2].

The 'elite' searchers are the most fortunate ones. They possess a maximum of IR and conceptual knowledge. According to Howard very experienced searchers with knowledge of the field have the lowest preparation and connect time and highest use of commands, terms searched, sets viewed and the best results [20]. This may indicate that this group is quite efficient in adapting to (learning) the setting and to process the conceptual KS of the system in relation to ASKs.

"It is up to the systems to support the searcher at the interface stage by presenting more simple, flexible and automated (intelligent) facilities. Otherwise, the situation will imply that only an elite group can be trained and operate the IR systems qualitatively, thus inhibiting online penetration to a mass market . . . . "

Klasén's investigation shows some interesting characteristics regarding group 2 and 3. Group 2 accounts for approximately 50% of the online community, all being intermediaries. Group 3 accounts for 35% — out of which 55% are intermediaries and 45% real end-users.

Individual intermediaries (in Sweden) search for approximately two and a half hours per month, with knowledge on approximately four databases and processing two files per search. Within six months they apply two and a half different IR systems, i.e. two and a half command languages and other system's settings. Individual end-users search approximately 0.8 hours monthly, being familiar with approximately 2.3 databases and applying 1.8 database per search. They use less than two host systems half yearly [16]. As may be seen many searchers, especially from group 3, are so called 'casual users.' They show individual search frequencies far too low to keep up with changes in the search language facilities and other setting structures.

Search results then become more expensive and less valid than if carried out by means of other — yet unperceived — system features.

In view of the present state of increasing complexity of IR systems with respect to the contents of overlapping databases, the rapid growth of document representations and full text items as well as the growing complexity of search languages, casual searchers may run into serious, qualitative retrieval problems.

It is up to the systems to support the searcher at the interface stage by presenting more simple, flexible and automated (intelligent) facilities. Otherwise, the situation will imply that only an elite group can be trained and operate the IR systems qualitatively, thus inhibiting online penetration to a mass market, as commented by the UK survey [21].

5. The command language and information

Each command language consists of commands and sub-commands that constitute a consecutive range of search facilities. The complexity of IR, the number of features and the nature of information needed make it extremely difficult algorithmically to establish a 'natural' order or sequence. However, it is possible to piece certain individual command sequences automatically.

Although the rules for command execution are less stringent than in the past, online IR requires exact knowledge on the syntax and impact of the single language. The design presents at a maximum only explanations in relation to (parts of) the setting. In cases of the incorrect use of commands and/or rules the computer responses are far less informative than replies based on correct input. Today's error messages are often inadequate and un-refined, designed to support searchers with a vast IR background — groups 1 and 2, section 4.

With the broadened scope of searchers in mind the online industry has very recently initiated some new ways of applying information technology to the interface situation.

The following approaches — sometimes contradictory, often complementary — are visible:

— the traditional
  (a) amending new command facilities
— the instructive
  (b) intensifying the application of command and search instruction
— the automated
  (c) automating the IR process on semi or full scale.

Almost all hosts supply the traditional view, which is needed to meet the demands for specific searching. This was the reason for introducing free text operators and to a certain extent also crossfile searching. However, each necessary improvement increases the effort needed to interpret the impacts of the new facilities. Casual users, beginners and potential customers are in weak positions. Although search-motivated they seem to stick to basic routines and commands and avoid the more sophisticated ones [22]. As indicated by the surveys, a great deal of system setting data never becomes information [16, 20, 21] — see Figure 1a. Potential users see the languages as a barrier.

The instructive approach opens up a new spectrum. It is mainly caused by the microcomputer environment which in two respects imposes a challenge to the online industry.

Firstly, several independent micro-training packages dealing with online command languages and searching are now on the market. Sequol is a simulation program applied to ESA-Quest consisting of modules and a test quiz [23]. Other packages deal with both commands and the search process [24]. They rely on diagnostics and contain more or less informative instructions in case of errors. They derive from the decade-long experiences with Computer Aided Instruction (CAI).

Secondly, intermediary or expert systems have emerged [17]. Applied during current searching they offer both training and actual IT instruction. However, the complexity of IR makes their diagnostics simple and didactically insufficient for the time being.

One logical result of this approach is to put online training programs on mainframes and to convert diagnostics and instructions into the host computers. Thus, the command language is augmented by more informative computer responses — whereby the demands of the different user groups can be satisfied. By operating on several levels of IR knowledge it is assured that experienced searchers don't get bored and beginners perceive more sufficient information.

The automated approach is only in gestation in relation to operational IR systems. Again, the micro evolution may rapidly change the scenario. Downloading and the manipulation of setting and conceptual data from mainframes to extended microcomputer memory spaces for internal use is a fact today. What is lacking behind this technological enhancement is our full understanding of the IR interface, as stated in relation to Figure 2. It may seem fruitful to recognise downloaded material as representative knowledge structures, which in machine-readable form may serve as manipulative, conceptual networks, usable via upload as basis in iterative retrieval processes. A close R & D collaboration between AI (Artificial Intelligence) and information science may open up new perspectives [25].

What is already becoming operational is semi-automation applied to the setting. Basically two kinds of command features exist; the 'multi-option' type, containing several parallel, independent functions and the 'sequential' type, which — to fulfil their aim — consists of consecutive functions. Only the latter, or combinations of the two, may result in automation and minimisation of commands.

Free text searching using positional operators is a typical example of the 'multi-option' kind of facility. The find command — incorporating Select and Combine functions — and Questindex are 'sequential' features. Zoom shows characteristics of both types.

5.1 Free text searching

Communication and information requests are normally made in natural language and in context. Free text searching offers in principle options by which the searcher may search more specifically, i.e. on broad contexts, expressions and composite words from titles, sources, descriptor fields and most important; from the abstracts.

It is the aspect of contextualisation which makes free text searching attractive, since it decreases the degree of concept ambiguity. It ought to improve relevance because searching and retrieval of a context should assure that the retrieved text means the same to both indexer, author and searcher. Thus, 'potential information' becomes useful 'information' — see Figures 1 and 2. Free text operations are carried out by means of contextual logic using positional operators, truncation and masking. In short they allow for searching on adjacent words, words with up to eight words distance in fixed order, words or expressions placed in the same sentence, searchable, field or citation. Each option has its individual operator. Unfortunately for searchers the operators are designated differently from system to system. Adjacency may for example be communicated correctly to ESA-Quest, CCL and BRS using respectively: (W) — 'space' character — ADJ.

To use the operators to the full requires a vast IR knowledge background. The impact of the different options must be well known beforehand and linguistical elements like spelling, singular/plural forms, common/scientific language and language syntax, synonyms — well understood.

Although adequate in IR nearly all online search studies show a rather low application frequency of free text with positional operators. This may also be due to the fact that a great number of surveys are based on the Eric database and similar well-equipped online files, which are rare on the market [20, 26].

" . . . the lack of online support in the interface situation when applying free text searching. A semantic gap exists between authors' and searcher language. To come up with synonyms and/or closely related concepts and other linguistic elements within a subject area requires in-depth conceptual analysis . . "

As illustrated by studies of search outcome in relation to controlled vocabulary and free text, the only way to maximise the return is to apply both controlled and free text terms [27].

The problem then is the lack of online support in the interface situation when applying free text searching. A semantic gap exists between authors' and searcher language. To come up with synonyms and/or closely related concepts and other linguistic elements within a subject area requires in-depth conceptual analysis [28]. This may suit the 'elite' and 'end-user' categories.

If no controlled vocabulary exists in a file and few feedback mechanisms are available on the system, the outcome depends solely on the searcher's conceptual and setting knowledge. In the Hseline file — for example — produced in Britian, the search statement

S HARBOR ? ?

retrieves eight references while the English version (S HARBOUR ? ?) results in fifty-eight records. With free text and natural language the searcher will retrieve *some* — but not all potentially relevant references. Obviously, it is only by sheer chance that the alphabetical expand list feedback may present adequate help.

If controlled terms are present in a file — but no online thesaurus, i.e. the normal case at present — free text searching will include retrieval of the identical index terms. The problem is slightly different from that above because the searchers may use the record display as feedback mechanism, obtaining preferred terms from the descriptor field.

It is the work of the indexer which may ease the intellectual burden for the searcher. To search optimally, however, the searcher must possess well-developed IR skills — related to both free text searching methods and indexing theory. Thus, this applies to intermediaries and end-users possessing IR knowledge.

To search on controlled vocabulary alone — without an online thesaurus — is conceptually similar to free text searching. The outcome, however, may be of less quality since the entered search statement may give even zero postings — if not present in the controlled vocabulary. For example, by entering 'micro computer' as keyword in NTIS, Pascal and EI Meetings retrieves zero, zero and three references, respectively.

To search on keywords applying online thesaurus seems conceptually easy but requires more IR knowledge than simple use of index terms. The commands and the search methods necessary to apply in combination with free text operations to maximise the result, create very complex search situations. The thesaurus is in itself cognitive structures artificially designed by indexers. Although recognition of concepts and concept relations is more easily performed than memory recall and consequently the searcher conceptually is under less stress than in free text searching alone, he is also further away from the contextualisation in the records and even more remote from the documents themselves. Hence, the intellectual burden required to cope with the complex search process is shifted from conceptual to IR knowledge. This may suggest an explanation as to why the very experienced searchers (the 'elite') achieve the qualitative best results in search environments rich on online aids — as demonstrated by surveys [20].

In the present analysis the impact of factors such as (in)-consistency and specificity have not been explored.

When used at all, the free text functions most frequently used are those basic ones related to adjacency and truncation — other operations are avoided or rarely applied. Not surprisingly, 'quick and dirty' search patterns are applied rather often, as stated by Wormell [22]. A wide gap seems to exist between the free text's operative potentiality and its exploitation.

5.2 Crossfile searching (Questindex)

As for free text searching this facility is designed within the framework of the traditional view and the 'best match' principle. In the ESA-IRS version some automation is built into the feature. Its aim is to support the searcher in the choice of valid databases and additionally to indicate the validity of conceptual search statements in respective files. Crossfile searching is a typical spin-off effect of a rapidly growing information system.

However, like all other operational interface facilities this option is not prompted to the searcher, for example at initiation of a search, although normally at least two files should be searched for exhaustivity reasons. More 'instructive' command languages, e.g. on microcomputer intermediary systems or future mainframes, will hopefully include the prompt and automatically perform the crossfiling. Opposite free text searching crossfile or multifile searching is sequential in its scope. Basically it consists of the following steps:

1. to enter the facility (into e.g. a separate file)
2. to obtain a list of standard topic areas that contain predefined files
3. to choose a topic area (= group of files)
4. to enter the search statement to be crossfiled automatically.

"More 'instructive' command languages, e.g. on microcomputer intermediary systems or future mainframes, will hopefully include the prompt and automatically perform the crossfiling"

If no separate file exists the system must be prompted alternatively to perform the multifile searching. On ESA-Quest this is assured by entering a 'Q' or 'Quest' preceeding the rest of the command string on each step. Time consuming file switching is thus avoided and the searcher remains in his current database. Forgetting to enter the prompt signal, however, results in search execution solely in the current file.

Normally the search statement may consist of a single term, a free text context or an entire saved search.

Basically one has two options for crossfiling (on IRS):

1. to perform an immediate and quick cross-searching
2. to carry out a specially defined cross-searching.

The quick crossfile searching is performed automatically in all multidisciplinary IRS databases, i.e. approximately sixteen million records are scanned in seconds by the computer. This basic option reduces the steps to one — step 4. More important to casual or potential users, the automation adds only one command (including impacts) to the searcher's IR knowledge structures. The second option — the defined one — requires at most the knowledge of three commands if performed correctly: — one to enter the facility and display the topic area list (step 1–2) — one to enter a chosen area from the list — one to enter search statement and execute cross-searching in the predefined files. All three commands must be preceeded by the prompt symbol. In case of a separate file the symbol is avoided — but the file switching command is added.

From a cognitive view two aspects are of importance:

1. the IR knowledge required to enter and to use the facility
2. the conceptual knowledge required.

To enter the feature. Performing quick cross-searching (and entering of search statements) the common search commands must be current knowledge. To display and select a valid topic area from the list and to perform the search across the databases always requires the preceeding prompt symbol. Without computer driven prompts only full-time searchers use the facility sufficiently to know it by heart.

One may note that it is not always mandatory to enter the facility via the command to display the list of topic areas. If headings from the list are known beforehand, it is sufficient to enter the command for selecting a topic, i.e. the predefined group of databases. Consequently, the enlightened searcher has only to rely on two commands — three commands if separate file exists.

To select an appropriate topic area from the standard list requires that the ASK need statement(s) is sufficiently categorised to be related to one of the topics. This problem of matching is in line with situational and categorial classifications of objects as described by Ingwersen [3].

It is in addition a standard index design problem in librarianship among others dealt with by Keen in his investigations on information processing of printed subject entries during searching.[29].

At present the number of topic headings in alphabetical order is small — thirty to forty, i.e. two screen pages at maximum. It does not seem a difficult task for all searcher groups to relate their abstract or situational ASK statements — when defined — to a heading.

To use the feature. Output data from the computer contains database name and number, timespan and number of postings related to the entered search statement for each file.

This implies that the searcher only gets new information concerned with the hit figure and perhaps with the file name/number. No indication of file size is given in any system and it is thus evident that the searcher has to possess knowledge on subject coverages, indexing policies, file structures and sizes, to be able to use the computer reply. Otherwise the output remains data or contains very limited information value.

The file number may be of use for file switching — but only when the posting figure is interpreted. The example below will demonstrate the problem and, in addition, show that common linguistical aspects, discussed under free text searching, are involved.

If a searcher crossfiles, for example, the statement 'micro computer' in the standard 'computer' files, even in free text mode, and obtains few or zero postings in file A and B, he may avoid using these files.

If A and B are the Inspec and NTIS databases, can he afford to eliminate them? In fact the result is very few or zero postings depending on the mode of searching. About seven to ten cross-searched databases, might be avoided, if their names do not trigger any part of the searcher's database knowledge. They become information leading to possible wrong decisions made by the K(S)* — see Figure 1a–b. 1a–b.

Furthermore, cross-file searching is designed — not only to demonstrate the large, perhaps known databases in a field — but certainly also to underline the often rather specialised newcomers. Unfortunately file producers have a tendency to name their products using uninformative acronyms or abbreviations. Besides online novices and casual searchers even experienced customers run into troublesome decoding tasks. Consequently, the easiest way out is to make one's choice among the already known files, whereby potentially useful databases (and information) remain untouched.

In addition to file size, one may suggest that online services add a subject coverage heading to each file name — in order to provide intelligible feedback.

Today, crossfile searching requires a vast IR background on system setting and the retrieval process as well as conceptual knowledge, for example on linguistical matters such as spelling and terminology. Only if possessing the proper KS (the elite searcher), is one able to use the feature to select valid databases and to judge the relevance of the statement(s) presented to the computer. For group 2 intermediaries it becomes merely a tool for file selection and only to some extent an aid as term relevance feedback mechanism.

The end-user with scarce IR knowledge may apply it to choose files. However, he may choose the wrong ones, if a valid file does not (or rarely) contain the search statement. The identical gamble is the concern of the layman.

5.3 Term frequency analysis (Zoom)

As a feedback mechanism Zoom is a consequence of both free text and crossfile searching. Although the feature is a typical 'multi-option' facility, some automation is incorporated.

Zoom is designed to analyse the frequency of single words, phrases or codes appearing in a selected set of references. Up to a maximum of 200 records can be analysed. It is based on the converted versions of the large scale IRS databases. For each database a selected number of searchable fields are made analysable. The following fields — common to all files — can be zoomed:

> Titles, Authors, Corporate sources, Abstracts, Classification codes, Controlled and Uncontrolled terms (=Index terms). In certain databases additional special fields are zoomable, e.g. Molecular formulae and Registry numbers on Chemabs.

Zoom was invented and applied to the ESA-Quest search language in order to improve and support the casual user's search performance [30]. By analysis of a selected set of references the outcome is a ranked list of designations of concepts which in one way or another are associated with each other within the selected pool of documents.

The interesting aspect of Zoom is the fact that the frequency analysis displays a conceptual knowledge structure representing the actual, conceptual relations between the entered search statement and the ranked terms.

If a searcher for example, selects 'micro computer(s)' on Inspec [see the cross-searching sample above] and analyses the fifty latest entered records for index terms, he is sure that *all* the listed concepts are associated with 'micro computer(s)'; what he does not know is the relationships between the listed terms:

```
SELECT MICRO(W)COMPUTER? ?  - then:
ZOOM LATEST                     - CPU reply:

Frq Words/Phrases
----------------------------------------------------

11 MICRO COMPUTER
11 MICROCOMPUTERS
 9 MICRO COMPUTERS
 5 ARCHITECTURAL CAD
 5 COMPUTERISED INSTRUMENTATION
 5 MINI COMPUTERS
 4 COMPUTER SOFTWARE
 4 MICROCOMPUTER
 3 COMMUINICATIONS COMPUTING
 3 COMPLETE COMPUTER PROGRAMS
 3 COMPUTERISED MONITORING
 3 EDUCATIONAL ADMINISTRATIVE DATA PROCESSING
 3 EDUCATIONAL COURSES
 3 ENGINEERING SERVICES
 3 POWER STATION COMPUTER CONTROL
 3 SOFTWARE
 2 ADMINISTRATIVE DATA PROCESSING
 2 BUILDING
...Pages.Lines: More= 22.6
```

Figure 3. *Zoom on index terms in latest 50 refs. on "micro computer(s)". INSPEC, Sept. 1983 (22 screen pages plus 6 lines are still available).*

The ranked list — or conceptual KS — shall be interpreted as follows:

— 'MICRO COMPUTER' and 'MICROCOMPUTERS' occur eleven times each as index terms in the analysed set — together with the search term 'micro computer(s).'
— 'ARCHITECTURAL CAD' occurs five times, 'EDUCATIONAL COURSES' three times.

Since we have zoomed on index terms we can be almost sure that the frequency is equal to the number of references containing the relationship. In other words, five references seem to be concerned with 'micro computer(s)' and 'architectural CAD', etc. Just as important, the terms 'microcomputer(s)' seem also to be used in this set as preferred terms. Thus, by selecting them, the searcher may improve his recall in OR combinations with the original 'micro computers(s).' The feedback makes heuristic searching available, pointing to for example, spelling variants or more specific conceptual relations within a database. The list, however, gives no indications for the existence of a 'building' — 'power station' relationship.

Hence, Zoom may support the searcher in several ways. It may:

* suggest potentially useful, new search entries by displaying
  — preferred index terms [see Figures 3, 4, 6, 7]
  — potential related conceptual aspects [see e.g. 'personal computing' — Figure 4]
  — new terminology [see e.g. 'architectural CAD Figure 3]
  — natural language concepts [see e.g. 'micros' — Figure 5]
  — different spelling (British/US) & variants [see e.g. 'micro computer(s)' Figure 3 & 4 — or the display Figure 8]
  — identical concepts in different languages (translation) [see e.g. Figure 8]
  — synonyms [see e.g. 'human' — Figure 6]
* give ideas and associations in relation to precision of the ASK — serendipity effects
  — make the searcher discover ambiguities and homographs [see e.g. 'animal' — Figure 6]
* work as a relevance feedback tool for conceptual contents of the final set of references [see e.g. Figure 7]
* be an information research tool
  — display conceptual changes in a field [see e.g. 'educational courses' — Figure 3]
  — discovery of new terms to be included in a controlled vocabulary [see e.g. 'home' — Figure 5]
  — bibliometrics
* allow for specific searching [see e.g. 'zero gravity' sample — Figure 6]

The frequency analysis is by default carried out on the basis of fifty records across the last selected set. The computer working time is thus minimal and tests show that fifty records are enough to obtain relevant results. By default the analysis is made on the index terms. In its simplest form the command

ZOOM or Z

thus implies frequency analysis of all index terms — phrases or single keywords — applied to fifty records across the last set. This basic default command results in the feedback device required to obtain potential preferred terms associated with a search statement, see Figures 3 & 4.

Other command functions are for example;

ZOOM LATEST	— analysis of index terms of the fifty latest references in last set, [see Figure 3]
ZOOM TITLE WORDS	— analysis of single title words from fifty records across the last set [see Figure 5]
ZOOM LATEST AUTHORS	— analysis of author names of the latest fifty records in last set
ZOOM 4 (200)	— analysis of the index terms from 200 records across set 4
Z AB, TI WORDS	— analysis of single abstracts and title words.

Below are shown some examples of ZOOMing. This is followed by a discussion of the impacts in relation to the searcher's KS.

5.3.1 Zoom samples

To fill the semantic gap between the searcher's KS and that of the system, one could — prior to ZOOM — look up in the expand list of controlled or free text terms, consult a thesaurus — or display a few records, scanning the title, abstract and index term fields. This gives very limited opportunities as stated previously under free text searching. Let us consider the Questindex result on entering 'micro computer' in Inspec. [It is assumed that Inspec is known to contain in-depth coverage of electronic engineering and consists of 1.9 million references].

First we enter the search argument in free text mode:

FIND MICRO COMPUTER? ? - CPU reply:

1 342 MICRO(W)COMPUTER? ?

Then we Zoom set 1: ZOOM or Z

Computer reply:

```
   Text Analysis Results
Frq Words/Phrases
--- --------------------
 16 MICROCOMPUTERS
  7 MICRO COMPUTER
  6 MICRO COMPUTERS
  6 MICROCOMPUTER
  4 COMPUTER NETWORKS
  4 COMPUTERISED
     INSTRUMENTATION
  4 MINICOMPUTERS
  4 SYSTEMS ANALYSIS
  3 AUTOMOBILES
  3 COMPUTERISED CONTROL
  3 DATA PROCESSING
  3 MEDICAL DIAGNOSTIC
     COMPUTING
  3 MICROCOMPUTER
     CONTROL
  3 PERSONAL COMPUTING
...Pages.Lines: More=20.16
```

Figure 4. *ZOOM on index terms, Inspec, Sept. 83.*

Because of the high frequency of the preferred terms — for example 'microcomputer(s)' — it is rather easy even for beginners or persons with scarce subject knowledge to pin-point the adequate terms and add them to the search.

In addition, some associated terms may be grasped — if their conceptual relation to the search argument is known to the searcher. For example, the term 'personal computing' which may trigger the concept 'personal computer(s)' by memory recall. Hence, entering:

FIND (1 OR MICROCOMPUTER? OR PERSONAL COMPUTER? ?)

where 1 is set 1, results in:

2 16017 MICROCOMPUTER? ?
3 1260 PERSONAL(W)COMPUTER? ?
4 16943 1+2+3

Note, that the OR combination of set 1 and 2 alone gives 16176 items.

The difference between the ranked lists in Figure 3 and 4 derives from the different ZOOM functions applied to the same search statement. In Figure 3 the analysis is carried out on the fifty most recent records. In Figure 4 the list is a result of analysing fifty records *across* the 342 references in set 1. Thus, only each seventh record has been analysed in Figure 4 which — to a certain extent — mirrors the subject coverage in Inspec in relation to 'micro computer(s).'

To be able to obtain new concepts in natural language generated by authors one may zoom for example the set on 'personal computer(s)' — set 3 — on the most recent title words:

```
ZOOM 3 LATEST TITLE WORDS - CPU reply

Frq Words/Phrases
--- -------------------------
 17 PERSONAL
 14 COMPUTERS
 12 A
 12 COMPUTER
  3 HOW
  3 INVITATION
  3 MICRO
  3 MICROCOMPUTER
  2 BIT
  2 FALL
  2 GRAPHICS
  2 HOME
  2 ITS
  2 MARKET
  2 MICROS
  2 NETWORK
  2 NETWORKS
  2 NEW
...Pages.Lines: More= 12.6
```

Figure 5. *ZOOM on INSPEC, Sept. 83.*

Notice for instance 'home' (for home computers) or 'micros' — the slang version of 'microcomputers.'

Synonyms can be spotted and more precise ASK expressions obtained whereby ambiguity can be avoided, via Zoom.

We consider a situation in which the ASK is about 'body factors under zero gravity conditions in space experiments.' It might, however, be expressed in more general terms by the user, for example 'man and zero gravity in space.' We assume that the end-user/intermediary tries the NASA file. Prior to Zoom he might have entered the 'quick and dirty' strategy: Select man; Select Gravity; Combine 1 and 2 to be sure to obtain references. With Zoom in mind, specific searching in free text can easily be performed with a good result:

FIND MAN AND ZERO GRAVITY - reply:

```
1  9240 MAN
2   399 ZERO(W)GRAVITY
3     6 C 1*2
```

Considering the size of NASA (over one million records), its space science coverage and the impact of the entered concept, one may decide to supplement the argument.

Zooming set 3 gives the following list in shortened form:

```
Frq.  Word/phrases

  6   WEIGHTLESSNESS
  2   GRAVITY
  2   HUMAN BODY
  2   HUMAN FACTORS ENGINEERING
  2   MAN
  2   MOTION

  :   :
  :   :
  2   ZERO
  1   ANIMAL
```

Figure 6. *ZOOM on index terms on NASA, Feb. 83.*

As in the former example another preferred term seems relevant: 'weightlessness'. As a synonym to 'man' 'human' seem obvious. Perhaps the more specific term 'human body' catches the eye of the user and maybe he wishes to exclude 'animal(s)' from the search. His ASK may very well be elaborated during this learning and processing process — see Figure 1b — resulting in for example 'human body factors under zero gravity or weightlessness conditions in space, human experiments. (In fact he might change his ASK concept rather radically from the physical 'body factors' into the psychological 'emotion' — triggered by the term 'motion'. Such serendipity effects are not so rare in IR performed by end-users).

After having manipulated and combined sets the searcher may finally apply ZOOM as a relevance tool — not yet displaying any records:

```
Frq Words/Phrases
--- --------------------
 46 HUMAN BODY
 41 WEIGHTLESSNESS
 27 AEROSPACE MEDICINE
 18 SPACE FLIGHT STRESS
 16 PHYSIOLOGICAL
     EFFECTS
 10 PHYSIOLOGICAL
     RESPONSES
 10 WEIGHTLESSNESS
     SIMULATION
  7 BLOOD CIRCULATION
  7 CARDIOVASCULAR
     SYSTEM
  7 HEMODYNAMIC
     RESPONSES
  7 HYPOKINESIA
  6 BED REST
  5 ADAPTATION
...Pages.Lines: More= 4.2
```

Figure 7. *ZOOM on index terms on NASA, Feb. 83.*

"On Chemabs, for instance, one may obtain CAS registry numbers directly from molecular formulae — and vice versa"

(The sample is from a search on 'human body * (zero gravity + weightlessness) — animal(s) 'resulting in 145 references).

5.3.2 Discussion of ZOOM: advantages and limitations

It is the basic default function that allows for initial specific searching that can later be expanded. In databases containing index terms the searcher may enter his statement in a specific way applying free search mode and positional operators. He is now searching all fields including the keyword field, as mentioned in section 5.1 above. Even if the number of postings is very low, he will obtain relevant index terms from the default zoom and if recombining his initial statement with the obtained keywords — again using free text searching — the result is always improvement of relevance [see Figures 6, 7]. The searcher is consequently applying the index terms as a switching language between the entered natural language concepts and preferred term phrases that later in the search session eventually may be zoomed for title and abstract terms or entered as natural language concepts [see Figure 5].

In files only consisting of abstracts and titles, this 'switching' is more difficult to perform. Hence, it is difficult to obtain spelling variants and synonyms via a direct zoom. This is a limitation which may be overcome — to some extent — by using classification codes (if existing in the file) as 'switching mechanism'.

As such, the knowledge of the impacts of the Zoom functions may change the online search patterns radically.

If a searcher, for example, is applying the analytical search mode put forward by Vigil [31], operating on discrete displays formulated by using Boolean NOT logic, he easily obtains via Zoom information the relevance of the intermediate set combinations — as well as new vocabulary ideas.

It is interesting to observe that some Zoom functions may serve as an instrument for direct fact retrieval in bibliographic files. On Chemabs, for instance, one may obtain CAS registry numbers directly from molecular formulae — and vice versa. On all files zooming a set on authors, corporate sources or classification codes can quickly provide the searcher with example clusters of prominent authors within the subject, their name form variations, the institutional names related to a particular research area or the important class codes assigned to a specific topic. A high frequency will in general indicate a close relationsip to the search argument in the zoomed set or demonstrate main topic areas within the field of interest. Preferred terms which may supplement the searched one(s) are normally placed on the top of the list.

This makes Zoom quite relevant for all user groups to apply, especially intermediaries without specialist knowledge of the subject.

An advantage is the language translation aspect for bilingual (mostly European) databases, such as CISDOC on occupational health produced by ILO which contains abstracts, titles and controlled terms in French and English — see Figure 8. However, persons with scarce subject or language potential can only apply the high ranked terms, i.e. to them zoom is merely a tool for recall improvements.

```
                      Text Analysis Results

 Frq Words/Phrases                Frq Words/Phrases
 --- ---------------------        --- ---------------------
  39 CAS 1332 21 4                       VALUES
  30 POUSSIERES                     7 ANIMAL EXPERIMENTS
       RESPIRABLES                  7 RECHERCHES SUR L
  30 RESPIRABLE DUST                     ANIMAL
  27 AMIANTE                        7 USA
  27 ASBESTOS                       6 CHRYSOTILE
  11 ASBESTOSE                      6 COMPARATIVE STUDY
  11 CARCINOGENS                    6 DIAGNOSTIC
  11 SUBSTANCES                     6 ETUDE COMPAREE
       CANCEROGENES                 6 INDUSTRIE DE L
  10 ASBESTOSIS                          AMIANTE
   9 VALEURS SEUILS                 6 MESOTHELIOME PLEURAL
   8 CANCER DU POUMON               6 MESURE DE L
   3 CAS 12001 29 5                      EMPOUSSIEREMENT
   8 CONFERENCE                     6 PLEURAL MESOTHELIOMA
   8 LEGISLATION                    6 TALC
   8 LUNG CANCEF                    5 ASBESTOS PROCESSING
   8 THRESHOLD LIMIT                     INDUSTRY
...Pages.Lines: More= 12.13
```

Figure 8. *ZOOM on index terms on CISDOC, March 84. Search terms: Asbest? (755 items).*

Low frequencies must not be overlooked. It is down on the lists that sub-areas and aspects are hidden — not only in relation to the executed arguments but also associated with the other concepts displayed.

However, as special subject idea bank, Zoom only serves conceptually specialised persons, i.e. the elite searchers and end-users.

Furthermore one must be aware of a certain unreliability concerned with the lower and low frequency terms — especially in relation to search statements with high postings in the database. When the analysis is executed across fifty references of the selected set, the terms ranked below ca. 40% of the number of analysed references are not forming a stable pattern but may change for each update. This is naturally evident when the analysis is executed on the fifty (or 200) most recent records of the set.

Figure 9 shows the same search as performed in February 1983 on NASA — with the displayed frequency analysis shown on Figure 7 — now executed in March 1984.

As can be observed, 'space flight stress' has fallen down and 'physiological responses' disappeared from this first part of the list. Instead 'manned space flight' has entered. The differences between the two analysis displays derives from the fact that two different groups of fifty references are analysed, since the retrieved set of 1984 contains three records more than the set of 1983.

```
Fr. Words/Phrases
47 HUMAN BODY
40 WEIGHTLESSNESS
26 AEROSPACE
   MEDICINE
11 PHYSIOLOGICAL EFFECTS
11 WEIGHTLESSNESS
   SIMULATION
10 HUMAN
11 BODY
 9 SPACE FLIGHT
   STRESS
 8 MANNED SPACE
   FLIGHT
 7 CARDIOVASCULAR
   SYSTEM
 7 PHYSIOLOGY
 6 BED REST
```

Figure 9. *ZOOM on index terms NASA, March 1984: same search as for figure 7; 148 refs.*

The most obvious limitations for users is the layout design. It has been suggested that data concerned with the database in which the analysis is carried out ought to be shown automatically together with the analysis results. Likewise for the analysed set number and the analysis parameters. At present it is necessary to prompt the computer to give this basic information.

It has also been suggested by users to supply the frequency list with accessible sequential numbers, similar to those for expand lists, i.e. Z1-Z2-Z3 . . . There are two reasons for not doing this:

(a) free text operations using relevant term phrases are impossible to execute without a separate selection of the phrases that include adjacency and (sometimes) truncation. Adjacency might be performed algorithmically though.

(b) just as important is the kind of search logic to apply in relation to the already entered and retrieved concepts and the question of which (parts of) expressions or derivations to use in the consecutive search process. It is not as simple as the purely alphabetically ordered terms in expand lists.

A final obstacle in employing a feedback mechanism like Zoom may be a tendency towards 'information overload' for intermediary searchers whereby relevant data with lower frequency might be overlooked — never becoming information. It is, however, the recognition facets which make Zoom such a powerful interface mechanism on a operational system. As stated earlier in relation to Figure 1, recognised concepts may carry with them new, related ones or trigger memory recall activity and thinking..It is the receiving searcher's KS which must establish the valid associations. Besides the basic and simple zoom command a great number of command combinations exist. To grasp the impact of them all is a heavy task. Hence, Zoom is a thrilling but also comprehensive intellectual challenge to searchers at all levels.

"Present IR systems are designed almost exclusively for IR experienced end-users and intermediaries. However, the scope of command languages and other interface devices is being expanded . . . . "

7. Concluding remarks

Present IR systems are designed almost exclusively for IR experienced end-users and intermediaries. However, the scope of command languages and other interface devices is being expanded embedded in instructive and automative approaches to information. Although the traditional approach is not outdated, it is the former viewpoints which most likely provide more constructive, flexible and usable answers to how to design IR systems and apply information technology — on mainframes or micros — diminishing the knowledge gap between the experienced and the casual user or the new customer.

It is not only a question of awaiting new generations of computers, expecting new hardware and software to solve contemporary problems. The Alvey report [32], for example, is rather traditional in this respect, correctly aiming at intelligent knowledge-based systems, but merely overlooking the human factor — the socio-psychology of information.

After all, IR is a problem solving tool — the means to access data and obtain the right piece of information at the right time for the right person.

8. References

[1] R.C. Shank and R.P. Abelson, 'Scripts, plans, goals and understanding: an inquiry into human knowledge structures,' Hillsdale, N.Y., Erlbaum, 1977.

[2] M. De Mey, 'The relevance of the cognitive paradigm for information science. Theory and application of information research,' *Proceedings of the SIRE 2 conference*, 1977, Copenhagen. ed. by Ole Harbo and Leif Kajberg. London, Mansell, 1980, pp. 48–61.

[3] P. Ingwersen, 'Search procedures in the library analysed from the cognitive point of view,' *Journal of Documentation*, **38**, 3, Sept. 1982, pp. 165–191.

[4] B.C. Brookes, 'The developing cognitive viewpoint in information science,' International Workshop on the Cognitive Viewpoint, 1977, ed. by Marc De Mey. University of Ghent, 1977. pp. 195–203.

[5] D.M. Mackey, 'What makes a question,' *The Listener*, **63**, 5 May 1960, pp. 789–790.

[6] N.J. Belkin, 'Information concepts for information science,' *Journal of Documentation*, **34**, 1, March 1978, pp. 55–85.

[7] B.C. Brookes, 'The foundations of information science, Part I: Philosophical aspects,' *Journal of Information Science*, **2**, 3, 4, 1980, pp. 125–133.

[8] R.C. Shank, 'The role of memory in language processing,' The structure of human memory, ed. Charles N. Cofer. San Francisco, 1975, pp. 162–189.

[9] D.A. Norma and D.G. Bobrow, 'On the role of active memory processes in perception and cognition,' The structure of human memory, ed. by Charles N. Cofer, San Francisco, 1975, pp. 114–132.

[10] J.D. Novak, 'Concept-based learning,' Presented at the Amherst Symposium on the Optimum Utilisation of knowledge, Amherst, Mass. 5–8 Nov. 1981, 22pp.

[11] D.B. Gowin, 'The structure of knowledge,' *Educational Theory,* **20**, 4, 1970.

[12] K.R. Popper, 'Objective knowledge: An evolutionary approach,' Oxford, Clarendon Press, 1973.

[13] M. Merleau-Ponty. 'Le visible e l'invisible,' Paris, Gallimard, 1964.

[14] R.S. Taylor, 'Question-negotiation and information seeking in libraries,' *College and Research libraries,* **29**, 3, 1968, pp. 178–194.

[15] O. Harbo, P. Ingwersen and P. Timmermann, 'Cognitive processes in information storage and retrieval,' International Workshop on the Cognitive Viewpoint, 1977, ed. by Marc De Mey, University of Ghent, 1977, pp. 214–218.

[16] L. Klasen, 'Online-använding,' Rapport. DFI, 1982, Stockholm, Delegationen för vetenskapling och teknisk informationsförsörjning, 1982, 1, pp. 80–128. ISSN 0349-9324 (Engl. title: Online use).

[17] C.T. Meadow *et al.*, 'A computer intermediary for interactive database searching, part I: Design,' *Journal of the American Society for Information Science,* Sept. 1982, pp. 325–332. 'Part II: Evaluation,' *Journal of the American Society for Information Science,* Nov. 1982, pp. 357–364.

[18] S.E. Preece, 'An online associative query modification methodology,' *Online Review,* **4**, 1980, pp. 375–382.

[19] N. Belkin, R.N. Oddy and H.M. Brooks, 'ASK for information retrieval: Part I & II,' *Journal of Documentation,* **38**, 2/3, 1982, pp. 61–71, 145–164.

[20] H. Howard, 'Measures that discriminate among online searchers with different training and experience,' *Online Review,* **6**, 4, 1982, pp. 315–327.

[21] J. Deunette and Hall. 'Survey of UK online users: A report on current online usage,' London, The Online Information Centre, Aslib, 1983.

[22] I. Wormell, 'Computerised statistical research study on ESA-Quest command use,' Interim Report, Unpublished, ESA-IRS, 1981.

[23] J. Scoltock, 'Sequol: A simulation of ESA-Quest on-line,' University of Sheffield, 1982, 199pp. (Thesis + computer program).

[24] R. Waldstein, 'Diatom: A Dialog simulator,' *Online,* **5**, 3, 1981, pp. 68–72.

[25] L.C. Smith, 'Artificial intelligence applications in information systems,' *Annual Review of Information Science and Technology,* **15**, 1980, pp. 67–105.
R. Hjerppe, 'What artificial intelligence can, could, and can't, do for libraries and information services,' 7th International Online Information Meeting, 1983. Oxford, Learned Information, 1983, pp. 7–25.

[26] K. Markey, P. Atherton and C. Newton, 'An analysis of controlled vocabulary and free text search statements in online searches,' *Online Review*, **4**, 3, 1980, pp. 225–236.

[27] M.L. Calkins, 'Free text or controlled vocabulary? A case history step-by-step analysis . . . plus other aspects of search strategy,' *Database*, **3**, 2, 1980, pp. 53–67.

[28] T.E. Doszkocs, 'Implementing an associative search interface in a large online bibliographic database environment,' New Trends in Documentation. 39th FID Congress, Edinburgh, 1978. Aslib, 1980, pp. 295–297.

[29] E.M. Keen, 'On the processing of printed subject index entries during searching,' *Journal of Documentation*, **33**, 4, 1977, pp. 266–277.

[30] W.A. Martin, 'Helping th less experienced user,' *Proceedings of the 6th International Online Information Meeting, 1982*. Oxford, Learned Information, 1982, pp. 67–76.

[31] P.J. Vigil, 'Analytical methods for online searching,' *Online Review*, **7**, 6, 1983, pp. 497–514.

[32] Alvey Committee. 'A programme for advanced information technology,' Report of the Alvey Committee, London, HMSO, 1982.

This article has been cited by:

1. Orland Hoeber, Xue Dong Yang. 2008. Evaluating WordBars in exploratory Web search scenarios. *Information Processing & Management* **44**, 485-510. [CrossRef]
2. Kalervo Järvelin, Peter Ingwersen, Timo Niemi. 2000. A user-oriented interface for generalised informetric analysis based on applying advanced data modelling techniques. *Journal of Documentation* **56**:3, 250-278. [Abstract] [PDF]
3. XIAOWEN FANG, GAVRIEL SALVENDY. 2000. Keyword comparison: a user-centered feature for improving web search tools. *International Journal of Human-Computer Studies* **52**, 915-931. [CrossRef]
4. Ying Ding, Gobinda G. Chowdhury, Schubert Foo, Weizhong Qian. 2000. Bibliometric information retrieval system (BIRS): A web search interface utilizing bibliometric research results. *Journal of the American Society for Information Science* **51**:10.1002/1097-4571(2000)9999:9999<>1.0.CO;2-Z, 1190-1204. [CrossRef]
5. Harry Bruce. 1998. User satisfaction with information seeking on the internet. *Journal of the American Society for Information Science* **49**:10.1002/(SICI)1097-4571(1998)49:6<>1.0.CO;2-J, 541-556. [CrossRef]
6. Mohamed Abdel-Mottaleb, Hsiang-Lung Wu, Nevenka Dimitrova. 1996. Aspects of multimedia retrieval. *Philips Journal of Research* **50**, 227-251. [CrossRef]
7. Efthimis N. Efthimiadis. 1995. User choices: A new yardstick for the evaluation of ranking algorithms for interactive query expansion. *Information Processing & Management* **31**, 605-620. [CrossRef]
8. N.J. Belkin, P.G. Marchetti, C. Cool. 1993. BRAQUE: Design of an interface to support user interaction in information retrieval. *Information Processing & Management* **29**, 325-344. [CrossRef]
9. Barbara A. Norgard, Michael G. Berger, Christian PlauntThe Online Catalog: From Technical Services to Access Service 111-148. [Citation] [PDF] [PDF]
10. PETER INGWERSEN. 1992. Information and Information Science in Context. *Libri* **42**. . [CrossRef]
11. Mary M. Huston. 1991. Windows into the search process: an inquiry into dimensions of online information retrieval. *Online Review* **15**:3/4, 227-243. [Abstract] [PDF]
12. E. MICHAEL KEEN. 1991. THE USE OF TERM POSITION DEVICES IN RANKED OUTPUT EXPERIMENTS. *Journal of Documentation* **47**:1, 1-22. [Abstract] [PDF]
13. EFTHIMIS N. EFTHIMIADIS. 1990. ONLINE SEARCHING AIDS: A REVIEW OF FRONT ENDS, GATEWAYS AND OTHER INTERFACES. *Journal of Documentation* **46**:3, 218-262. [Abstract] [PDF]
14. ROY DAVIES. 1989. THE CREATION OF NEW KNOWLEDGE BY INFORMATION RETRIEVAL AND CLASSIFICATION. *Journal of Documentation* **45**:4, 273-301. [Abstract] [PDF]
15. PETER INGWERSEN, IRENE WORMELL. 1988. Means to Improved Subject Access and Representation in Modern Information Retrieval. *Libri* **38**. . [CrossRef]
16. Marcia J. Bates. 1986. Subject access in online catalogs: A design model. *Journal of the American Society for Information Science* **37**:10.1002/asi.v37:6, 357-376. [CrossRef]
17. Tamas E. Doszkocs. 1986. Natural language processing in information retrieval. *Journal of the American Society for Information Science* **37**:10.1002/asi.v37:4, 191-196. [CrossRef]
18. Roy Davies, Alastair G. Smith, Anne MorrisChapter 4. Experts systems in reference work . [CrossRef]