Syracuse University

From the SelectedWorks of Robert Oddy

1992

Towards the use of situational information in information retrieval

Robert N Oddy, *Syracuse University*
Elizabeth D Liddy, *Syracuse University*
Ann Bishop, *Syracuse University*
Eileen Martin, *Syracuse University*
Joseph Elewononi, *Syracuse University*, et al.

Available at: https://works.bepress.com/robert_oddy1/18/

TOWARDS THE USE OF SITUATIONAL INFORMATION IN INFORMATION RETRIEVAL

ROBERT N. ODDY, ELIZABETH DUROSS LIDDY, BHASKARAN BALAKRISHNAN, ANN BISHOP, JOSEPH ELEWONONI *and* EILEEN MARTIN

School of Information Studies, Syracuse University Center for Science and Technology, Syracuse, NY 13244, USA

DOI: 10.1108/eb026894 **(Permanent URL)**

This paper is an exploratory study of one approach to incorporating situational information into information retrieval systems, drawing on principles and methods of discourse linguistics. A tenet of discourse linguistics is that texts of a specific type possess a structure above the syntactic level, which follows conventions known to the people using such texts to communicate. In some cases, such as literature describing work done, the structure is closely related to situations, and may therefore be a useful representational vehicle for the present purpose. Abstracts of empirical research papers exhibit a well-defined discourse- level structure, which is revealed by lexical clues. Two methods of detecting the structure automatically are presented: (i) a Bayesian probabilistic analysis; and (ii) a neural network model. Both methods show promise in preliminary implementations. A study of users' oral problem statements indicates that they are not amenable to the same kind of processing. However, from in-depth interviews with users and search intermediaries, the following conclusions are drawn: (i) the notion of a generic research script is meaningful to both users and intermediaries as a high-level description of situation; (ii) a researcher's position in the script is a predictor of the relevance of documents; and (iii) currently, intermediaries can make very little use of situational information. The implications of these findings for system design are discussed, and a system structure presented to serve as a framework for future experimental work on the factors identified in this paper. The design calls for a dialogue with the user on his or her position in a research script and incorporates features permitting discourse-level components of abstracts to be specified in search strategies.

1. INTRODUCTION

INFORMATION RETRIEVAL RESEARCH in the context of scholarly or professional work is driven by a difficult practical problem: that of extricating a manageable number of texts from a very large body of documents (such as the literature of a discipline) in response to an information need in the mind of a scientific, professional, technical, or scholarly worker. The purpose of an information retrieval system is to facilitate the communication of knowledge; but it is not concerned with direct communication between source and destination. An information retrieval system is like a gigantic bulletin board: authors deposit their messages, recipients search for useful ones. Research on the information retrieval problem in the past has, not unnaturally, been predominantly topic-oriented. That is it assumes that the enquirer will express an information need as a topic, and looks for ways of finding mention of that topic in texts. Beginning with a tiny trickle which surfaced during the 1960s, a stream of thought has developed, which holds that users' problems and situations are essential ingredients of information needs, which we should attempt to recognise in retrieval system design (see reviews by Dervin and Nilan [1] and Hewins [2]). Perusal of recent information retrieval literature will show that topic-oriented research is, nevertheless, still predominant.

In the course of the work that we shall describe in this paper, it has become clear that the use to which information will be put, which is intimately related to the situation in which a user finds him- or herself, is an important factor in properly understanding statements made by the user, and is an important determinant of relevance judgements. User-modelling has received significant attention in recent literature, for example Daniels [3] and Kosba and Wahlster [4], and there are some experimental IR systems that incorporate them [5, 6]. However, it is important for IR to differentiate between the concepts of situational information and user-modelling.

It seems, therefore, that we need to develop models of situations, and also ways of relating texts and natural language utterances to those models. This aspect of the problem appears not to have received significant attention: how can textual documents be represented in automatic systems in such a way that their relationship to users' problems is made clear? In the information retrieval context, the texts are usually abstracts, and the other utterances of importance are users' problem statements (including, but not limited to, queries). We have made use of some of the techniques of discourse linguistics to study the problem from this perspective. Much of human discourse is concerned with sharing experience of the situations with which we must cope. It is directed towards improving our ability to recognise common situations, and to respond effectively to them. Specifically, document abstracts are written with the intention of informing people who belong to the same community as the authors and are engaged in similar work. They should, therefore, be able to recognise common situations. The situations of interest to an author are discernable in the discourse-level structure of an abstract. Similarly, we should look in users' utterances for indications of their problematic situations. Then users can be put in touch with related situations of others as represented in texts.

In order to approach this problem on a broad front, we set up a one-year, multi-faceted project.* We decided that our first priority was to gain insights on a number of potentially troublesome aspects, before investing our efforts in system implementation. In this paper, therefore, we will cover a wide territory, sometimes much less thoroughly than we would wish, but hoping to convey our impressions of the scope of the problem and to give the reader a basis for judging the prospects of the approach. Our account is organised into three major sections dealing with: the structure of document abstracts; users' expressions of situations and information need; and implications for information retrieval system design work. We will present a discourse-level structure that is apparent in the abstracts of empirical research papers, describe how this structure is revealed by lexical clues in texts, and present our first attempts to recognise the structure automatically. Although we limit our attention to the one type of document, namely empirical abstracts, it should be noted that this covers an enormous domain of human knowledge. Empirical abstracts typically contain brief statements of hypotheses, methods, results, conclusions and other matters related to the conventional conduct of empirical research. We suggest that the representation of documents in terms of structural components like these will enable us to introduce models of the problematic situations of users, who are presumably working within the same conventions. Our efforts to follow a parallel path with transcripts of users' oral problem statements did not succeed. We did not discover a discourse-level structure describing the information search problem of individuals. (There are discourse-linguistic explanations for this that occurred to us in retrospect.) We therefore took the path of in-depth open-ended interviewing of end-users and search intermediaries in order to learn how the user's situation in the (empirical) research process influences search strategies and relevance judgements. We represent the research process as a generic empirical research script, which has a great deal in common with the discourse-level structure of empirical abstracts.

We will not elaborate an information retrieval system design in this paper, but we have some suggestions for features and general structure of systems which incorporate a discourse-level treatment of texts and cognisance of users' situations through a research script. We are not suggesting at this juncture an increase in the intelligence exhibited by IR systems, but rather the incorporation of new types of information with suitable tools for users to appreciate and manipulate them. Then new behaviours can be observed and we can evaluate the potential of these ideas *in the IR situation.*

2. STRUCTURE OF EMPIRICAL ABSTRACTS

2.1 Discourse linguistics

In our analysis, we take a linguistic approach that emphasises discourse level processing and the pragmatic interpretation of discourse, rather than the syntactic and semantic levels of processing usually found in IR systems.
Discourse linguistics is concerned with all forms of natural language that involve multiple sentences or utterances. It may involve both syntactic and semantic processing, but most importantly, sentences are interpreted in light of the other sentences that comprise the discourse [7]. Pragmatics is concerned with analysis of language as it is used by individuals in specific contexts to achieve specific purposes [8]. Discourse structure can be defined as the textlevel syntactic organisation of semantic content. Structure is the recognizable form or organisation that is filled with different meaning in particular examples of that text type. A structure consists of components and the relations among these components. In discourse structure, the components are those categories of content which define the text type. The reason that text types have perceivable structures is that conventions of form evolve over time and it is efficient for both producers and receivers of these conventional text types to make use of their rather fixed structure [9-11]. It has also been observed [7, 10] that for some text types the discourse structure reflects the structure of the problems or tasks in which their originators are engaged.

The early emphasis in information retrieval on choosing individual index terms to represent the contents of a document deflected attention from the more holistic view of a document that discourse analysis suggests. Recent research, however, has seen the emergence of a concern for the natural structure of informative content at the discourse level [12-16]. Liddy's [17,18] analysis of the structure of empirical abstracts is a precursor to the work reported in this paper, and will now be outlined.

TABLE 1. *Components ordered using data from Liddy's Phases I and II*

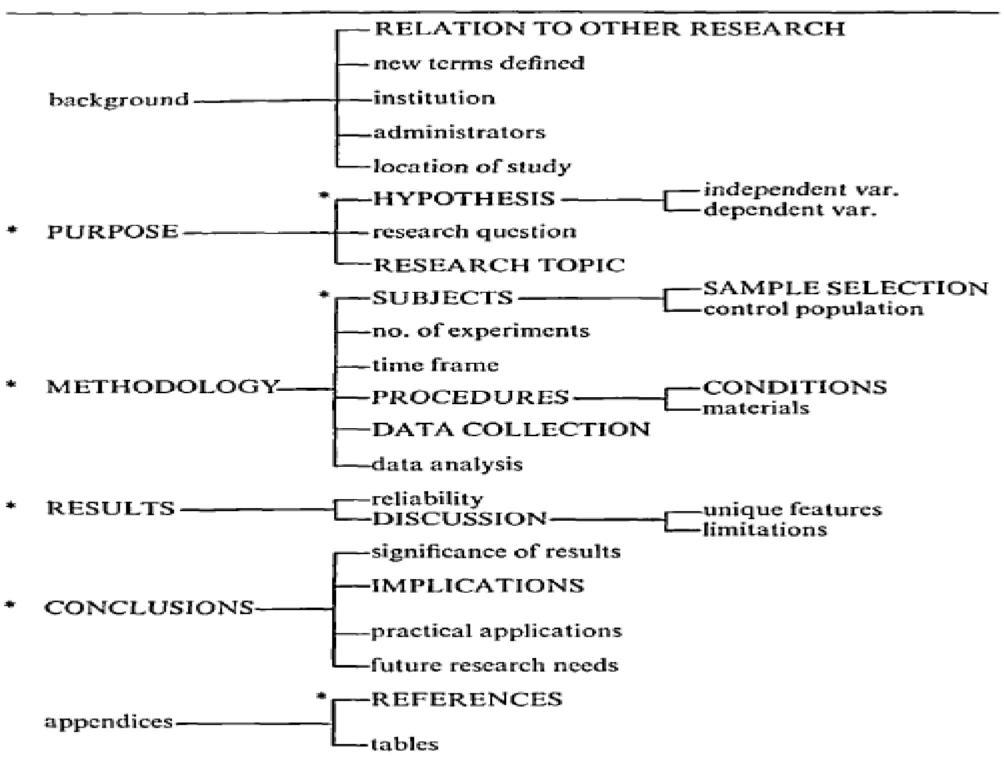
S	*Component name*	*Typicality* Phase I	*Typicality* Phase II
C	subjects	.96	.95
C	results	.82	.89
C	purpose	.84	.86
C	conclusions	.77	.45
C	methodology	.89	.26
C	references	.57	.59
C	hypothesis	.93	.18
P	procedures	.42	.57
C	relation to other research	.60	.22
C	implications	.62	.17
C	conditions	.45	.28
E	data collection	.24	.47
A	research topic	—	.72
C	discussion	.54	.09
C	sample selection	.53	.10
E	dependent variable	.27	.28
E	independent variable	.25	.28
E	data analysis	.30	.21
P	tests	.50	.01
E	location of study	.25	.24
C	practical application	.44	.05
P	no. of experiments	.32	.16
P	research question	.32	.15
P	drugs administered	.44	—
E	time frame of study	.30	.14
E	background	.24	.16
E	appendices included	.25	.13
E	future research needs	.34	.04
P	significance of findings	.34	.04
P	control population	.32	.04
E	tables included	.23	.13
P	apparatus	.31	—
E	institution	.26	.01
P	scope	.23	—
E	unique features	.23	.01
E	intended audience	.21	—
P	materials	.19	.02
E	confounding variables	.19	—
E	limitations	.18	.01
E	administrators	.18	.01
E	new terms defined	.14	.03
E	reliability	.15	.01
E	subsequent research	.15	—

S column: Source of component, indicates whether generated by:
P – PsycINFO only; E – ERIC only; C – common; A – added

2.2 Liddy's study

In order to follow the work reported in this paper, it is necessary to know something of the research done by Liddy [17, 18] on the discourse-level structure of empirical abstracts. Her investigation was conducted in three phases. Phase I made use of four cognitive tasks, performed by expert abstractors, from the ERIC and PsycINFO services, which provided insights into their internalised knowledge of the structure of empirical abstracts. Phase II was linguistic analysis of a sample of 276 empirical abstracts, from the ERIC and PsycINFO databases, aimed at determining whether and how the experts' suggested structure was evidenced in abstracts. Phase III served as a validation of the model of the structure of empirical abstracts that arose from a synthesis of the results of Phases I and II.

Table 1 presents a combined view of the outputs from Phases I and II. The Phase I column provides for each component a relative indication of abstractors' tendencies to generate it and regard it as typical. The Phase II column shows the relative frequency with which each component generated by the abstractors in Phase I actually evidenced itself in the 276 abstracts analysed in Phase II. Based on these figures, and relationships identified in Phase I by the abstractors, Liddy built up a hierarchical structure at three levels of detail, or typicality, which is reproduced in Figure 1. This diagram includes all components generated by either access service which occurredmore than once in the sample of abstracts. The *Prototypical* model (starred components only) is the most general model, but perhaps of limited use because it lacks sufficient detail. In contrast, the *Elaborated* model (all components) may prove useful for fields whose research is similar in methods and style to that covered by ERIC and PsycINFO, but is too detailed to generalize to less similar disciplines. In addition, the results of our preliminary attempts to automate the detection of structure in individual abstracts suggest that the Elaborated model may be too refined a structure for automatic detection.



* – Prototypical component; UPPER-CASE lettering – typical component; lower-case lettering – elaborated component.

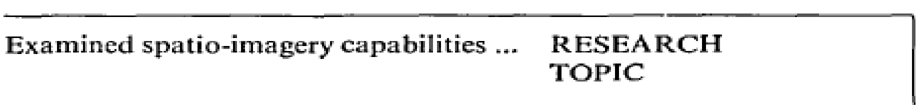
FIGURE 1. *Structure of empirical abstracts*

The *Typical* structure of empirical abstracts (all upper-case components in Figure 1) suggests itself as the structure which would be most useful in applying the model to most areas of empirical research. It provides enough structure for the specific components needed in other disciplines to be easily added in appropriate places. In terms of specificity, it also appears to be the appropriate level at which useful representations could be provided to users of a retrieval system. (However, all our work on recognition of discourse-level structure of abstracts so far has focussed on the Elaborated structure.)

As part of Liddy's manual analysis of the 276 abstracts, concordances of lexical clues were developed for each component having ten or more occurrences in the sample of abstracts. These word stems serve as clues to the fact that a segment of text is functioning as a particular component. There is a core semantic factor in each component's set of clue terms because these clues embody in some sense the activity germane to a particular aspect of the research process. This semantic similarity is exemplified in the set of lexical clues observed in the sixty-nine instances in which the HYPOTHESIS component occurred in the sample. 90% of the occurrences of the HYPOTHESIS component can be detected by the presence of one of a set of clues sharing the semantic feature of the act of conjecturing (e.g. hypothes-, predict-, test-). Liddy found that a relatively small number of such clue stems account for most component occurrences.

Having developed the linguistic model, Liddy conducted a validation experiment using some of the abstractors. In terms of a somewhat complex measure of the differences between the structuring based on the linguistic model and that intuitively preferred by abstractors, the rate of disagreement was only 14%. Some patterns were evident in the differences. In particular, 44% of the differences arose from the abstractors not using multi-level structuring. For example, a common structuring based on the linguistic model is such as this:

Here, the text begins with a few words belonging to the PURPOSE component (e.g. 'Tested...'; 'Examined...'; 'Assessed...'; etc.) before dropping to the lower level of RESEARCH TOPIC. Abstractors, however, usually did not assign PURPOSE and included the introductory terms as part of the
RESEARCH TOPIC, as follows:



Abstractors' preference for middle hierarchy labels supports theoretical work done in classification of other objects by Rosch and colleagues [19] who found that within taxonomies there is generally one level of abstraction at which most individuals tend to make category cuts. This is referred to as the basic level and it tends to occur in the categories at the middle of the hierarchy.

2.3 Automatic recognition of structure

Liddy's work laid a foundation for the incorporation of the discourse-level structure of documents in information retrieval system databases. For abstracts of empirical research papers, at least, this structure exists in the minds of those who write them, and is associated with a number of recognisable linguistic clues. However, the step to a fully automatic structure recognition program is not a small one, and thus exploration of this problem is an important part of the present study. We continued to use Liddy's corpus of abstracts in order to take advantage of her substantial investment of effort in marking up all the texts with their Elaborated structure.

The clues to discourse-level structure seem

1. to span a full range of levels (from morphological to lexical to relationships between structural components);
2. to be probabilistic;
3. to interact with each other; and
4. to be very numerous, even within limited domains.

Lexical clues are essentially templates. For example: 'indicate that ...' usually introduces the results; 'appeared to ...' sometimes signals an interpretation of a result. Morphological clues can also be used: for example, tense endings of verbs can sometimes distinguish between well-founded statements (results of past research) and speculation (hypothesis). Other clues are semantic classes (e.g. animate nouns often indicate experimental subjects, a sub-component of methodology), and expected orderings of components within the text. Another aspect of the structural analysis is to establish the scope of a component - how much of the text should be included in it. Sources which are useful in these decisions are both 'cohesion' clues which indicate the flow from one component to another, and 'continuation' clues which indicate that a sentence is a continuation of the component started in the previous one.

If we were to assemble a set of deterministic rules for this analysis, we would expect a very large number of complex rules, most of which were needed to deal with exceptions to simpler ones. Moreover (extrapolating from experience in linguistics and information retrieval), there would probably be a 'diminishing returns' situation, in which the addition of more complexity produced very little gain in effectiveness. Another approach is to view rules in this domain as probabilistic. They contribute evidence toward the acceptance of hypotheses about the components of a text. The processing of the text terminates, in successful cases, when the overall likelihood of a candidate structure achieves dominance over others. The goal is to evaluate many alternative hypotheses concurrently, using a multitude of rules about clues. Two approaches to programming the required text analysis, which are appropriate to this situation, will be given here. The first is a probabilistic method, making use of Bayes' theorem to incorporate the evidence afforded by lexical and other features. The other is a 'neural network' approach which, in principle, allows the many different factors simultaneously to influence the outcome of the analysis.

2.3.1 Probabilistic text analysis

The clues to discourse-level structure of empirical abstracts, which seem to be at the same time the most important and the most straightforward, are single words or stems occurring in the text. We consider an abstract, *A,* to be divided into a sequence of text-pieces: *A1, A2,... An.* Each text-piece *Ax* is composed of one or more words, some of which are potential clues, such as those identified by Liddy [17]. We suppose that *Ax* can belong to exactly one of the thirty-seven possible components, C1, C2, ...C37, in the Elaborated model. (As we shall see later, this is a somewhat problematic assumption.) We shall denote the potential clues for a component Cj in the text-piece *Ax* by e1, *e2, ...er,* where there is no constraint on the *e,s* to be unique. The identification of a set *e1, e2,... er* in *Ax* is an event that we shall call *E.* Our problem is to estimate, for each *Ax,* the thirty-seven probabilities, P(CJ|E), that *Ax* belongs to each component, *Cj*($1 \leq j \leq 37$).

Applying Bayes' Theorem, we have:

$$P(Cj|E) = P(Cj) \times P(E|Cj)/P(E)$$

In this expression, *P(Cj)* is the prior probability that a text-piece is contained in component Cj, *P(E|Cj)* is the probability that the event *E* occurs, given that the text-piece *Ax* is in component Cj, and *P*(*E*) is the probability of event *E* occurring. If we assume that the clues comprising *E,* namely e1, *e2, ...er* occur independently of one another, then:

$$P(E|Cj) = P(e1|Cj) \times P(e2|Cj) \times \ldots \times P(er|Cj)$$

which can be estimated by:

$$\frac{n_1}{N_1} \times \frac{n_2}{N_2} \times \ldots \times \frac{n_r}{N_r}$$

where ni is the number of occurrences of word *ei* in all components of type *Cj,* and *Ni* is the frequency of *ei* in all abstracts, both taken from the corpus of 276. Our estimate for P(Cj) is the proportion of components in the corpus which are of type Cj. The values obtained from our data for this prior probability are given in Table 2. Given the mutual exclusivity assumption above, *P(E)* is the sum of the joint probabilities *P(Cj).P(E|Cj)s.*

To incorporate clues of a different kind, we can treat the posterior probability, *P(Cj|E),* as a new prior probability, and go through a similar process to produce a new posterior probability. For example, we know that components do not occur haphazardly in an abstract, but tend to occur in certain orders. This information can be incorporated to, hopefully, improve the probabilities. This order information is not clear-cut, but can be captured to some extent by observing the tendencies of components to occur in grossly defined regions of the abstract. We chose to divide the abstract into three approximately equal parts in terms of word count, and compiled a table of frequencies of occurrence of each component type in each third (see Table 2).

TABLE 2. *Components: prior probability distribution and order information from 276 abstracts*

j	Component C_j	Prior $prob(C_j)$	Order: Position in thirds 1st %	2nd %	3rd %
1	background	.01781	96	4	—
2	new terms defined	.00364	70	—	30
3	purpose	.09630	91	7	2
4	research topic	.08050	92	7	1
5	research question	.01660	82	5	3
6	hypothesis	.02023	45	28	27
7	independent variable	.03120	62	23	15
8	dependent variable	.03120	63	22	15
9	administrators	.00080	100	—	—
10	institution	.00080	100	—	—
11	time frame of study	.01538	69	29	2
12	location of study	.02670	90	10	—
13	sample selection	.01133	63	37	—
14	subjects	.10560	68	27	5
15	methodology	.02954	76	22	2
16	no. of experiments	.01821	60	31	9
17	conditions	.03076	51	44	5
18	control population	.00484	50	43	7
19	procedures	.06390	60	36	4
20	materials	.00243	40	50	10
21	data collection	.05260	45	50	5
22	test	.00040	100	—	—
23	data analysis	.02350	10	72	17
24	results	.10000	8	44	48
25	reliability	.00080	—	75	25
26	signif. of findings	.00486	—	17	83
27	discussion	.00973	6	4	90
28	limitations	.00121	—	40	60
29	unique features	.00162	—	62	38
30	conclusions	.05058	4	24	72
31	relat. to other research	.02430	46	23	31
32	practical application	.00566	7	10	83
33	implications	.01862	2	3	95
34	future research needs	.00405	—	—	100
35	tables included	.01457	—	—	100
36	appendices included	.01416	—	—	100
37	references	.06560	—	—	100

Part of an abstract:

ERPs to light flashes were recorded from 24 undergraduates during and immediately after auditorily presented ambiguous target sentences. Each target sentence was preceded by either a relevant or a neutral context sentence.

Division into text-pieces by human:

[...ERPs to light flashes were recorded from]

[24 undergraduates]

[during and immediately after auditorily presented ambiguous target sentences. Each target sentence was preceded by either a relevant or a neutral context sentence]

Division by punctuation:

[...ERPs to light flashes were recorded from 24 undergraduates during and immediately after auditorily presented ambiguous target sentences]

[Each target sentence was preceded by either a relevant or a neutral context sentence]

FIGURE 2. *Dividing an abstract into text-pieces*

Below, we present the effect of incorporating this particular information into the probabilistic text analysis.

Implementation of the probabilistic analysis

Some compromises have been made in the process of implementing this analysis on a computer. The first problem encountered is to determine how an abstract may be divided into text-pieces, satisfying the requirement that each is contained in exactly one component. We have not performed a thorough syntactic analysis of the discourse-level components in our corpus, but hold the opinion that it would be very difficult to separate text-pieces, under this constraint, by syntactic means. (Single words obviously satisfy the requirement, however the majority of text-pieces would in that case contain no clues.) In the work reported here, punctuation symbols (period, question and exclamation marks, comma, semi-colon, and colon) were used to separate text-pieces. An example of how this process differs from the decisions of a human expert is shown in Figure 2.

The clues used in this work were not precisely those nominated on semantic grounds by Liddy, although most of the latter are included. The new list was derived from statistical analysis of the collection of abstracts from both ERIC and PsycINFO. A summary of the main features of the corpus is as follows:

TABLE 3. *Probabilities from lexical clues: example*

Example text-piece:

'ERPs to light flashes were recorded from 24 undergraduates during and immediately after auditorily presented ambiguous and unambiguous target sentences'

Potential clues (for all components):

were, recorded, from, undergraduates, during, after, presented, target, sentences

I State C_j	*II* Prior prob(C_j)	*III* $P(E\|C_j)$	*IV* Joint $P(C_j).P(E\|C_j)$	*V* Posterior $P(C_j\|E)$	*VI* Standardised score
subjects	.1056	.900	.09504	.303	1.000
results	.1000	.900	.0900	.287	.947
procedures	.0639	.900	.0575	.183	.604
data collection	.0526	.726	.0382	.122	.403
conditions	.0308	.250	.0077	.025	.083
hypothesis	.0202	.333	.0067	.020	.066
independent variable	.0312	.200	.0062	.018	.059
research topic	.0805	.058	.0047	.014	.046
conclusions	.0506	.088	.0045	.013	.043
data analysis	.0235	.058	.0014	.004	.013
research question	.0166	.058	.00098	.003	.010
sample selection	.0113	.058	.0006	.002	.007
reliability	.0008	.058	.00005	.0001	.0003
.	.	0	0	0	0
.	.	0	0	0	0
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
references	.0656	0	0	0	0
			$P(E) = 0.0314$		

Note: The components are ranked by column VI.

Number of abstracts: 276
Number of components: 2,834
Number of words (tokens): 49,283
Number of distinct words (types): 6,478
Number of punctuation marks: 4,461

The properties that we desired of a good potential clue word are:

1. that it should occur in relatively few distinct types of component; and
2. that is should occur in relatively many abstracts.

TABLE 4. *Effect of order information: example*

I State C_j	VI Prior $prob(C_j)$	VII $P(E\|C_j)$	VII Posterior $P(C_j\|E)$	IX Standardised score
subject	.303	.68	.376	1.000
conditions	.025	.51	.232	.617
procedures	.183	.60	.199	.529
data collection	.122	.45	.100	.266
results	.287	.08	.042	.112
research topic	.014	.92	.024	.064
independent variable	.018	.62	.020	.053
hypothesis	.020	.45	.016	.042
research question	.003	.82	.004	.011
conclusions	.013	.04	.001	.0027
data analysis	.004	.10	.001	.0027
sample selection	.002	.10	0	0
reliability	.0001	.01	0	0
.	0	0	0	0
.	0	0	0	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
references	0	0	0	0

Notes: Column VI in this table is the same as column V in Table 3, ordered differently.
The event E is, in this case, the fact that the text-piece is in the first third of the abstract.
The components are ranked by column IX.

After some experimentation, a balance between these factors was found in the weight formula:

$$W = N \times N_a/N_c^{2.7}$$

where N is the total number of occurrences of the word in the corpus, Na is the number of abstracts containing the word, and Nc the number of types of components in which it occurs. The 6,478 distinct words in the corpus were ranked according to W (descending), and the top 1,192 became the list of potential clues with which we worked.

Tables 3 and 4 show how the probability calculations work out for a specific text-piece. The final column of each table contains the standardised scores attained by each component-type in relation to the given text-piece. These indicate the relative likelihoods that the text-piece belongs to the corresponding components. They are obtained by dividing the column of posterior probabilities by the largest value in that column, and therefore vary between 0 and 1. This is done to enable us to summarise the performance of the method over many text-pieces. Table 3 takes the calculation as far as the application of lexical clues, and in Table 4 information about the position of the text-piece in the abstract is incorporated. In this example, the CONDITIONS component has moved up noticeably in the ranking, which is appropriate for the text-piece in question.

Evaluation of the probabilistic method

A preliminary evaluation of the method has been done using a sample of ten abstracts drawn from the corpus. Each of these was divided into text-pieces automatically using punctuation, and also by a person familiar with the discourse-level structure. We should note here that, in addition to the linguistic compromise involved in the formation of the text-pieces, there are two deviations from the conditions of the probabilistic model: firstly, a textpiece may contain parts of more than one component (i.e. the mutual exclusivity assumption is not met) and secondly, the probability estimates used to recognise automatically separated text-pieces were derived from the frequencies of occurrence of clues in manually identified, full components.

In this evaluation, we have two treatment variables:

1. Text break-up method, either 'by human' or 'by punctuation';
2. Use of order information, 'order' or 'no order'.

Our measure of performance in relation to a single text-piece is the standardised score of the target component. This is, ideally, the component to which the text-piece had been assigned when the abstract's discourse-level structure was determined (by a human being). In cases where the text-piece overlapped with more than one component, the one with the highest standardised score was used. If the target component is first in the ranked list, its standardised score will be one, and we might say that the text piece has been correctly identified. The simple counts of correctly identified text-pieces in the sample are:

BY HUMAN: number of text-pieces correctly identified out of ninety-nine text-pieces.

No order: 24 (24.2%)
Order: 26 (26.3%)

BY PUNCTUATION: number of text-pieces correctly identified out of ninety-one text-pieces.

No order: 19 (20.9%)
Order: 37 (40.7%)

For our purposes, we think it is more important to know where the mean standardised score for target components lies in the distribution of standardized probabilities for all components with respect to all text-pieces (the 'overall' distribution). In other words, we want to know whether target components are getting relatively high standardised probabilities. The following chart describes the overall distribution:

	BY HUMAN	BY PUNCTUATION
NO ORDER	A ave = 0.0388 N = 1,406 S.D. = 0.1672	B ave = 0.0348 N = 1,591 S.D. = 0.1833
ORDER	C ave = 0.0482 N = 1,961 S.D. = 0.1809	D ave = 0.0400 N = 1,396 S.D. = 0.1959

Standardised probabilities for all components

The next chart describes the standardised scores for target components:

	BY HUMAN	BY PUNCTUATION
NO ORDER	A ave = 0.3268 N = 99 S.D. = 0.4111	B ave = 0.3317 N = 91 S.D. = 0.4178
ORDER	C ave = 0.3432 N = 99 S.D. = 0.4372	D ave = 0.4828 N = 91 S.D. = 0.4583

Standardised probabilities for target components

Our first observation from these figures is that under all treatments (A, B, C and D), the average score for target components is one order of magnitude greater than the overall average. In addition, it lies approximately two standard deviations away from the overall mean. We interpret this as a strong indication that the probabilistic algorithms are effective in identifying good *candidate* components.

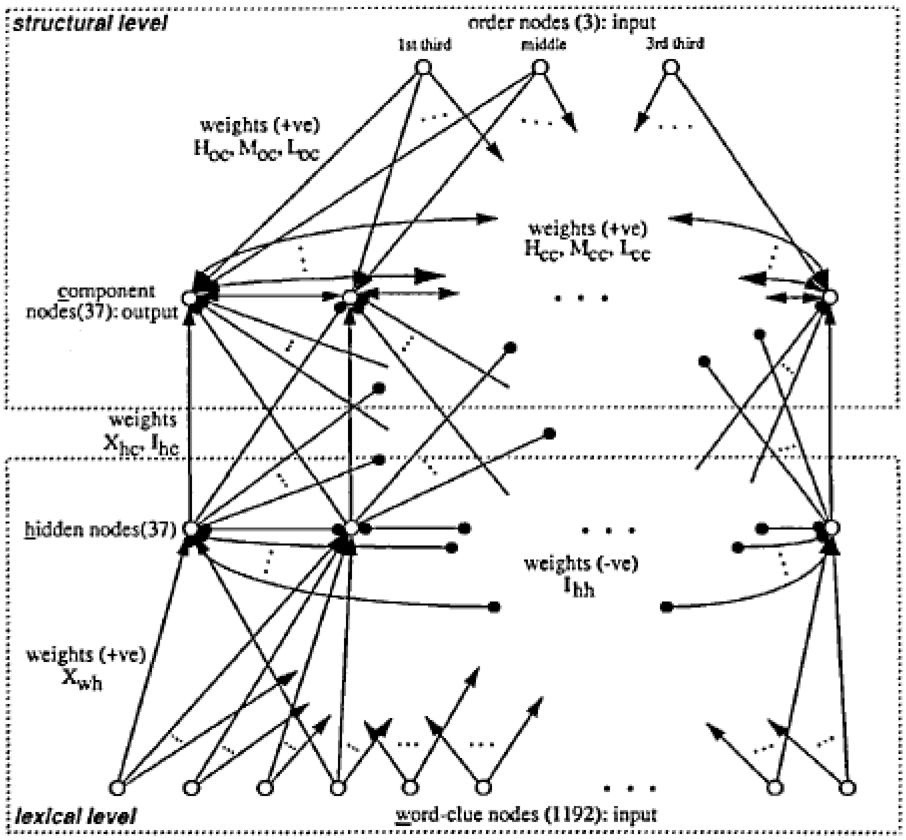
Another observation is that the average for target components in box D is higher than the others. We conducted a t-test on the difference between D and B for target components, and obtained $t = 2.324$, and *a* p-value < 0.021. We conclude that, in the case of automatically divided text-pieces, incorporating the order information into the Bayesian model has a significant beneficial effect. The t-test applied to boxes A and C (i.e. human text division) did not reveal a significant difference ($t = 0.272$, p-value < 0.786).

If we simply interpret these probabilities in the sense that the highest probability indicates the most likely correct component identification, then we have a large number of erroneous identifications. However, the statistically measured performance of a good probabilistically motivated model will be consistent with the measurements which were used to estimate the probabilistic parameters of the model. It can be nothing more! Hence, in our situation, we can expect no more than that a good probabilistic model will indicate which components are worth considering by some further processing. The next phase (which we have yet to tackle) would contain some merging of text-pieces and choosing final assignments to components, using rules concerned with juxtaposition and embodying the hierarchy in the discourse structure of empirical abstracts. For this purpose, it should be sufficient that the correct identification of a text-piece has a probability well above the mean.

2.3.2 Neural network-based text analysis

In a number of domains, it has been found that a suitably designed, richly connected network of simple processors (real or simulated) can be used to perform recognition, or diagnostic tasks. Certain nodes, referred to as input nodes are given interpretation as observable features of entities in the domain. Those corresponding with the features of a specific entity to be recognised are made active. Active nodes send signals along links to other nodes in the network. Each signal received by a node serves to increase its activation level, and if this level exceeds a given threshold, the node will fire, i.e. send signals along all the links emanating from it to other nodes in the network. This sequence of events is called an activation cycle and, repeated many times, the process is termed spreading activation. Some of the nodes in the network are interpreted as classes to be recognised. These are called output nodes. After many activation cycles, the activation levels of the output nodes are read off, and the high values determine the class assignment of the entity presented for recognition. In the application that we shall describe here, the machine should assign passages of text (the entities) to discourse-level components (the classes). Our input nodes correspond to clue words and textual order information, and the output nodes represent the thirty-seven component types.

There are many ways of constructing a neural network, and many parameters to be set for satisfactory performance. These have to do with the patterns of connection (the architecture) of the network, the strengths of the impulses from node to node, and the algorithm used to aggregate the incoming signals and modify the activation level of a node. In addition, most neural network models contain a learning mechanism whereby parameters (particularly connection strengths) can be automatically adjusted to improve the network's performance on a set of training instances. In this paper, we do not intend to discuss these very complex issues, and will limit ourselves to a brief account of our own application of this approach to computation. There is a large literature on the topic, e.g. Rumelhart and McClelland [20] and Khanna [21], and Balakrishnan [22] develops this aspect of our work. What we present here is the result of a 'trial-and-error' series of efforts to define a suitable architecture. We regard the work as at an early stage of development, and there are many intriguing ideas to pursue in the future.



Notes: 1. Each order node is connected to every component node.
Weights: $H_{oc} = 0.015$, $M_{oc} = 0.005$, $L_{oc} = 0.001$
2. Each component node is connected to every other component node.
Weights: $H_{cc} = 0.05$, $M_{cc} = 0.01$, $L_{cc} = 0.002$
3. Hidden nodes and component nodes are in one-to-one correspondence.
4. Each word-clue node is connected to some hidden nodes.
Weights: X_{wh} based on frequency of clue in components
5. Each hidden node is connected to every other hidden node.
Weights: $I_{hh} = -0.015$
6. Each hidden node is connected to every component node.
Weights: $X_{hc} = 0.4$ (corresponding nodes), $I_{hc} = -0.015$ (others)

FIGURE 3. *Neural network for discourse-level component recognition*

The network

The architecture of the neural network for discourse-level component recognition is shown in Figure 3. In that diagram, nodes are represented by open circles, and the lines stand for connections between nodes. At the ends of the lines are drawn arrow heads or dark circular blobs. These indicate the direction in which signals traverse the links (towards the arrow head or blob), and also whether the signal is positive (excitatory - arrow heads) or negative (inhibitory - blobs). Some lines convey signals in both directions, and not necessarily of the same strength. The strength of a signal is determined by a numerical weight associated with the link.

Although all the nodes in the network have the same basic properties (i.e. they are identical processors), they are grouped into four classes which determine their interpretation and patterns of interconnection. The four classes are shown in distinct rows in Figure 3, and are as follows:

1. *Word-clue nodes.* Lexical clues to the components are the same in this analysis as in the probabilistic one described above: there are 1,192 of them. One input node is assigned to each word in this list. When a word-clue is observed in a fragment of text, the corresponding input node is activated.
2. *Order nodes.* These three nodes are also input nodes in the network. Exactly one of these is activated for each text-piece to be recognised, depending on whether it occurs in the first, middle or last third of the abstract.
3. *Component nodes.* These are the output nodes in this network design. There are thirty-seven, corresponding to the component types in the Elaborated model (see Figure 1). After many activation cycles in the network, the activation levels of these nodes are examined to determine the outcome of the computation.
4. *Hidden nodes.* These nodes are neither input nor output nodes, and it is probably best not to try to give them a simple interpretation. They arose in the development of the model as a means of allowing the network to integrate confirming and disconfirming indications for each component. There is a hidden node corresponding to each component node in our network. A discussion of the role of hidden layers of nodes can be found in Smolensky [23, pp. 216-220] and Khanna [21, pp. 69-74].

The weights of the links between nodes are parameters of the model. We give here a brief description of the way these are set in the current (interim) version of the model:

1. Each order node is connected to every component node with an excitatory link. The weight on the link can be one of three values, depending upon how frequently the component was found to occur in the respective thirds of the 276 abstracts in the corpus (see Table 2). For instance, the link between the 1ST order node and the METHODOLOGY node is *Hoc* (high), between 2^{ND} and METHODOLOGY is *Moc* (medium) and between 3RD and METHODOLOGY is *Loc* (low).
2. Each component node is connected to every other component node with an excitatory link. These links reflect the *structural* aspects of the break-down into discourse-level components, i.e. the relationships between components. The presence of one discourse-level component lends evidence that certain other components exist in the vicinity. For example, if there is some indication that the DEPENDENT VARIABLES component exists, it suggests the existence of the INDEPENDENT VARIABLES component. In this case, the relationship follows from the hierarchical structure (Figure 1). In this model, we have assigned one of three positive weights to each component-component link:

a. a relatively high weight *(Hcc)* from lower level components to higher level ones;
b. a medium weight (Mcc) from higher level components to lower level ones within the same branch;
c. a low weight *(Lcc)* to all other links.

3. Each word-clue node is connected to the hidden nodes corresponding to those components for which the word is considered a lexical clue. The weights on these links are based on the frequency of occurrence of the words in the components in the corpus, and the formula is analogous to the '*tf* x *idf*' weight used in information retrieval [24]. The weight, *Xwh,* between word-clue node *w* and hidden node (corresponding to component) *h* is:
Xnh = *Fnh* x log[(*N* -*fw*)/*f*w], if *w* is a clue for *h* (there is no link otherwise)
where: *Fuh* = frequency of occurrence of *w* in all components of type *h;*
fw = number of components containing *w;*
N = total number of components in the corpus.

These weights were scaled by an experimentally determined factor, so that they become of a magnitude comparable to other weights in the network.

4. Each hidden node is connected to every other hidden node with a mildly inhibitory link.
5. Each hidden node is connected with a strongly excitatory link to its corresponding component node, and an inhibitory link to every other component node. In this way, activation spreads to components for which clues exist in the text fragment, but activation is dampened in components for which no clues are present (in competition with excitation at the level of discourse structure).

Activation of the network
Firstly, we describe what happens in the network in a single activation cycle.
All nodes behave in the same manner, as follows:

(a) A collector variable in each node accumulates the sum of the weights of those incoming links which emanate from active nodes (i.e. nodes whose activation level is above a positive threshold). This simulates the notion of active neurons firing and contributing to the excitation of other neurons according to the strength of their connections.
(b) The new activation level *Anew* is calculated in each node from the old one *Aold,* using the formula:

if *collector* > collector threshold
Anew = *Aold* + *collector* x *(Amax - Aold)*
else
Anew = *Aold* + *collector* x *(Aold - Amin)*

where *Amax* is the maximum activation level (1.0) and *Amin* is the minimum (— 0.2). This process closely follows the model described by McClelland and Rumelhart [25].

Next we discuss the ways in which the network can be triggered. Given a text fragment, the clues that it contains can be treated either separately (in sequence) or simultaneously. In what we call the clue-by-clue (CBC) approach, the first clue in the fragment is identified and the corresponding input node in the network activated. The network is allowed to run for a certain number of activation cycles (twenty to thirty in our trials). Then the activation levels of hidden nodes and word-clue nodes are reset to zero in preparation for the next clue. Activation levels of component nodes are retained. The whole process is repeated for each clue in the fragment. We call the other method of triggering the network the fragment-by-fragment (FBF) approach. Here, all the clues in a text fragment are identified before any triggering occurs. The corresponding word-clue nodes are then activated simultaneously, and the network allowed to run for a set number of cycles (thirty to forty). Clearly, this is computationally less costly than the CBC approach.

Evaluation of the neural network method
The neural network activation programs were implemented on the massively parallel Connection Machine [26], using the programming language C*. This is, in itself, an interesting exercise in the application of a promising new technology to information retrieval problems. However, discussion of the technical details is beyond the scope of this paper.

We are able, at this stage, to report only very preliminary results of these methods of text analysis. Early versions of the program made use of the CBC approach, and many of the parameters of the network were set at this time. The test that we report here, however, was done with the FBF (fragment-byfragment) approach. A sample of twenty abstracts (ten from ERIC and ten from PsycINFO) were used in this test, and the text fragments were whole sentences (167 of them). Two hundred and sixty-one components were identified manually in this sample. Two treatments are compared: 'without order information', in which the order input nodes in the network were not activated, and 'with order information', in which one was activated, depending upon where in the abstract the text fragment began. Two measures of success in component identification are reported: (i) the number of instances where the component node with highest activation level is one of the components identified by hand as overlapping the text fragment, and (ii) the number of instances where one of these manually identified component nodes had a high activation level (above 0.85). The results can be summarised as follows:

		Identification criterion	
		maximum activation	*activation > 0.85*
Order information	YES	69 (41.3%)	114 (68.3%)
	NO	67 (40.1%)	109 (65.3%)

This performance is clearly not yet adequate for an operational system, although it does indicate that a neural network approach to the problem is feasible. It should be borne in mind that recognising discourse components is a relatively high-level task for this technology at the present time. A failure analysis has suggested a number of improvements. For instance, the RESULTS component node often erroneously achieves a high activation level. The reason for this is that RESULTS components are usually by far the longest in empirical abstracts, and this leads to comparatively high weights when our frequency-based formula for *Xuh* is used. If the average lengths of the various component types were appropriately incorporated into the weight, we would expect significant improvement. The incorporation of order information, through the three order nodes, has so far given us no significant benefit overall, although it has proved useful in some cases where the evidence from clue words is weak.

3. USERS' EXPRESSIONS

People engaged in empirical work can be viewed as members of a community: they complete similar tasks, espouse shared social and technical norms, and operate within a unique reward structure [27, chapter 7]. We first attempted to find evidence of this common information use environment [28] in the discourse-level structure of users' problem statements, which, we anticipated, would be similar to that found in abstracts. It emerged that although discourse analysis had provided very appropriate techniques for analysing empirical abstracts, understanding users' information needs required an approach more akin to that used in the area of linguistics known as pragmatics. In the study of pragmatics, the situational aspects of a communication are paramount in understanding the meaning of a text. What is required is a means of capturing from users information about their work situations that cause them to seek information about other research. This type of information is rarely embodied very fully in users' problem statements. What we found during extensive interviews with end-users and intermediaries, and an analysis of problem statements, was not so much an inability on the part of the users to present their information needs, but a tendency for users, intermediaries and current systems to concentrate mainly on the topic aspects of an information need. Perhaps due to expectations about the capabilities of current systems, the situational aspects that gave rise to the users' needs and which were major determinants of their relevance judgements seldom come into consideration when conducting searches.

3.1 Discourse-level structure of problem statements

Our initial aim in studying users' problem statements was to describe the essential components of a prototypical problem frame structure (as in Liddy's study of empirical abstracts). The search for a source of expertise in this area began with a comprehensive review of the literature describing how information seekers and search intermediaries (both machine and human) recognise, specify, and interpret information needs. The problem statement texts chosen for examination were transcribed from audiotapes made during an earlier project [29]. We selected and studied twenty of the statements offered by empirical researchers. Subjects in the earlier study had been asked to give a statement of their information need and the context or problem from which it arose. The resulting problem statements typically consisted of about 300 words.

The first coding scheme that we developed to describe the structure of these statements was based on discourse analyses of dialogues between information seekers and search intermediaries conducted at the City University, London [15]. This set of codes was adapted to describe those elements which the subjects spontaneously mentioned in their statements (as opposed to those that arose in dialogues). There were about 150 codes in our scheme. Some component definitions turned out to be somewhat ambiguous when coders attempted to assign them to pieces of text in the problem statements. Generally, it was easier to understand and identify specific elements (e.g. 'news', defined as 'the user would like newspaper articles') than broader codes (e.g. 'problem mode', defined as 'the appropriate mode of response'). There was a tendency to apply more than one code to a given piece of text.

It seemed that a more reliable and distinct partitioning of a problem statement text might be achieved if the taxonomy were collapsed into a few broad categories, namely: background, information traits, position of user in problem treatment process, non-topic aspects of user's research, subject literature, topic, and user model. Definitions for these seven components were provided to nine coders, each of whom coded the same group of ten problem statements. Intercoder consistency was fair, but discussion revealed that the definitions were still ambiguous, leaving much to the interpretation of the individual coder. Text elements appeared either to fit easily into several of the broad categories, or not to fit into any of them. In addition, such lexical clues as suggested themselves were more ambiguous and inconsistent than those in empirical abstracts. Thus, we did not succeed in delineating a discourse-level structure for problem statements which would lend itself to automatic detection. We shall devote no further space in this paper to detailed reporting of this effort.

From the work on delineating the discourse-level structure of users' problem statements several conclusions emerged. One was that problem statements do not seem to exhibit the same sort of predictable structure that can be discerned in empirical abstracts. We attributed this to the fact that while empirical abstracts are texts which summarise documents thatthemselves have a standard structure, problem statements are texts which attempt to describe a user's unique problem. Another property of our collection of problem statements is their origin in spoken statements. They exhibit the typical characteristics of more or less spontaneous spoken discourse: informality, disjointedness, redundancy, immediacy, and the speaker's expectations of the listener's reaction. Thus, we came to regard problem statements as a type of spoken discourse, rather than as a text-type. Though incomplete as an expression of the user's requirements, they can be quite rich in what they reveal, both implicitly and explicitly, about the nature of the user's problematic situation. However, the problem statement cannot serve as the sole basis of the entire retrieval process, but should be viewed, rather, as one possible 'opening move' in a conversation between the user and system.

With the wisdom of hindsight, we are not surprised that problem statements do not exhibit the same degree of predictability either in structure or vocabulary that was found in the empirical abstracts. Both discourse theory and sublanguage theory [30] suggest that a community of users which produces texts of a particular type will, over time, develop a shared notion of what their texts should be. Frequently this is not discussed, but rather becomes the implicit norm for that community. Since the disparate subjects who provided the problem statements do not, as far as we are aware, comprise a community of text producers, and do not regularly produce problem statements, they have developed neither a sublanguage for their statements nor norms for problem statement content and structure.

In addition, it now seems likely that we were confounding two distinct discourse types: one to do with researchers' work problems, or information use environments, and the other to do with their information search problems. Abstracts are part of the work-related discourse, and problem statements are from the search-related discourse. Hence, in the context of our investigation of the use of situational information, problem statements appear not be a very appropriate focus of attention.

3.2 Users' problem situations

In order to establish a context for the discourse on work situations, we conceptualised both the author's and the user's situation as occupying a particular stage in the research process [27, chapter 6; 31, 32]. The research process can be viewed as a script. A script provides an outline of the steps needed to complete a particular type of task. Mavor and co-workers [33] developed detailed scripts for doing research in three particular scientific subdisciplines. Moreover, they made assertions, based on discussions with researchers in those disciplines, about the type of information usually helpful to complete each task in the research script. We decided to investigate the use *of generic* scripts. That is, one script would outline the stage of work pursued by the typical empirical researcher. A generic script for empirical research can be derived from a variety of sources. The empirical abstract components delineated by Liddy [17,18] (e.g. sample selection, data collection, discussion) provided a guide to the steps which make up the research process. These needed to be interpreted more as processes than products to reflect the tasks which a researcher must still complete rather than those which an author is reporting as completed. Abstract components were supplemented by some of the components identified as part of our investigation into the discourse-level structure of problem statements. These directly reflect users' interpretations of the tasks they must complete. Finally, the entire list was converted into a generic empirical research script by presenting it from the perspective of 'the scientific method' as it is described in many research design and statistics textbooks. Table 5 contains the resulting script. This script describes an idealised, linear process. Researchers do not always progress through the stages represented in the script in a strictly linear fashion; nor would a system, of the kind we envision, force users to proceed in any prescribed manner.

TABLE 5. *Generic empirical research script*

1. Choose research topic.
2. Complete background research to become familar with topic.
3. Complete background research to become familiar with related theory.
4. State purpose of research.
5. Develop research question/hypothesis.
6. Identify variables.
7. Complete comprehensive literature search to see if results of identical research effort already published.
8. Operationalise variables.
9. Decide on time frame for study.
10. Decide on location of study.
11. Decide on type of subjects desired for study.
12. Choose methodology/research design.
13. Decide on number of experiments.
14. Decide on conditions/treatments to be applied.
15. Delineate detailed outline of experiment procedures.
16. Acquire needed materials/equipment/technical expertise.
17. Select sample.
18. Collect data.
19. Perform tests.
20. Complete statistical analysis.
21. Report findings/results.
22. Evaluate findings (reliability, significance).
23. Interpret findings (limitations of study, unique features, etc.).
24. Formulate conclusions (generalisability of results, relation to other research).
25. Consider practical applications.
26. Weigh impact or consequence of findings on the field.
27. Formulate recommendations for future study.
28. Write research report, dissertation, etc.

Our work with problem statements led us to believe that, especially in the case of novice researchers, the user lacks knowledge about the composition of the shared research script and/or is unaware of his or her position in the script. This is one aspect of the typical researcher's information need that has received little attention up to this point. As a first step toward exploring the notion of research scripts and their utility for IR systems, we interviewed several researchers and search intermediaries.

3.2.1 End-user interviews

Before proceeding with a design for an IR system, we felt the need to seek some evidence of the validity of the notion that a research script would be useful ininformation retrieval. To this end, in-depth interviews were conducted with four researchers (the IR system users) and five search intermediaries. We shall begin with an account of the user interviews.

Our objectives in interviewing users were:
1. to learn more about how a problem statement can be interpreted, by exploring the work context of the user's problematic situation;
2. to ascertain whether researchers recognise the existence of a research script and their position in such a script, by discussing the sort of tasks comprising their research;
3. to find out what criteria they had used to make relevance judgements about the documents retrieved by the online search.

We hoped thus to discover connections between users' problem statements, relevance judgements, and research scripts. We wanted to explore the notion that scripts could be used in the prediction of relevance judgements.

User interview methodology

Four subjects were selected from those who had given the problem statements in the earlier project. (More than a year had elapsed since their earlier participation.) Subjects were selected who 1. had been engaged in empirical research; 2. represented a variety of academic disciplines; and 3. had been at different stages of the research process when they approached the online search service. Literature on conducting research interviews was reviewed [34, 35], and our interviewing instrument made use of techniques employed by Nilan and Fletcher [36] in their interviews with information seekers. These techniques elicit descriptions of researchers' problematic situations and information gathering behaviour as a series of steps. Kwasnik [37] developed inductive methods for eliciting descriptions of cognitive processes involving information-related interactions, taking into account the interpretative processes that occur when people are asked for verbal reports. Our interviews with researchers were open-ended, lasted about 1½ hours each, and were audiotaped and later transcribed.

Overview of results

While the subjects had not explicitly discussed the nature, methodology, and so on, of their empirical research in their original problem statements, i.e. they had not talked about a research script and their placement in it at the time of the original online search, this type of information was remarkably easy for them to provide. Also, their view of where they were in the research script consistently revealed itself as a strong predictor of relevant documents. What these data suggest is that if users had been asked 'Where are you in the research process?' and 'How will the information help you to accomplish your next research task?' they could have provided important insights into the intended use of the information they hoped to retrieve, which in turn could be used to improve the focus of the online search. Particular aspects of the results will be amplified below.

Awareness of research script

As one part of the interview, each subject was asked to describe his or her research as a series of steps that most people who were doing that kind of work would probably go through. The users described, quite glibly, fairly specific scripts, differentiated somewhat according to discipline and type of research design. The research agendas of the four subjects were as follows;

S1: qualitative research, based on interviews, on issues related to volunteerism;
S2: multiple regression on an existing data set for a social science dissertation on the criminal sentencing of women;
S3: applied policy research on adopting the 'Effective Schools Model';
S4: the production of an ethogram for reporting the habits and movements of the two-toed sloth.

As can be seen in the full scripts (Tables 6-9), the ability of users to elaborate research scripts suggests that they do exist for them at some level of awareness, at least in retrospect, and that perhaps a more conscious awareness of them at time of search would have been helpful.

> S4: (In response to question: 'Do you think that you were sort of aware of this whole process [the particular research script] at the very beginning?') 'The tendency is to say no because, I mean, I guess I didn't realise. I knew there would be many steps, many stages, but I didn't realise how much lack of definition I had in the beginning... it's gotten a little more complicated, a little more in-depth than I could have possibly imagined the first time I really did anything this thoroughly in terms of scientific research ...' (When asked whether it would have been helpful to have known about the steps in the beginning)'... it gives you a sense that you're on the right track . . . you can anticipate better where you're going...'.

TABLE 6. *Script from user interview (S#1)**

Qualitative research: perceptions of and about volunteers

1. Get funding for next two years.
   - Send for specific applications.
   - Talk to specific people.
   - Search online funding database.
   - Contact appropriate offices, organisations.
2. Data collection.
   - Ask contact person for address list.
   - Set up meetings, contact various people (to discuss instrument?)
   - Organise interview tapes.
3. Dissertation Committee administrative tasks.
4. Data analysis.
   - Write one page statement of topic.
   - Code interviews.
   - Create logic variables.
   - Transcribe related interviews.
   - Write memo on structuralist methods.
5. Dissertation literature review.
   - Order ERIC report.
   - Find copy of [particular item]
   - Do full-scale online search.
   - Ask about list of descriptors.
   - Order dissertations.
   - Enter current references into ArtFile database.
   - Go through materials in *Social Work Journal*.
   - Read [particular items]
   - Develop outline for review.
6. Miscellaneous stuff.
   - Site visits.
   - Keep up particular personal contacts.
   - Write abstract.
   - Contact particular people.
   - Get reimbursed for conference.

*Note: This script was not given orally by subject during interview. It is derived from the written plan for dissertation work that the user referred to in the interview. The dissertation is on a different topic from, but uses the same research methodology as, the online search topic.

TABLE 7. *Script from user interview (S#2)*

Multiple regression on an existing data set: criminal sentencing and women [dissertation]

1. Read something or talk to people to get an idea.
2. General reading (classics in the field, books, etc.) to get background information.
3. Read review articles.
   — Use bibliographies to find new sources.
4. Choose a particular perspective.
5. Read more specific articles.
6. Find out who's done what you plan to do;
   — Decide what modifications you can, should make.
7. Have hypothesis.
8. Do some observing (i.e. watch court proceedings).
9. Discuss topic with expert, others with experience.
10. Develop support group with other students, etc.
11. Alert librarians to topic.
12. Find data set (in library basement);
    — Identify, evaluate, options; select data.
13. Choose variables.
14. Get data up on computer;
    — Clean up data.
15. Find someone to teach you computer techniques.
16. Learn new computer techniques (SAS).
17. Add variables to chosen data set.
18. Re-code variables.
19. Understand sample and its relationship to population.
20. 'Run it' . . . Play with data.
    — Look for relationships;
    — See which variables seem to influence others.
21. Data might give new ideas, insight to test or play around with.
22. Data might suggest other projects.
23. Check for oversights, errors.
24. Conclusions:
    — Highlights of research;
    — Future directions;
    — State of knowledge on topic.
25. Re-check literature.
    — Keep current (esp. if following particular case).
26. Hand in dissertation.

The specific scripts elicited from Subjects 1 to 4 were analysed in order to assess their congruence with the generic research script described above. Table

TABLE 8. *Script from user interview (S#3)*

Applied policy research: adopting the 'Effective Schools Model' (S#3)
1. Get contract.
2. Contact principal and key people on committee.
3. Conduct needs assessment.
4. Discuss key issues, problems with principal.
5. Determine relationship of problems to effective schools model.
6. Design survey.
7. Administer survey.
8. Analyse data (Likert scale).
9. Semi-structured interviews to explore particular problems identified through survey.
10. Analyse interview data.
11. Write up findings and recommendations.
12. Follow-up visit to supervise implementation of recommendations.

10 provides a view across the four users of the particular steps in the generic script which were included in each person's description of her or his own research process. There are some obvious differences due to the subjects' specific research areas, but also some clear patterns – some steps that were common to all four research scripts.

In addition, the interviewees were aware in a very general sense that different information was required at each stage.

> S1: '. . . I've also given indications, I think, in talking about this that, that there are perhaps steps or phases that are kind of points along this path as you describe it, that have different needs. For example, I talk about the point I was at with this research and that somehow having to do with what my need was. I was at an early stage, my information need was somehow different from what it would have been perhaps at a much later stage and . . . there are contextual clues here . . . just in the whole way I've talked about it, that, sort of, even despite my resistance to thinking about the specific or discrete set of steps that are involved . . . I think that there are different kinds of information that are useful at different points for different, sort of, parts of the research process. . .'.

However, when the notion of a generic script was suggested to the subjects, there was some concern expressed over whether a user's individual nature, flexibility, and control would be sacrificed if research scripts were built into information retrieval systems.

Effect of position in script on relevance judgements

The interviews suggest to us that the research scripts were related in a number of fruitful ways to what was expressed in the problem statements and to the

TABLE 9. *Script from user interview (S#4)*

Ethogram: reporting the habits and movements of the two-toed sloth (S#4)

1. Formulate research questions;
   — Revise, add throughout ⟶
2. Initial observations
   — Get to know about sloth's basic behaviour, in general;
   — Learn to identify individual subjects.
3. Prepare description of each sloth.
4. Conduct information search
   — in order to understand what you're observing ⟶
5. Identify which behaviours to include in ethogram;
   — Label behaviours;
   — Subdivide into specific facets.
6. Choose type of ethogram to use.
7. Carry on discussions with other experts ⟶
8. Develop conclusions
   — Revise as you go along; change, rewrite ⟶
9. Develop coding scheme.
10. Prepare data sheets for data collection (incl. floor plan, date/time, comments area)
    — Adapt existing form.
11. Conduct time scan (Gather observations every 15 mins. for 4 hrs., for 6 mos., in 2 settings)
12. Group, reorganise data.
13. Prepare computer coding scheme.
14. Enter observations into computer.
15. Refine, elaborate, resolve inconsistencies in coding scheme.
16. Manipulate data.
17. Analysis;
    — Compare individual sloths;
    — Look for trends, relationships.
18. Create time budget (i.e. main portion of ethogram)
19. Go back to zoo for follow-up observation.
20. Come up with way to describe behaviours for qualitative discussion.
21. Finalise, write up conclusions;
    — Summary.

[22. Go into wild, in case of extra behaviours]
    — Use knowledge of wild to validate your findings.

relevance judgements made by the users at the time of the search. Thus, it seems that they may be useful in formulating more effective search strategies for document retrieval. Although different users described different criteria for their relevance judgements, the relevance criteria suggested by each user

TABLE 10. *Comparison of generic and specific scripts*

A specific script received an 'X' in the row for a particular generic research step if that step – or a similar, but more specific task – was mentioned by the subject.

	Generic script research steps	*Specific scripts**			
		1	*2*	*3*	*4*
1.	Choose research topic		X		
2.	Complete background research (literature review)	X	X		X
3.	Complete background research (discussions with other researchers and experts, familiarisation with research setting)	X	X	X	X
4.	State purpose of research				
5.	Develop research question/hypothesis		X	X	X
6.	Identify variables		X		
7.	Complete lit. search to see if results of identical research already published		X		
8.	Operationalise variables				
9.	Decide on time frame for study				
10.	Decide on location of study				
11.	Decide on type of subjects desired for study				
12.	Choose methodology/research design		X		X
13.	Decide on number of experiments				
14.	Decide on conditions/treatments to be applied				
15.	Delineate detailed outline of experimental procedures	X	X	X	X
16.	Acquire needed materials/equipment/technical expertise		X		
17.	Select sample				
18.	Collect data	X		X	X
19.	Perform tests		X		
20.	Complete data analysis	X	X	X	X
21.	Report findings/results			X	X
22.	Evaluate findings (reliability, significance)		X		X
23.	Interpret findings (limitations of study, unique features, etc.)				
24.	Formulate conclusions (generalisability of results, relation to other research)		X		X
25.	Consider practical applications				
26.	Weigh impact or consequence of findings				
27.	Formulate recommendations for future study		X	X	
28.	Write research report, dissertation, etc.			X	X
29.	Follow-up (further discussion or research, implementation, etc.)			X	

* The specific script presented by subject #1 was actually a list of 'things to do' which the subject had compiled previously.

TABLE 11. *Users' relevance criteria*

1. 'Aboutness' or 'topic-related' criteria
   a. Right kind of volunteerism (S1)
   b. Right setting (S1)
   c. Examples of other settings (S1)
   d. Inclusion of all three factors: women, race, sentencing (S2)
   e. On topic (S3)
   f. On topic: sloth behaviour (S4)
   g. Life history features: habitat (S4)
   h. Narrower topic: adult sloths (S4)
2. 'Traditional non-topic' criteria
   a. Particular author (S2)
   b. Publication source (S3)
   c. Place of publication (S4)
   d. Language (S4)
   e. Either retrospective or current (S4)
3. 'Problem-oriented' or 'situational' criteria
   a. Help to define what I was doing (S1)
   b. Research methodology used (S1)
   c. Right issues addressed (S1)
   d. Filling 'gap' in current info; only new info needed (S2)
   e. Research method (S2)
   f. Something that 'triggers a thought' (S2)
   g. Where I was in my own work (S2)
   h. Unusual variables (S2)
   i. Congruence with own experience, point of view . . . confirmatory (S3)
   j. Breaking new ground (S3)
   k. Discussion of results implementation (S3)
   l. Appropriateness of focus (*not* how-to or anecdotal) (S3)
   m. Empirical research method (S3)
   n. Not too technical (S4)
   o. Anatomy or physiology aspect (S4)

included both 'aboutness' and 'aspect' or 'problem-oriented' elements (see Table 11). Problem-oriented elements are those which relate to an immediate problem which a researcher has encountered in attempting to accomplish a particular task in the research script. In giving their original problem statements, users devoted considerably more attention to aboutness criteria, perhaps because, as suggested by Weinberg [38] and others, these correspond with the indexing mechanisms found in most current systems. Moreover, they think of current systems as document retrieval, rather than 'problem solution' systems. Their statements are typically expressed in terms of a perceived information need, e.g. 'I need information about...' rather than as a perceived work problem, e.g. 'I'm having trouble with . . .'.

The interviews, however, revealed that individual differences in users' relevance judgements were tied to the placement of the users in the research scripts they described. For example, the individual doing qualitative research on volunteers noted in his problem statement that he was 'just getting involved in' the project. During the interview he said that the most important thing in his problem statement was the 'background information' on all the various facets of volunteerism that would be of interest to him, because he had not yet become familiar with theory, formulated a hypothesis, designed the research, or collected data. His relevance judgements seem related to his position at a formative stage in the research script. He subsequently noted that some documents were relevant because they dealt with a variety of issues, some because they gave him ideas for a methodology. On the other hand, the individual doing research on the criminal sentencing of women stated that the keywords: 'women', 'race', and 'sentencing' were the most important aspect of her problem statement. The first two keywords actually referred to her study's independent variables and the third was her dependent variable. And, in subsequent discussion, she revealed important information that was *not* in her problem statement: that she was well along in her research, was doing regression analysis using a statistical software package, and was really only interested in re-checking the literature to see if anyone else had done a regression analysis with her chosen variables and, if so, what they had come up with. This specific purpose was clearly in evidence in the relevance judgements she had made at the time.

Figure 4 is an example of the charts we produced for each subject. They reflect our preliminary attempts to analyse more systematically the connections between problem statements, research scripts, users' place in the research script, intended use of retrieved information, and relevance criteria. In each of them, it is instructive to note the type of information that can be gleaned from problem statements directly once one is aware that non-topic aspects are of significance. It is equally informative to note that the users could have embellished their original problem statements with much useful data, had they been aware that it would have helped to produce more relevant results, even, perhaps, with the traditional 1R system they used.

Another finding was that several subjects included in their research scripts steps related to the use of various informal information networks. They noted that they formed support groups with other students who had gone through the same kind of work, or sought out other researchers who had done work pertinent to theirs 'just to discuss common interests, problems, etc.' The preference for these informal networks for communicating with others about their common tasks and purposes would seem to validate our notion that the most useful retrieval would exploit this search for *process* commonalities, by matching the discourse-level components of empirical abstracts with the information seeker's placement in the research script. It was quite apparent that current IR systems and intermediaries are not considered by users as able to provide the type of information available through the informal network. Users set up a dichotomy in their minds about what type of information can be

RESEARCH SCRIPT/TOPIC: Multiple regression on an existing data set for a social science dissertation on the criminal sentencing of women.

Clues from PS: 'I'm from Sociology and I study Criminology. Basically I try to determine which factors affect the sentencing of women, criminal sentencing . . . I would particularly like information that deals with race, women, and sentencing.' [No clue to precise research method]

Interview comments: 'It was my dissertation . . . I was using a canned data set ... from the Bureau of Justice... and I needed to know which variables might be significant . . . I intended to do a regression'.

PLACE IN SCRIPT: Subject had already been working on the research for about a year. Most of literature review was done, subject had done background reading, had tentative hypothesis, had selected methodology, and located data set. Now needed to select variables [Step 13 in S #2 Scripts]

Clues from PS: 'Basically I'm trying to determine which factors affect the sentencing of women . . . and [compare] the different variables that come into play and how it affects each group of women . . . '

Interview comments: 'most of my literature review was done . . . the literature search would allow me to choose which independent variables to use in my model'

CHIEF CONCERN (identified by user in interview): 'I was having trouble finding information . . . I had used what resources I knew were available, and so I thought by using the computer I'd be able to search different databases that

I wasn't able to do by myself or do it quicker... I had basically given up on finding any more articles or an article which more closely paralleled what I was doing . . . It had to deal with women and race and sentencing. Two out of three really wasn't good enough; I was able to find two out of three when I did it myself.

Clues from PS: 'What I would particularly like to deal with is how the issue of being black or white affects the sentencing of women . . . there is very little information about just studies of women who are black or white and compares the different variables that come into play . . . so I would particularly like information that deals with race, women, and sentencing'

INTENDED USE (identified by user in interview): 'I had the data set of so many women; I wanted to know which variables to choose... I intended to do a regression'

Clues from PS: None specifically about choosing variables.

FIGURE 4. *Relationships among scripts, problem statements (PS) , and relevance judgements from user interview S#2*

Figure 4 (*continued*)

RELEVANCE CRITERIA (identified by user in interview):
- On all three facets of topic: women, race, sentencing;
- Followed work of one particular author because her work paralleled own work:
- Filling 'gap in my information,' i.e. needed only new, specific information;
- Research method (same, regression, would be good; historical overviews not helpful);
- '. . . make sure they're not doing what you're doing with same data set';
- Something that 'triggers a thought';
- Unusual variables;
- 'Probably the most important thing was where I was'.

Clues from PS (identified by user in the interview): 'coming from Sociology suggests that I'm interested in certain things . . . "particularly would like information that deals with race, women, and sentencing" . . . [research question]: are women of either race getting longer sentences'.

---

provided by formal systems and what type must be sought within the informal network. An ideal information system would integrate these types. (We should note that an environmental change that is beginning to blur these roles is the widespread accessibility of electronic networks for mail, conferencing and scientific communication [39, 27, chapter 6]).

3.2.2 Intermediary interviews

In order to gain a different perspective, we interviewed five intermediaries about the nature of the discussions which typically occurred when they conducted searches for researchers engaged in empirical work. We wanted to know what problem statement components searchers think are most important in the formulation of a search and whether they use the notion of research scripts in any way to formulate or modify search strategies. This information would allow us to compare the expectations which searchers and users hold, especially in regard to current IR systems, and might also shed light on the design of an effective user interface in the future.

Intermediary interview methodology

The five intermediaries interviewed worked with users in a variety of disciplines (Agriculture, Social Science, Physical Sciences, and Applied Sciences). They had no connection with the subjects of the user interviews described above. All five had some experience in both manual searching, as a part of library reference work, and online searching. Three of the five subjects noted that the end-users were generally *not* present during the actual online searches they conducted. Our interviewing approach was based on a series of open-ended questions. The first group of questions dealt with the intermediaries' perceptions of users' problem statements. The intermediaries were asked

1. what they thought were the most important components of users' problem statements;
2. which components were typically mentioned spontaneously by users: and
3. how they used these components, either to understand users' problematic situations or to formulate or refine a specific search strategy.

The second group of questions dealt with the idea of research scripts. Intermediaries were asked:

1. whether they thought the same problem statement components were important when conducting a search for those people engaged in empirical research;
2. whether researchers discussed different aspects of their problematic situations more than did other types of information seekers;
3. whether they could describe the steps most researchers go through in the course of their work;
4. whether they thought that researchers were aware of these scripts;
5. if they thought particular kinds of information might be helpful at particular steps; and
6. whether the intended use of the information sought or the user's place in the research process (two critical contextual or situational indicators) were typically used by them in formulating searches or were brought up by users in their relevance judgements. Each interview lasted about 1½ hours, was audiotaped and transcribed.

Overview of results

For the purposes of this discussion, problem statement components may be grouped into three categories. The first category comprises topic-related keywords. This includes those words that information seekers use in describing the specific topics or broad subject area of their search. Intermediaries usually take these terms from the user's natural language problem statement and use them as the keywords in Boolean queries to the IR system. Other problem statement components may be grouped together as non-topic information. These aspects of the information use context encompass demographic data pertaining to the user (educational level, previous search experience, etc.) and a description of desirable document characteristics (language or date limits, level of difficulty, etc.). Traditional IR systems and human search strategies often make use of this type of information from and about the user. Finally, it is useful to distinguish a third category, the situational aspects of a user's information need. Information seekers may include in their problem statements comments about how or why the information need arose, or about the intended use of the retrieved documents.

The results obtained from our interviews with intermediaries allow us to offer tentative answers to questions about interactions between researchers engaged in empirical work, search intermediaries, and current information retrieval systems:

1. Do intermediaries solicit input about the end-user's situation or purpose/ reason for seeking information? Our results suggest that intermediaries do ask about the intended use of the information being sought, but only in very general and limited ways. They tend to focus simply on the end goal, for instance, whether the researcher is writing a journal article as opposed to a dissertation. Questions about the user's situation do not become a focus of the search.
2. Are there ways in which these data from the users might be used to modify or improve a search? Our interviews with researchers revealed that their place in the research script was linked to their relevance criteria; the general opinion of the intermediaries, however, was that this kind of data was difficult to elicit from users and was not easily applied to the enhancement of searches using current online IR systems.
3. Are intermediaries aware of a common research script followed by empirical researchers? The intermediaries we interviewed described scripts, but they were much more general than those described by users. There was a fair amount of variation among intermediaries in their opinions about
(a) whether researchers were aware of the scripts; and (b) whether, or how, knowledge about the users' place in the script helped them during the search process. The intermediaries did not seem to make a strong connection in their own minds between research scripts and contextual understanding of the users' problematic situations *or* between contextual understanding and the improvement of online search results.

These results suggest that both users and intermediaries are constrained by their expectations about current systems and by their views of their own roles in the search process. Both groups of subjects were aware of research scripts but failed to use this awareness as a productive means of communication with each other and with the system. Both groups seemed constrained by their view of current IR systems: they tended to view the system, functionally, as a mechanism for the provision of bibliographic citations and not as a mechanism for assisting in the resolution of problematic situations. Thus they paid scant attention to situational variables and the attainment of research goals (i.e. had a limited view of the value of research scripts), and perhaps retrieved fewer relevant and/or more irrelevant documents than might otherwise be possible. Particular aspects of the interview results will be amplified below.

Intermediaries' views on situation components

When asked which problem statement components were most important, those components most readily and frequently mentioned by the intermediaries were topic-oriented keywords (and term synonyms) or other traditional online search parameters, such as language and date. Also commonly mentioned were the traditional user-oriented factors: user's 'level', purpose and previous search experience. Several of the intermediaries noted the importance of inquiring whether the user desired a comprehensive or narrow search. Almost all of the intermediaries required users to fill out forms which asked for this type of information. These forms also asked users to describe their request in a few sentences; the intermediaries seemed to consider this to be 'contextual' information. Thus, when the intermediaries noted that they considered the 'context' component important, they seemed mainly to be referring to broadening the context of the search topic. Although 'intended purpose' also appeared on the form used by many of the intermediaries, the user merely checked one of several categories, such as 'dissertation', 'article', and so on. Relatively little attention was paid by intermediaries to situational data. They seemed more inclined to inquire after situational factors when they were helping users at the reference desk than when they were preparing to conduct online searches.

All of the intermediaries said that 'topic' was the only component of a problematic situation that users consistently discuss spontaneously. Although the intermediaries generally maintained that other components were mentioned by empirical researchers (in addition to those they had previously listed for general users), there was not much consensus on what those other components were. Four of the subjects used phrases that all seemed to allude to the notion of the user's subject expertise (SI1 - 'knowledge of literature'; SI2 - 'critical areas'; SI3 - 'user sophistication'; SI4 - 'depth'). Two intermediaries mentioned that researchers sometimes predict the amount of information they expect to retrieve. One subject mentioned cost as a component of researchers' requests, another mentioned the population being studied, two noted that researchers might more often mention particular authors in the field whose work was relevant to their topic. The intermediaries seemed to feel that situational variables played a more dominant role in their interactions with empirical researchers than with users in general, but the nature of that role seemed problematic. S,l (Agriculture) said that 'how they intend to use it' came up more often with researchers. S,3 (Physical Sciences) said that context was more important, but that the researcher's purpose or agenda was very difficult to elicit and often was revealed only after the search. SI4 felt that it was common for researchers to give more contextual information, but seemed to consider this type of exchange merely as 'preliminary conversation' which did not have much to do with the search itself. SI5 said that researchers often talked less, in general, and attributed this to their sense of 'pride', i.e. they did not want to appear ignorant, particularly about their supposed area of expertise. SI5 was the only searcher who mentioned specific stages of the research script as part of the 'context' or 'purpose' components which were important when working with empirical researchers; she mentioned that users might note whether they were 'trying to come up with a theory', or 'developing a method'.

Search strategy formulation

Search intermediaries were generally unable to specify the manner in which they formulated or revised their search strategies according to the data in the problem statement components that they typically elicited from users. They discussed search strategies in very general and quite traditional terms, with the focus once again remaining on topic keywords and other common search parameters. Two intermediaries said, for example, that they used the keywords generated by the user to find synonyms in controlled vocabulary thesauri. One complained about the inability to specify the desired relationships among topic keywords or to locate documents in which a topic keyword is playing a specific role. Non-topic, situational components such as purpose were used by different searchers differently: in the selection of databases, the determination of the appropriate scope for the search, and in the consideration of the level of service to be delivered. Of special significance is the fact that the same set of demographic-like variables (user's educational level, search expertise, etc.) were used by different intermediaries to form different types of search strategies, given different situations. This suggests that perhaps searchers are aware of, and make use of situational data from users, even though they are unable to articulate the exact process involved. Perhaps they are simply not used to considering this type of data as part of the traditional search process. One reason for their tendency to undervalue and underutilise situational data is that they apparently find it difficult to convert such data into concrete search terms or to apply them to a search in other useful ways, given the limitations of current systems.

Two intermediaries (SI2 and SI5) noted that researchers in certain disciplines tended to have certain 'experimental styles' that were transferred to their information-seeking behaviour. S,5 noted that the component 'investigative preferences' was one she considered important when dealing with empirical researchers. The place of the empirical researcher in the research script is one special instance of situational data. It will be considered below.

Use of research scripts

Each of the intermediaries interviewed was able to describe a 'generic research script' that represented the steps or stages that most people who are engaged in empirical work must go through in order to complete their work. The scripts provided by intermediaries, however, were more general than those provided by users; script elaboration also seemed a more difficult task for the intermediaries. These scripts are listed in Table 12. The language and terminology used by the searchers in elaborating these scripts varied noticeably from one person to another. It may well be that the use of certain terms is discipline-specific. An analysis of the language used to describe the research process might be a helpful first step in the selection of terms for the presentation of a generic script online, or in the development of an interactive mode for aiding users in delineating their current position in the research script.

TABLE 12. *Generic research scripts from intermediary interviews*

S_I1(Agriculture):	1. Creative thinking;
	2. Background reading;
	3. Discussion with others;
	4. Developing methodology;
	5. Choosing equipment;
	6. Doing the work:
	a) gathering equipment, graduate assistants,
	b) gathering data,
	c) getting results/conclusions;
	7. Writing up report;
	8. Submitting report;
	9. Peer review.
S_I2(Soc. sci.):	1. Making sure it 'hasn't been done before';
	2. Literature review;
	3. Choosing procedures and instruments;
	4. Choosing means of data analysis;
	5. . . . [I suggests a few later steps; conversation proceeds in another direction]
S_I3(Phys. sci.):	1. Explore the field; (Applied sci.)
	2. Gather background information on techniques, topic;
	3. Choose specific research problem;
	4. [Do experiments (prompted by I)]
	5. Analyse results (S gave wordier description of this stage);
	6. Follow up unexpected results.
S_I4(Soc. sci.):	1. Have a clear idea of what they want to do;
	2. Find out what related work has been done by others;
	3. Find out how to do desired work;
	4. Find out 'how they'll know if they've done it', i.e. how to test validity and reliability of results.
S_I5(Phys. sci.):	1. Identify desired research problem; (Applied sci.)
	2. Make sure it hasn't been done before;
	3. Make sure it's significant;
	4. Trial and error in the lab;
	5. Library research to 'get to next step';
	6. Conclusions;
	7. Find out how and where to publish.

The intermediaries were divided in their opinions about whether or not the researchers themselves were aware of following a script (3 'yes'; 1 'maybe'; 1 'probably not'). The intermediaries also varied in their descriptions of how they would, could, or should use knowledge of the researcher's place in the script to formulate or modify a search strategy. They all felt that this was very difficult to do given current online systems. The intermediaries generally associated the search for certain types of information, which would be helpful at particular research stages, with the informal networks of empirical researchers, although several noted that they wished they could be more involved with the actual work of researchers. They thought that most researchers prefer to go to colleagues to resolve problematic situations related to current work. It is clear that both users and searchers make a definite distinction between the kind of help which current systems can offer (i.e. bibliographic tasks such as literature review) and the kind of help which the informal information network is capable of offering (i.e. selecting instruments, data analysis, etc.). We believe that IR systems could be profitably expanded to serve both these functions (though not exclusively).

3.2.3 Conclusions on users' situations

Typical search behaviour is far from optimal. Users seem to be constrained by the forms they fill out and the questions they are asked, while intermediaries are constrained by the capabilities of current systems. Users have some idea that the situational aspects of their information needs are important, as evidenced by their relevance judgements and other comments during our inverviews, but they generally do not spontaneously provide, and are not encouraged by intermediaries to delve into and discuss these aspects. Intermediaries cannot articulate strategies well. Therefore, gathering the expert knowledge for automatic search strategy formulation by a system would be very difficult. The critical importance of understanding situational aspects in order to enhance IR performance has been stressed recently by Saracevic and colleagues. Reporting on their extensive experimental results, they conclude that: 'the users' context (the problem at hand and the intent) is a most powerful element in the potential effect on retrieval effectiveness, and that exploring the context has a large potential payoff, while doing the search on the basis of question terms only (without elaboration) is the poorest way to go about it' [40, p. 203].

There is a good deal of evidence from our interviews that both users and intermediaries recognise that people engaged in empirical work follow a generic research script. It is also apparent that some databases currently provide search capabilities that are tangentially related to the user's position in the research script and that these capabilities are appreciated by those searchers who know how to exploit them (both in terms of possessing the technical knowledge of specific search techniques and of possessing the ability to draw out from users the needed situational information). We speculate, therefore, that a system which promoted a heightened awareness of the research script and allowed one to retrieve documents based, in part, on the user's place in the script, would offer substantial improvements over current systems.

4. IMPLICATIONS FOR SYSTEM DESIGN

Our studies of the discourse-level structure of empirical abstracts and of the various expressions of problematic situation and information need lead us to some suggestions for the design of information retrieval systems. At this point in time, we must still confine ourselves to somewhat general statements and high-level features.

1. *An IR system should include search features related to situational variables.*

Relevance judgements are based on both content- and context-oriented, or situational, factors. Besides topic, three related notions have emerged as major influences on relevance decisions:

1. a person's position in a research script;
2. what the person is (or will soon be) trying to do;
3. how the person intends to use the information (this is basic).

We assume certain relationships between these three:

1. What a person is trying to do can be described as taking one or more steps in a research script;
2. The use that a person intends to make of information is at a lower level of detail than a step in a research script, but there are typical uses at certain steps, so knowledge about intended research script steps conveys some information about intended information use.

Given these assumptions, we maintain that a system that makes use of a research script in its interaction with the user will bring the above three relevance determinants to bear in the search process. Documents should be described in both content-oriented and situational terms.

2. A generic research script should be presented to the user.
Our study indicates that a heightened awareness on the part of users of the existence of a research script, and of their current position in the script, could aid retrieval by encouraging them to focus more attention on the important situational aspects of their information needs. Script awareness seems helpful to users because it allows them to direct their search efforts to the accomplishment of the particular task at hand. A user would presumably be able to develop a more precise search if database abstracts were also structured into components related to the tasks in the research script. In addition, a user with this awareness should be better able to recognise relevant documents. Thus, even if a system only made very rudimentary use of script information, the simple act of raising the issue with the user should make a positive impact on retrieval effectiveness.

3. A new type of system should foster experimentation and evaluation.
We would design a new system in such a way that it can teach us something about the assumptions upon which it was built. This seems especially important as our assumptions have so far been supported only by a very small number of interviews with current system users. We would prefer to go ahead with the construction of a modest system whose use will offer a means of evaluating our ideas, rather than to spend a great amount of time and effort continuing to gather data to validate our ideas from users and searchers who are constrained by current systems. A new IR environment needs to be created in order realistically to test new ideas about IR system design and use.

4.1 Expert systems?
A system with the characteristics we have outlined might well be credited with intelligence. Intelligent retrieval has been described as that which 'involves the system's stored knowledge of its "world" (documents, users, topics, etc.) and of information about the user and his or her problem to infer which documents would enable that particular user to resolve or better manage his or her problem' [41, p. 367]. Along with the revival of interest in artificial intelligence generally, beginning in the last decade, came a substantial amount of research into expert systems for IR and a proliferation of prototype systems, most of them functioning within a rather limited domain [42-44]. With few exceptions, however, these systems fail to take into account non-topic aspects of the user's problem. Those that do attempt to incorporate this sort of information (e.g. Croft and Thompson [5]) are often able to make very little use of it. In most cases, expert systems are seen as a 'new generation' for doing the same thing that was done by traditional online systems: topic-based retrieval.

It was clear from our interviews with search intermediaries that they found it exceedingly difficult, if not impossible, to specify how or why they arrived at a particular search strategy given certain user input. As Feigenbaum [45] has noted: 'Experience has shown us that [an expert's] knowledge is largely heuristic knowledge, experiential, uncertain - mostly "good guesses" and "good practice" in lieu of facts and rigor... much of this knowledge is private to the expert, not because he is unwilling to share publicly how he performs, but because he is unable'. Others in the information science field have echoed these views [41, 46, 47]. Several of the intermediaries we spoke with averred that they would use identical demographic-type information about different users differently, depending on the situation, but they gave only broad, general examples of this and did not appear to follow any systematic 'rules of thumb' that they could articulate. (In addition to indicating a lack of rules, this would also suggest caution when adopting a user model as part of a system design.) Thus, we are inclined to agree with Brooks [41] that: it is not feasible, at present, to think of building an expert system that carries out intelligent retrieval . . . The great benefit of expert systems research to IR has been . . . at the conceptual level; that is, it has stimulated interest in the exploration of what knowledge is required toachieve intelligent retrieval, in the problem-solving strategies used by human retrieval experts, and in how documents could be represented so that the system would 'know' something about their contents. The influence of expert systems has shifted IR research from a paradigm concerned with retrieval algorithms to one in which users, retrieval heuristics, knowledge, and human-computer interaction are key themes.

We do not mean to say that it would never be possible to have an outstanding expert system for information retrieval, but only that we have found one parameter which is basic to users' relevance assessments, namely situation in their work efforts, and that the purported experts in the field of searching (intermediaries) are not expert in how to make use of this information.

Our final comment in relation to an 'expert systems' approach is methodological. We considered, in some detail, a frame-based, rule-driven system design, and put it to one side because its complexity was so great that, if it worked much better than a conventional IR system, we would not know precisely why, and if, as is much more likely, it worked about equally well, we would have little idea how to improve it. So, our strategy is to move ahead cautiously, and suggest an evolving family of systems, to be used as a vehicle for exploiting what we think we know, and for conducting sound experiments to resolve new questions.

4.2 Suggested system - overview

We have come to the conclusion that a system which makes use of a research script and discourse-level structure of documents in its interaction with the user will introduce factors which are important influences on relevance: information uses and the user's situation. However, we cannot yet go much further in specifying the mechanisms whereby a system could use research script information in predicting relevance. We are prepared to advance some hypotheses, which can be explored through the use of a quite straightforward progression of systems:

1. Awareness of a research script helps a user to understand his or her information need. This, in turn will help a user formulate a better query, and be better able to recognise relevant documents.
2. Thus, even if a system only made very rudimentary use of script information, the simple act of raising the issue with the user will make a positive impact on retrieval effectiveness.
3. Discourse-level structure of abstracts is a suitable vehicle for interacting with a user on the matter of research script.
4. Discourse-level structure constitutes a new informational dimension in the description of documents, distinct from topic, which is understood (and therefore usable) by searchers.
5. A search that takes account of the discourse-level structure of abstracts will be more effective than one that doesn't.
6. Any sensible use of discourse-level structure in a query language will improve retrieval.

We suggest that the retrieval system be based on a three-module design, and that development of versions take the form of evolution in the functionality of these modules and the relationships between them. Each of the three modules interacts with the user. They are as follows:

RS: research script dialogue. This will interact with the user on the matter of his or her situation, in terms of steps in the research script (see Table 5). The complete, generic research script is a large structure; too large for a user to appreciate easily without a carefully designed interface, for which we can suggest some helpful features: a graphic (map) representation, with typical -chronological relationships indicated; alternative vocabulary available (gleaned initially from our interviews) to enable the system to refer to places in the script in the user's terms; a set of suitable questions to help the user focus on parts of the script that are of the most immediate importance.

SF: search statement formulation. This will provide helpful tools to assist the user in constructing a formal query, involving not only keywords, but the occurrence of words and phrases in selected abstract components, and the presence or absence of specified component-types in abstracts.

RP: retrieval program. This searches the database of *structured* abstracts for those matching the formal search statement. There are a number of retrieval models (e.g. Boolean, probabilistic) which can be adapted for use with document representations arising from analysis into discourse components. This module also interactively manages display of retrieved abstracts, indicating their structural relationship to the search statement, and collecting relevance feedback from the user.

The required database will be built with the aid of a discourse-level analysis program. Each abstract is divided into a hierarchy of discourse-level components, from which a list of term/component pairs is constructed. For instance, (elephant/RESULT) signifies that the word 'elephant' occurs in the RESULT component of the abstract under consideration. On the basis of our work so far (including Liddy's empirical work), we know that such a description can be produced by humans, and are prepared to assume that it can be done automatically with reasonable accuracy.

The complexity of these modules, and the relationships between them (control structure and communication) would be low in *version one,* and increase only as we learn more. Version one consists of the three major modules listed above, with no system-controlled links between them, other than the essential, and conceptually straightforward SF to RP (i.e. formal queries constructed in SF can be passed to RP for searching). The overall control structure in version one is to begin with the sequence RS. SF, and then allow the user to switch from one module to another at will. There is no direct link between RS and SF (i.e. no use is made *by the system* of research script information), although we expect to see user behaviour affected simply by having to go through RS. The ability to switch back and forth between modules allows the user to adjust the search formulation in the light of retrieved abstracts (which are annotated with their structure), but version one has no facilities for suggesting improvements.

All of the hypotheses listed above can be addressed to some degree through behavioural experiments using this primitive version.

SUMMARY

We set out to find ways of introducing cognisance of the user's problematic situation into information retrieval system designs. Currently, systems do not have explicit representations of situations, and making inferences in this domain from the data they do maintain would seem to be extremely difficult. We have explained our conception of information retrieval as part of a communication process between people who share experience of similar or related tasks (rather than as a question-, or query-answering process). We noted the evidence that, at the discourse-level of analysis, language often exhibits structure that reflects the structure of the tasks in which its users are engaged. Thus, we formulated a plan to determine the discourse-level structures of a common IR text-type and of users' problem statements, and to find then predictable and usable relationships between them. We offer the following summary and remarks on our progress.

Using the structure of empirical abstracts delineated by Liddy [17, 18], we have demonstrated two methods of automating the discourse-level analysis of these texts. One program performs a probabilistic analysis, and the other uses a neural-like network. Both are in early stages of development at the time of writing and show promise rather than display what we would accept as good performance. It is our view that reasonably accurate automatic analysis of empirical abstracts is feasible.

Problem statements do not lend themselves to this kind of analysis. We expended much thought and employed a variety of techniques, and yet were unable to detect in (oral) natural language problem statements, the discourse level components we had presumed were present in them. They were neither conceptually nor linguistically apparent. Hence, we decided to approach the problem from a different direction. We began by defining a model of the users' work patterns, the generic research script, and investigated its validity and potential utility through open-ended interviews with users and intermediaries. The research script bears an obvious similarity to the discourse-level structure of empirical abstracts. It appears from the interviews that the generic research script is indeed a meaningful construct to both users and intermediaries. Moreover, a researcher's position in the research script is a predictor of relevance judgements on retrieved documents. It also emerged that, because current systems offer no appropriate facilities, intermediaries make very little use of situational or research script-position information from the user. Not surprisingly, the indications are that there is no expertise on the use of situational information in searching which could be drawn upon for the design of a rule-based system.

From this evidence, we predict that the introduction of two new categories of information into information retrieval systems will improve the quality of their responses. Firstly, the discourse-level structure of texts should be available for searching and display, and secondly, a research script should be invoked during the user's dialogue with the system, even if only very superficially.

ACKNOWLEDGEMENTS

We should like to thank Inho Chang, Hannah Francis, Pradeep Singh and Marilyn Smith for contributions made in the course of the project, and Becky Freeland for her painstaking clerical work, including managing the transcription of lengthy interviews. We are also indebted to the Northeast Parallel Architectures Center at Syracuse University for access to a Connection Machine and for technical support. Various aspects of the work were funded by grants from USWest Advanced Technologies and the Council on Library Resources.

REFERENCES

1. DERVIN, B. *and* NILAN, M. Information needs and uses. *In:* WILLIAMS, M.E., *ed. Annual Review of Information Science and Technology, 21.* Amsterdam: Elsevier Science Publishers, 1986, 3-33.

2. HEWINS, E.T. Information need and use studies. *In:* WILLIAMS, M.E., *ed. Annual Review of Information Science and Technology, 25.* Amsterdam: Elsevier Science Publishers, 1990,143-172.

3. DANIELS, P.J. Cognitive models in information retrieval: an evaluative review. *Journal of Documentation, 42* (4), 1986, 272-304.

4. KOBSA, A. *and* WAHLSTER, W., *eds. User models in dialog systems.* Berlin: Springer-Verlag, 1989.

5. CROFT, W.B. *and* THOMPSON, R.H. I3R: A new approach to the design of document retrieval systems. *Journal of the American Society for Information Science, 38* (6), 1987, 389-404.

6. BRAJNIK, G. *et al.* IR-NLI II: Applying man-machine interaction and artificial intelligence concepts to information retrieval. *Proceedings of the Eleventh Annual International ACMSIGIR Conference on Research and Development in Information Retrieval.* Association for Computing Machinery, 1988, 387-399.

7. GROSZ, B. *and* SIDNER, C. Attentions, intentions, and the structure of discourse. *Computational Linguistics, 12,* 1986, 175-204.

8. SEARLE, J.R. *Speech acts.* Cambridge: Cambridge University Press, 1969.

9. BRITTON, B. *and* BLACK, J. Understanding expository text: from structure to process and world knowledge. *In*: BRITTON, B. *and* BLACK, J., *eds. Understanding expository texts: a theoretical and practical handbook for analyzing explanatory text.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1985, 1-9.

10. VAN DIJK, T.A. *Macrostructures: an interdisciplinary study of global structures in discourse, interaction, and cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

11. MALONE, T.W., GRANT, K.R., LAI, KUM-YEW, RAO, R. *and* ROSENBLITT, D. Semistructured messages are surprisingly useful for computer-supported coordination. *ACM Transactions on Office Information Systems.* 5(2), 1987,115-131.

12. FRANCIS, H. *and* LIDDY, E.D. Structured representation of theoretical abstracts: implications for user interface design. *Proceedings of the 1989 American Society for Information Science Mid- Year Meeting* (in press).

13. HAHN, U. *and* REIMER, U. *The TOPIC project: text-oriented procedures for information management and condensation of expository texts.* Constance, West Germany: University of Constance, 1985.

14. GRISHMAN, R., HIRSCHMAN, L. *and* FRIEDMAN, C. Natural language interface using limited semantic information. *Proceedings of the Ninth International Conference on Computational Linguistics.* Amsterdam: North-Holland, 1982,89- 94.

15. BROOKS, H.M., DANIELS, P.J. *and* BELKIN, N.J. Research on information interaction and intelligent information provision mechanisms. *Journal of Information Science, 1*2(1), 1986, 37-44.

16. HOFFMANN, E. Defining information: an analysis of the information content of documents. *Information Processing and Management, 16,* 1980, 291-304.

17. LIDDY, E.D. The *discourse-level structure of natural language texts: an exploratory study of empirical abstracts.* Unpublished Ph.D. dissertation, Syracuse University, School of Information Studies, Syracuse, NY, 1988.

18. LIDDY, E.D. The discourse-level structure of empirical abstracts: an exploratory study. *Information Processing and Management, 27* (1), 1991, 55-81.

19. ROSCH, E., MERVIS, C., GRAY, W., JOHNSON, D. *and* BOYES-BRAEM, P. Basic objects in natural categories. *Cognitive Psychology, 8,* 1976, 382-439.

20. RUMELHART, D.E., MCCLELLAND, J.L. *and* THE PDP RESEARCH GROUP. *Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, Mass: MIT Press, 1986.

21. KHANNA.T. *Foundations of neural networks.* Addison-Wesley, 1990.

22. BALAKRISHNAN, B. *A study of the use of connectionist networks to identify the discourse-level structure of empirical abstracts.* Unpublished PhD dissertation, Syracuse University, School of Information Studies, NY (forthcoming).

23. SMOLENSKY, P. Information processing in dynamical systems: foundations of harmony theory. *In:* RUMELHART, D.E., MCCLELLAND, J.L. *and* THE PDP RESEARCH GROUP. *Parallel distributed processing: explorations in the microstructure of cognition. Vol. 1: Foundations.* Cambridge, Mass: MIT Press, 1986, 194-281.

24. SALTON, G., YANG, C.s. *and* YU, C.T. A theory of term importance in automatic text analysis. *Journal of the American Society for Information Science, 26* (1), 1975, 33-44.

25. MCCLELLAND, J.L. *and* RUMELHART, D.E. An interactive activation model of context effects in letter perception: part I. An account of basic findings. *Psychological Review, 8* (5), 1981, 375-407.

26. HILLIS, W.D. *The connection machine.* Cambridge, Mass: MIT Press, 1985.

27. MCCLURE, C.R., BISHOP, A.P., DOTY, P. *and* ROSENBAUM, H. The *National Research and Education Network (NREN): Research and policy perspectives.* Norwood, NJ: Ablex, 1991.
28. TAYLOR, R.S. Information use environments. *In:* DERVIN, B. *and* VOIGT, M.J., *eds. Progress in Communication Sciences, Vol. X.* Norwood, NJ: Ablex, 1991, 217- 255.
29. ODDY, R.N., PALMQUIST, R.A. *and* CRAWFORD, M.A. Representation of anomalous states of knowledge in information retrieval. *Proceedings of the 49thannual meeting of the American Society for Information Science.* Medford, NJ: Learned Information, 1986, 248-254.
30. GRISHMAN, R. *and* KITTREDGE, R., *eds. Analyzing language in restricted domains: sublanguage description and processing.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1986.
31. GARVEY, W.D. *Communication: the essence of science.* Oxford: Pergamon Press, 1979.
32. KRAUT, R., GALEGHER, J. *and* EGIDO, C. Relationships and tasks in scientific collaborations. *Human Computer Interaction, 3* (1), 1988, 31-58.
33. MAVOR, A.S., KIDD, J.S. *and* VAUGHAN, W.S., *Jr. Cognitive models of scientific work and their implications for the design of knowledge delivery systems.* Final report on NSF grant DS 176-24385. Annapolis, MD: W/V Associates, 1981.
34. RICHARDSON, S.S., DOHRENWEND, B.S. and KLEIN, D. *Interviewing: its forms and functions.* New York: Basic Books, 1965.
-35. BRENNER, M. Intensive interviewing. *In:* BRENNER, M., BROWN, J. *and* CANTER, D., *eds. The research interview: uses and approaches.* London: Academic Press, 1985, 147-162.
36. NILAN, M.S. *and* FLETCHER, P.T. Information behaviors in the preparation of research proposals: a user study. *Proceedings of the 50th annual meeting of the American Society for Information Science.* Medford, NJ: Learned Information, 1987, 186-192.
37. KWASNIK, B.H. A method for investigating classificatory situations. *Fifth Symposium on Empirical Foundations of Information and Software Sciences.* Roskilde, Denmark: Riso National Laboratory, 1987.
38. WEINBERG, B. Why indexing fails the researcher. *Proceedings of the 50th Annual Meeting of the American Society for Information Science.* Medford, NJ: Learned Information, 1987, 241-244.
39. ROSENBAUM, H. and NEWBY, G.B. An emerging form of human communication: computer networking. *Proceedings of the 53rd Annual Meeting of the American Society for Information Science.* Medford, NJ: Learned Information. 1990, 300- 325.
40. SARACEVIC, T., KANTOR, P., CHAMIS, A.Y. *and* TRIVISON, D. A Study of information seeking and retrieving. I. Background and methodology. II. Users, questions, and effectiveness. III. Searchers, searches, and overlap. *Journal of the American Society for Information Science, 39* (3), 1988, 161-216.
41. BROOKS, H.M. Systems and intelligent information retrieval. *Information Processing and Management, 23* (4), 1987, 367-382.
42. POLLITT, A.S. CANSEARCH: an expert systems approach to document retrieval. *Information Processing and Management, 23* (2), 1987, 119-138.
43. COHEN, P.R. *and* KJELDSEN, R. Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management, 23,* 1987, 255-268.
44. MUKHOPADHYAY, U. *et al.* An intelligent system for document retrieval in distributed office environments. *Journal of the American Society for Information Science, 37,* 1986, 123-135.
45. FEIGENBAUM, E.A. Themes and case studies of knowledge engineering. *In:* MICHIE, D., *ed. Expert systems in the micro-electronic age.* Edinburgh University Press, 1979.
46. BATES, M.J. Information search tactics. *Journal of the American Society for Information Science, 30* (4), 1979, 205-214.
47. VICKERY, A. *and* BROOKS, H. Expert systems and their applications in LIS. *Online Review, 11 (3),* 1987,149-165.

(Revised version received 14 October 1991)