10 Feedback and interaction in information retrieval

Efthimis N. Efthimiadis and Stephen E. Robertson

Since the late, unlamented days of batch search systems, we have come to assume a degree of interaction in any search process (which indeed was always present when searching card catalogues and printed indexes). Online retrieval of whatever sort allows the searcher to make a stab at a search, and then try again if the results are not satisfactory. Whether an experimental or theoretical information retrieval researcher, or a manager or user of a modern online system, it is part of our paradigm that interaction is desirable – that nobody wants to go back to the situation where only one attempt could be made in a reasonable timescale. This paradigm is made explicit and extensively analysed in Belkin and Vickery (1985).

Nevertheless, much of the theoretical work on information retrieval contrives to ignore the interaction process (see, for example, Belkin and Croft, 1987). The aim of this chapter is to look at, and attempt to categorize, some of the kinds of interaction and feedback that can occur, at the micro level (e.g. which information elements are transmitted/received/expected at various stages in the information retrieval process).

This review is not intended to be exhaustive, either of the literature or of the systems. Certain items have been selected from the literature and from systems known to the authors, with a view to illustrating specific ideas.

Some background

Intermediaries and intermediary mechanisms

The study of interaction in information retrieval (IR) is complicated by the existence of various kinds of intermediary mechanism (human or machine). Many IR situations can be conceptualized in something like the form indicated in Figure 10.1. The 'intermediary' may be a human being, or a user interface, front-end system or expert system. There may, indeed, be more than one intermediate stage.

In considering the interaction/feedback possibilities in Figure 10.1, the picture is decidedly complex. There may be interaction between the end

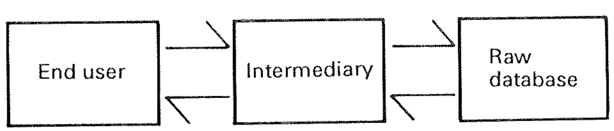

Figure 10.1 The two-stage model

user and the intermediary, without reference to the database; interaction between the intermediary and the database without reference to the end user; or interaction that spans all three parties. In this last case, messages from the end user may or may not be interpreted by the intermediary before being transmitted to the database; and similarly with return messages. 'Interpretation' may mean anything from a slight rearrangement to a complete change of form.

In fact much recent research is informed by the perception that some (human or machine) intermediary is desirable, specifically to encourage interaction and feedback. It is usually suggested that this desirability is simply the result of the inadequacy of the 'raw' database, though the conceptualization of IR as involving an intermediary mechanism seems to represent some deeper recognition of the centrality of feedback.

Systems

Taking a narrow view of what constitutes the IR system in a given IR situation, it would probably be seen as the raw database in the terminology of Figure 10.1, or the 'retrieval mechanism' in the terminology of Belkin and Croft (see below). But any appreciation of general systems theory should warn against such a narrow view. Such an appreciation would suggest taking at least the whole of Figure 10.1 as the 'system' to be examined. In that view, the user would not be seen as an independent entity, but as part of the system.

But looking at the situation from the user's point of view, it must be clear that users see the IR system as something outside themselves, which they approach and from which they expect something. In this view, the system must be everything in Figure 10.1 except the user (but including the interaction between the intermediary and the user).

For the IR theorist, it may be that some compromise between these last two views would be appropriate. Taking into consideration the model of communication described in the Belkin/Vickery review, in which each participant has a model of the other, the user's model of the system can be taken to be itself part of the system. Thus the boundary of the system (that is, the system of concern to the IR theorist or designer) would occur somewhere within the mind of the user.

In this chapter, 'feedback' refers mainly to information crossing the boundary of the system (in either direction). But within the total system under consideration there are subsystems, as indicated in Figure 10.1; feedback between the subsystems may also be of concern.

Retrieval mechanisms

Belkin and Croft (1987) have developed a classification of retrieval mechanisms, summarized in Figure 10.2. In their definition, a retrieval technique (RT) is a technique for comparing the query with the document representations. RTs are further classified according to the characteristics

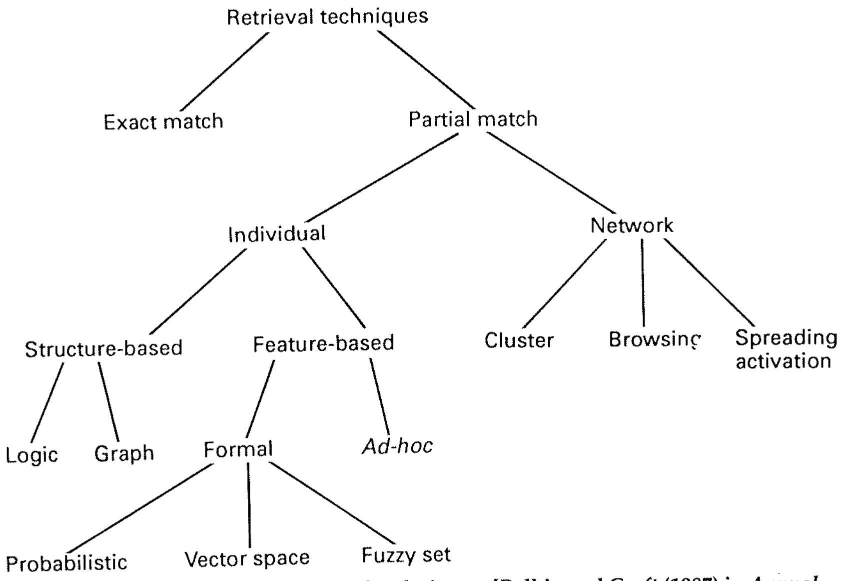

Figure 10.2 A classification of retrieval techniques. [Belkin and Croft (1987) in *Annual Review of Information Science and Technology* vol 22, Elsevier, by courtesy of the American Society for Information Science.]

of the retrieved set of documents and the representations that are used. Some RTs fall into more than one category, and others are a mixture of techniques from different categories.

The first distinction is between exact and partial match. In the former category are RTs that require any retrieved item to match the query exactly, such as Boolean systems, those using the pseudo-Boolean free text operators, and string-searching systems. Partial match RTs, by which items may be ranked according to degree of match, are further divided into individual and network RTs. The former are based on matching queries against document representatives treated individually, whereas the latter make direct user of inter-document links of some kind.

One of the things that a consideration of interaction and feedback does is to nibble away at the black-and-white distinctions in the Belkin/Croft model, in that, for example, some exact-match techniques which include facilities for feedback and query expansion begin to look much more like partial-match methods. Again, query expansion applied to individual RTs can make them behave rather like network RTs.

This is not to deny the validity of the model, but merely to recognize the inherent complexity of many retrieval situations, which is not always adequately represented by simple models.

Forms of feedback

Considering the communication between the user and the rest of the system (including any intermediary mechanism), several types of information might be identified. First to be considered is the information going from the system to the user.

Systems may provide information about their own facilities. Examples include error messages, command menus and help screens. They may provide subject-related information, such as menus of subjects or thesaurus or dictionary extracts. Finally, they may provide information deriving from an actual search. This last category includes partial or full records, and statistical information such as number of postings. Some types of information may come into more than one category, for example, an analysis of term occurrences in a retrieved set comes into the second and third groups.

In the reverse direction, users may provide commands summoning particular system facilities. They may provide subject descriptions (terms, phrases, natural language sentences). Finally, they may provide selections from explicit alternatives presented to them. This last includes selections from command menus, selections from subject menus, and responses to offered items or terms. Again, a particular user response may combine categories – e.g., a natural language statement may contain commands as well as subject terms.

Any of the above may occur at any stage of the dialogue. A simple sequence of user request/system response (as in a batch system), or a multiple sequence where all user requests have been decided prior to the search, are regarded as not involving feedback. Feedback can be said to occur when the user responds to a system response, i.e., where a second or subsequent users request is determined or modified by the preceding system responses.

The concern, therefore, is with ways in which (a) users may use system-generated information to help them reformulate some aspect of their problem; (b) systems may provide information intended to prompt or help the user to reformulate; (c) part of the reformulation process may be included as a system function, and the system may therefore invite user responses in an appropriate form.

The above description is phrased as though the 'system' (in the sense defined above) is a machine. If that part of the system visible to the user is human (e.g. a human intermediary), it is likely that a more complex analysis of information types is required (see, for example, Brooks, Daniels and Belkin, 1985). If it is a machine, then any query reformulation within the system can be regarded as a form of machine learning. Rada's (1987) distinction between 'knowledge-sparse' and 'knowledge-rich' learning is a useful one, implicit in much of the discussion below.

The literature

The literature of IR seems to divide into two. On the one hand there is IR theory literature, which is either concerned just with the retrieval mechanism, or incorporates only a very stylized form of feedback, e.g., simple relevance feedback, see below; on the other hand, there is the online searching literature, which is pragmatic rather than theoretical, but very often addresses feedback problems, e.g., Bates' search tactics. It is the intention of this paper to attempt some drawing together of these separate strands.

Also, some of the relevant work consists of descriptions of systems – experimental or operational. Particular aspects of these systems may be relevant to various stages of the present analysis. Thus no attempt is made to provide a complete description of, say, CITE (Doszkocs, 1983) or Plexus (Vickery *et al.*, 1986 and 1987), though specific characteristics of each system are mentioned at appropriate places below.

Traditional Boolean searching

The traditional Boolean searching has been the subject of many textbooks (Henry *et al.*, 1980; Markey and Cochrane, 1981; Meadow and Cochrane, 1981; etc.) and articles in journals as *Online Review*, *Online* and *Database*.

The figure describing 'steps in the presearch interview and the online search' has been borrowed from the Meadow and Cochrane (1981) book on online searching (see Figure 10.3).

Steps

1. Clarifying and negotiating the information need and search objectives.
   - Interviewing the information requester clarifies the narrative form of the request and determines search objectives:
     (a) retrieve all relevant items (high recall);
     (b) retrieve only relevant items (high precision);
     (c) retrieve some relevant items.
   - Identify constraints (e.g., books only as output or only in English, or only if published after 1975)
2. Identifying relevant online system and data bases.
   - Determining which online system and database to use first, which next, etc.
3. Formulating basic search logic and planning search strategies.
   - Analyzing the search topic into parts called facets or concept groups. Planning approaches to search strategy for combining concepts of the topic.
4. Compiling the search terms.
   - Choosing indexing terms from the databases thesaurus or other printed word lists.
   - Selecting terms for free text searching of the subject-conveying fields (title, abstract, etc.).
   - Deciding to use thesaurus and alphabetic word lists online.
5. Making output choices.
   - Choosing limits on, and printing of, output.
   - Selecting an approach to search strategy that best satisfies the search objectives expressed by the requester.
6. Conceptualizing the search as input to the retrieval system.
   - Arranging the search terms into concepts or facets for search strategies using features of the retrieval system, for example, word proximity.
   - Noting most important and less important concepts groups and deciding on sequence of input to access these concept groups efficiently.
   - Restricting or limiting output based on search objectives.
7. Evaluating preliminary results.
   - Reviewing search results, step by step.
   - Considering alternative search strategies to meet search objectives (recycling steps 1–6).
8. Evaluating final results.
   - Determining requester's satisfaction with search results.

Figure 10.3 Steps in the pre-search interview and the online search. [From Meadow and Cochrane (1981) *Basics of Online Searching*. ©John Wiley and Sons.]

These steps will now be examined and the different forms of feedback that arise from the user–intermediary interaction discussed. It is worth noting here that even if the end user searches alone, there will still be a need to clarify and negotiate the request and to know enough about the language of the database being searched to match the request with the basic index provided. In other words the user would have to follow all the steps on his or her own.

Steps 1, 2 and 4 could be said to belong in the pre-search procedures. Steps 3, 5, 6 and 7 are activities used in both pre-search procedures and during the search and step 8 is a post-search action. The user–intermediary/intermediary–system interactions would determine the order of the steps and any possible repetitions of them. The steps, thus, are not in a fixed order, except the first one, which involves user efforts to express and delegate their information need to the system, i.e. here the intermediary. The pre-search reference interview (for example, see Eichman, 1978; Knapp, 1978; Markey, 1981; Taylor, 1968) is the user's first interaction, and the feedback from it creates and links the conceptual analysis of the request (query) and the manipulation of concepts to develop a search strategy.

It is known that searchers have their own stereotypes both about how to deal with users and the online services, and also of how to conduct searches. At this pre-search stage the intermediary and the searcher are involved in a cognitive exchange. The stages of this interaction – e.g., steps 1–6 but especially step 3 – involve feedback that may change the searcher's – human intermediary's and/or end user's – view of the query. A Boolean search system has nothing that could be described as a dynamic cognitive structure (Ingwersen, 1984). Hence any evolution in the interaction has to involve changes in the cognitive structures of the human beings.

Having dealt with steps 1–6 the searcher then goes online. At this interactive session the search formulation is being submitted to the database and feedback is involved as soon as the searcher receives messages that would alter the search strategy. As a general rule, it would be fair to assume that any online search will involve feedback and modifications to the search formulation – except perhaps in the case of a one concept one statement search for say an exhaustive literature search or for a known item search. Thus, the searcher would have to iterate some if not all of the steps 1–6. The work by Bates (1979a; 1979b; 1981; 1986; 1987), Fidel (1985), Harter and Peters (1985) etc. deals with this problem and suggests various tactics (or moves) to be made during the search. All these tactics effectively are suggestions for feedback to the database as responses to its messages, e.g., postings, error messages, etc.

System help, command menus

Consider the form of feedback given in the interaction with a commercial host at a typical online session. The host computer would usually provide error messages, help facilities, and perhaps command menus.

Error messages are usually terse messages in response to a user command that the system cannot recognize or respond to. An error message may be just a number, e.g. E1234, which means nothing to the user apart from the understanding that something has gone wrong. More often, it would be a 'message in natural text', which is supposed to be self-explanatory. Thus, the error message may look like 'E1002 Unrecognized command – reenter or return to continue with last command' in DataStar or 'Invalid Argument' in ESA/IRS. So, if the computer does not recognize a command it will always respond in the same way.

Help facilities in online hosts, if they exist, tend to be unsophisticated and not at all context-dependent. For example, DataStar does not have an online help facility, whereas ?, ?EXPLAIN, HELP ?, ?HELP, .HE, would invoke the help facilities in ESA/IRS and Infoline, Dialog, Blaise, Dimdi, and Télésystèmes Questel respectively. This will introduce a 'master help menu' from which users choose the topic in which they need assistance. (Usually, unless one knows the exact name of the topic needed one would have to refer back to the master menu each time.)

As a contrast to the traditional online IR systems come the help facilities found in online catalogues. These are context-specific although still specific to system-context. For example, the CLSI OPAC offers different help facilities depending on which module of the system and/or stage of the search the user is in. So, if users ask for help before starting a search they will get different help information from that if they ask for help in the middle of, say, a title search. However, this context-dependence relates only to the context as perceived by the system. A request for help after a keyword search will produce the same screen whether the search produced 1000 items or none at all. An extensive discussion on OPACs is given by Hancock and Mitev elsewhere in this volume (Chapter 4).

Within the online interactive session, different support is necessary and required at different levels (Norman, 1984). A more careful design of menus (cf. Shneiderman, 1986) and of the user interface (cf. Draper and Norman, 1985) would have beneficial effects on the user–system interaction.

On the whole, from the aforementioned description of system facilities it is clear that these at present provide only a rudimentary form of feedback to the user. This (by intention) affects only the user's model of the system and therefore, should not be regarded as feedback in the sense defined in this chapter.

A note on query expansion

Query expansion refers to a process by which the system supplements the terms supplied by the user with additional terms. (More generally, query modification may involve deletion of terms as well.)

Query expansion as such is not a form of feedback. One could envisage a system which put query terms through some completely automatic process, resulting in new terms being added or additional matches being allowed. Indeed, such systems exist; an example is an automatic suffix stripping algorithm applied to any user-supplied terms, such as that in CUPID (Porter, 1982). Another example is an automatic matching of natural language terms against thesaurus entries, without reference back to the user, as is sometimes effected in Plexus (Vickery *et al.*, 1986 and 1987). In such a system, the expansion can be treated as part of the retrieval mechanism in the Belkin/Croft sense.

But query expansion may also either be a form of feedback to the user, or be the result of a feedback operation, or both. Thus query expansion crops up in several places in the discussions below.

Some general considerations apply to any form of query expansion. In particular, it is appropriate to ask the source of any extra terms supplied by the system. Possible sources include: a knowledge structure such as a thesaurus; an algorithmic process such as a suffix-stripper; characteristics of the collection of documents, such as term clusters; or terms extracted from documents retrieved in an earlier iteration of the search. The first three will be regarded in what follows as examples of 'knowledge structures' (taking the phrase broadly), independent of the search process. The last example clearly relates to a particular kind of feedback process.

Simple relevance feedback

In contrast to the manual/intellectual query reformulation described earlier, where the task falls to the searcher, it is possible to conceive of a system which takes over this task entirely, requiring only some simple yes–no judgements from the user. This automatic query reformulation process is called relevance feedback (Salton and McGill, 1983). Its aim is to improve the retrieved set by removing unwanted documents and adding more wanted documents without the user consciously constructing new search strategies, and by using relevance or nonrelevance information obtained from the user (McGill and Huitfeldt, 1979).

The typical automatic relevance feedback operation involves an initial search with a user-supplied query and an initial retrieval of certain documents. Then, from a display (usually of titles or abstracts of the retrieved documents) the searcher identifies/chooses some relevant documents. Those documents are used to modify the query by reweighting and/or adding terms that appear useful and by deleting terms that do not. This process creates a new query which resembles the relevant documents more than the original query does.

If we consider the information involved in the user-system communication in simple relevance feedback, it goes like this:

1. user gives system initial query;
2. system gives user document description;
3. user gives system relevance judgements.

The difference in the form between (1) and (3) suggests that there must be some kind of intermediary mechanism involved, since a raw database would normally accept queries in one form only. This intermediary mechanism may have to transform the initial query into a suitable formulation for searching, but will certainly have to transform the relevance information into such a form, and indeed combine the two kinds of information.

In its simplest form, feedback could be based on one document only. For example, after displaying a single document, the system could invite the user to see more documents like the one on display. Here the intermediary mechanism could use the classification scheme and present the user with books of the same class-mark as the first viewed. (The authors know of no system restricted to this simple method.)

Another simple form is when judgement is based on a set as a whole (as apparently suggested by Pietilainen, 1983). Here, a set derived from a previous search becomes the seed for the new query formulation. The method uses 'searchonyms', i.e., terms which might be regarded as synonyms for the purposes of a particular enquiry and derived from terms contained in the seed set.

Automatic feedback, generally, could be implemented in various ways depending on the retrieval technique used, e.g. vector space, probabilistic, etc., and also on the methods used to select terms for the feedback query. Four term selection methods for query reformulation can be distinguished. The first relies entirely on the original query and uses only those terms in the new one (Robertson and Spark Jones, 1976). This method has been successful implemented in a front-end system called CIRT (Robertson and Bovey, 1983; Robertson *et al.*, 1986). The second method uses terms from the original query and also adds terms from some other source, e.g., from all the terms adjacent to them in the maximum spanning tree (MST) (van Rijsbergen, Harper and Porter, 1981) or nearest neighbour (NN) terms (Smeaton and van Rijsbergen, 1983). The third method is a mixed method because it uses combinations of the terms derived from the original query and from the documents retrieved and judged relevant as found in the work of Salton and his colleagues (for example, Wu and Salton, 1981; Salton, Fox and Voohees, 1985) and in the latest version of the OKAPI online catalogue (Stephen Walker, 1988, personal communication). Finally, the fourth method abandons the terms from the original query and uses only terms found in the retrieved set of documents (Dillon and Desper, 1980; Dillon, Ulmschneider and Desper, 1983).

In all cases, after the initial query formulation, the only form of feedback to the user is items, and from the user is choices of items. The query reformulation is trusted entirely to the intermediary mechanism.

Prompted query definition/expansion

This section considers methods where the user is offered search terms (in one form or another) as part of the reformulation process. These could be based either on some form of knowledge structure (such as a thesaurus), or on the results of an earlier search.

Based on a knowledge structure

The knowledge structure (KS) of the intermediary mechanism, which as defined earlier is either human or machine, plays a major role in the interaction and is clearly of great importance for the success of the online search. In the case of a human intermediary, the model is determined by education, training and the development of skills, and experiences – both social and individual. Part of the work of Belkin and his colleagues is concerned with investigating the functions carried out by intermediaries (Brooks, Daniels and Belkin, 1986). They propose to proceed by building a number of different expert systems, each dealing with one aspect of the intermediary's KS, and all integrated as a distributed expert system based interface which would simulate the behaviour of the human intermediary (Belkin, Seeger and Wersig, 1983).

However, for most present machine-based intermediary mechanisms, the model would be static or passive depending on their level of sophistication. For example, the commercially available intermediary systems like Sci-Mate, Pro-Search, Easy-Net, etc. have automated only the mechanical operations of the online search (Hawkins, 1988). At best, with a fixed or passive KS (Ingwersen, 1984), the intermediary can select part of that structure to present to the user, thereby inviting the user to choose one or more elements.

Interpreting the phrase in a very broad way (as described above), the KS could be based in some way on the collection, or it could be independent of the collection.

Examples of feedback based on the collection are the EXPAND (or ROOT) command and the INSTRUCT term-clustering and morphological expansion modules. The EXPAND command provides a form of feedback from a KS of the 'raw' database which is the dictionary file. The user is given an alphabetical listing of descriptors and free-text terms to choose from and modify her/his KS. The INSTRUCT term clustering technique (Wade and Willet, 1988) identifies keyword stems which are most similar to the query stem. The morphological expansion (Freund and Willet, 1982; Henry, Willet and Wood, 1986) calculates a measure of string similarity (measured in trigrams) between a selected query stem and each of the stems in the dictionary file of the database. Then in both cases, the system displays the 20 most similar stems to the user who chooses the ones to be added in the query.

Query expansion examples based on a KS which is independent of the collection are found in CITE and in expert systems that use thesauri in one way or another. CITE automatically performs stemming on words input by the user, then selects medical subject heading (MeSH) terms and terms from the dictionary file by matching the user's words against them. Then it ranks the selected terms and displays them to the user for approval and feedback before commencing the search (Doszkocs, 1983).

All expert systems which have pre-search aid modules such as CANSEARCH, CONIT, etc., although they use KS independently of the collection and search results, help in the suggestion of terms. Some of these systems maintain a thesaurus as part of their structure but all provide some interesting means of the identification of terms.

EXPERT (Yip, 1981) helps the user to build a query formulation by suggesting splitting the topic into concepts and then suggesting terms to express each one of them. The system builds the query by developing these concepts and then it prints them out in a columnar manner with each column containing the terms that form one concept. Thus it has no thesaurus (or any form of semantic KS); its KS could be described as purely syntactic, relating to the structure of typical simple Boolean queries.

CANSEARCH (Pollitt, 1981 and 1987) uses a subset of the MeSH thesaurus, together with its own knowledge of how cancer queries tend to be structured, to help the user identify terms and incorporate them in the query. A system specifically designed to assist users in the selection of query terms has been developed by Shoval (1985). Its knowledge base consists of a thesaurus enhanced by additional information about terms, which is organized as a semantic network. The user inputs search terms and the system searches its knowledge base, selects terms that seem to be relevant, evaluates them and presents them to the user who chooses which are useful for the query.

Based on results of search

An alternative method for assisting users in the search process is by presenting them with information based on the results of the search. In general, this user–system interaction is as follows, depending on the various levels of sophistication and automation of the interaction. The system presents to the user a list of terms based on their occurrences in an identified set of documents. Then the user feeds back his or her choice of terms. The document set on which this analysis is based may either be simply a set retrieved in the usual fashion (and chosen or accepted by the user as a suitable *set* for this purpose), or it may consist of documents individually selected as relevant by the user.

The ZOOM (Martin, 1982; Ingwersen, 1984) feature on ESA/IRS is a helpful tool for online searching along these lines. It performs term frequency analysis on a number of records from the retrieved set(s) (on all or the first 50). The user is then presented with screen-displays which contain terms in a frequency ranked order. The searcher selects terms which then can use to expand the query. ESA/IRS offers also QUEST-QUORUM (D'Elia and Marchetti, undated) as a simple interface to its command driven system for inexperienced users, a similar service to DIALOG's KNOWLEDGE-INDEX or BRS's After-Dark, but which can also do a semi-automatic query expansion based on terms selected by the user from a ZOOM-like display.

There are also some other systems that have utilized the term frequency analysis function either in a ZOOM-like way or in some other form. Williams (1984) in his IT (Information Transfer system) uses the command EXPLORE to retrieve, organize and present index terms to the user, for addition to the search profile. Analogous is the command TERMS in the MUSCAT online catalogue (Porter and Galpin, 1988). It presents terms and UDC numbers extracted from the inspected and judged relevant documents and which are not included in the original query. With each term presented the user is asked whether or not to add it to the query. CITE (Doszkocs, 1983) utilizes the user feedback also in a similar manner. It automatically performs term frequency analysis on the records marked as relevant, and then it presents them in ranked order to the user for selection.

EUREKA, (Burket, Emrath and Kuch, 1979) is a full text retrieval system. It uses a user-specific thesaurus (there is not a system one) to help find synonyms for the query terms and which it consults throughout searching. Also, it uses a context specific vocabulary as a post-search aid. EUREKA presents either a histogram of term frequencies based on the retrieved documents, or word-lists of terms that are used in many documents or have high average frequencies. From these lists the user selects terms to refine the retrieved set.

Feedback based on the result of the search and with a dynamic form of search modification is given by Thomas (Oddy, 1977) which treats the document collection as a thesaurus and lets the user browse through it. The user starts with a simple natural language expression of interest, and the program responds by presenting a document representative (i.e., title, author(s), index terms) and asks for the user's judgements. At this point the user can make negative or positive judgements, on the document as a whole, and/or on any index term(s) or author(s) listed. Then Thomas iterates the search while keeping track of its actions by creating a model of the search in the form of a network.

Finally, there are a number of IR systems using AI techniques with the aim of improving their interaction style and effectiveness (Smith, 1987; Vickery and Brooks, 1987; Hawkins, 1988). Some of them use as their knowledge base a highly structured thesaurus, as described above. Also, some have incorporated knowledge of search heuristics for choosing alternative search strategies, and use browsing and other techniques to help the user choose terms and reformulate the query. Search heuristics drawn from the work of researchers such as Bates (1979a and 1979b) and Fidel (1985 and 1986) have been applied in IR-NLI (Guida and Tasso, 1983) and PLEXUS (Vickery *et al.*, 1986 and 1987). According to the user's responses, PLEXUS evaluates the search results then reformulates the query and resubmits it to the DBMS for processing. The result is again assessed and the process repeated if necessary.

It is clear that a reasonably full implementation of Bates-type tactics, for example, requires use of search result information. However, many of the expert systems are implemented as pre-online-search aids, and therefore depend on the use of other knowledge bases (as discussed earlier).

Concluding remarks

Somewhere between IR theory in the traditional sense, and human–computer interaction (HCI) theory applied to IR, lie the concerns of this chapter. What information should come back to the user from the system, and what information should the system request of the user, for effective

retrieval? It will be clear from this chapter that it is difficult to theorize about this question; the richer the situation is in feedback terms the more difficult it is.

Nevertheless, the whole area of interaction and feedback is clearly where the action is. IR researchers and system designers are rapidly becoming aware of all sorts of possibilities and all sorts of problems in this area. There is no consensus as to what the most important problems are, let alone where the best solutions might lie. There is, however, a plethora of approaches and ideas; also a plethora of opportunities to develop systems incorporating new ideas.

This last point is worth expanding upon. For a number of years, the major area of activity in real-world IR was the accessing of the big host systems, over slow telephone lines from dumb terminals. The host systems themselves offered searching software of a conventional kind, but on the whole did not go in for new ideas.

The present situation is different in a number of respects:

- There are many more local databases and packages for them. Some of the packages straddle the traditional text retrieval/DBMS divide, which makes for a cross-fertilization of ideas. The main benefit to text and bibliographic retrieval lies in the HCI domain.
- One major area for local database development is online library catalogues. Because these are designed for use by the general public, designers needs must pay attention to HCI and interaction aspects.
- With the traditional host systems, it is now feasible for a user to have a local front-end system, which may improve on the feedback facilities provided by the host.
- The front-end idea has made its way back to the host organizations, which now provide a variety of such systems on the networks. The heavy computing requirements of the raw database, which make for systems which are not responsive to the need for change, can now be separated from the interface aspects.

So the opportunities for more highly interactive systems are now excellent. The major problems lie in deciding which ideas to implement.

One reason for this difficulty lies in the difficulty of conducting adequate evaluation experiments on highly interactive systems. The more highly interactive the system, the more difficult experimentation is. Evaluation has not been discussed in this chapter, but good evaluation experiments on interactive systems are few and far between.

It is certainly not possible to conclude, on the basis of present knowledge, that the best form of feedback to design into a system is 'X'. Furthermore, some of the ideas discussed here can be (and have been) combined in one system. Nevertheless, there are some limitations to this process: certain combinations cannot be made, and in the end only a limited amount of information can be presented to the user or requested from the user at one time.

One can only suggest that, today, IR is a very open topic – perhaps (under the influence of new systems, new ideas and new opportunities) more open that at any time in the past.

References

BATES, M. (1979a) Information search tactics. *Journal of the American Society for Information Science*, **30**(4), 205–214

BATES, M. (1979b) Idea tactics. *Journal of the American Society for Information Science*, **30**(5), 280–289

BATES, M. (1986) Terminological assistance for the online searcher. In *Proceedings of the 2nd Conference on Computer Interfaces and Intermediaries for Information Retrieval* (Boston, 1986), C. E. Jacobson and S. A. Witges (compilers), Defence Technical Information Center, Alexandria, USA, pp. 285–293, (Report No DTIC/TR-86/5, NTIS: AD A174-000-0)

BATES, M. (1987) How to use information search tactics online. *Online*, **11**(3), 47–54

BELKIN, N. J. and VICKERY, A. (1985) *Interaction in Information Systems: a Review of Research from Document Retrieval to Knowledge-based systems*, British Library, London (LIR No 35)

BELKIN, N. J. and CROFT, W. B. (1987) Retrieval techniques. In *Annual Review of Information Science and Technology*, Vol. 22 (ed. Williams, M. E.), Elsevier, Amsterdam, pp. 109–145

BELKIN, N. J., SEEGER, T. and WERSIG, G. (1983) Distributed expert problem treatment as a model for information system analysis and design. *Journal of Information Science*, **5**, 153–167

BROOKS, H. M., DANIELS, P. J. and BELKIN, N. J. (1985) Using problem structures for driving human-computer dialogs. In *Proceedings of RIAO '85 Symposium* (Grenoble, 1985), IMAG, Grenoble, pp. 645–660

BROOKS, H. M., DANIELS, P. J. and BELKIN, N. J. (1986) Research on information interaction and intelligent provision mechanisms. *Journal of Information Science*, **12**, 37–44

BURKETT, T. G., EMRATH, P. and KUCK, D. J. (1979) The use of vocabulary files for online information retrieval. *Information Processing and Management*, **15**, 281–289

D'ELIA, S. and MARCHETTI, P. G. (undated) QUESTQUORUM: A new online search assistance tool from ESA-IRS. *Available online from ESA-IRS*

DILLON, M. and DESPER, J. (1980) The use of automatic relevance feedback in Boolean retrieval systems. *Journal of Documentation*, **36**(3), 197–208

DILLON, M., ULMSCHNEIDER, J. and DESPER, J. (1983) A prevalence formula for automatic relevance feedback in Boolean systems. *Information Processing and Management*, **19**(1), 27–36

DOSZKOCS, T. E. (1983) CITE NLM: Natural-language searching in an online catalog. *Information Technology and Libraries*, **2**(4), 364–380

DRAPER, S. W. and NORMAN, D. A. (1985) Software engineering for user interfaces. *IEEE Transactions on Software Engineering*, **SE-11**(3), 252–258

EICHMAN, T. L. (1978) The complex nature of opening reference questions. *RQ*, **17**(3), 212–222

FIDEL, R. (1985) Moves in online searching. *Online Review*, **9**(1), 61–74

FIDEL, R. (1986) Towards experts systems for the selection of search keys. *Journal of the American Society for Information Science*, **37**(1), 37–44

FREUND, G. E. and WILLETT, P. (1982) Online identification of word variants and arbitrary truncation searching using a string similarity measure. *Information Technology: Research and Development*, **1**(3), 177–187

GUIDA, G. and TASSO, C. (1983) An expert natural language interface to online databases. In *Proceedings of the Conference on Applied Natural Language Processing* (Santa Monica, 1983), Association for Computational Linguistics, USA, pp. 31–38

HARTER, S. P. and PETERS, A. R. (1985) Heuristics for online information retrieval: a typology and preliminary listing. *Online Review*, **9**(5), 407–424

HAWKINS, D. T. (1988) Applications of artificial intelligence (AI) and expert systems for online searching. *Online*, **12**(1), 31–43

HENRY, I. G., WILLETT, P. and WOOD, F. E. (1986) INSTRUCT: A teaching package for experimental methods in information retrieval. Part I. The users' view. *Program*, **20**, 245–263

HENRY, M., LEIGH, J., TEDD, L. and WILLIAMS, P. (1980) *Online Search: An Introduction*, Butterworths, London

INGWERSEN, P. (1984) A cognitive view of three selected online search facilities. *Online Review*, **8**(5), 465–492

KNAPP, S. D. (1978) The reference interview in the computer-based setting. *RQ*, **17**(4), 320–324

McGILL, M. J. and HUITFELDT, J. (1979) Experimental techniques of information retrieval. In *Annual Review of Information Science and Technology*, Vol 4 (ed. M. E. Williams), Knowledge Industry Publications, NY, USA, pp. 93–127

MARKEY, K. (1981) Levels of question formulation in negotiation of information need during the online presearch interview: A proposal model. *Information Processing and Management*, **17**(5), 215–225

MARKEY, K. and COCHRANE, P. A. (1981) *ONTAP: Online Training and Practice Manual for ERIC Database Searchers* (2nd edn), Syracuse University, Syracuse, ERIC Clearinghouse on Information Resources, USA

MARTIN, W. A. (1982) Helping the less experienced user. In *6th International Online Meeting* (London, 1982), Learned Information (Europe) Ltd, Oxford, pp. 67–76

MEADOW, C. T. and COCHRANE, P. A. (1981) *Basics of Online Searching*, John Wiley & Sons, NY, USA

NORMAN, D. A. (1984) Stages and levels in human-machine interaction. *International Journal of Man-Machine Studies*, **21**(4), 365–375

ODDY, R. N. (1977) Information retrieval through man-machine dialogue. *Journal of Documentation*, **33**(1), 1–14

PIETILAINEN, P. (1983) Local feedback and intelligent automatic query expansion. *Information Processing and Management*, **19**(1), 51–58

POLLITT, A. S. (1981) An expert system as an online search intermediary. In *5th International Online Information Meeting* (London, 1981), Learned Information Ltd, Oxford, pp. 25–32

POLLITT, A. S. (1987) CANSEARCH: An expert systems approach to document retrieval. *Information Processing and Management*, **23**(2), 119–138

PORTER, M. F. (1982) Implementing a probabilistic information retrieval system. *Information Technology: Research and Development*, **1**(2), 131–156

PORTER, M. F. and GALPIN, V. (1988) Relevance feedback in a public access catalogue for a research library: Muscat at the Scott Polar Research Institute. *Program*, **22**(1), 1–20

RADA, R. (1987) Knowledge-sparse and knowledge-rich learning in information retrieval. *Information Processing and Management*, **23**(3), 195–210

ROBERTSON, S. E. and BOVEY, J. D. (1983) *A Front-end for IR Experiments* Final Report to the British Library Research and Development Department, Project Number SI/G/569

ROBERTSON, S. E., BOVEY, J. D., THOMPSON, C. L. and MACASKILL, J. M. (1986) Weighting, ranking and relevance feedback in a front-end system. *Journal of Information Science*, **12**, 71–75

ROBERTSON, S. E. and SPARCK JONES, K. (1976) Relevance weighting of search terms. *Journal of the American Society of Information Science*, **27**(3), 129–146

SALTON, G., FOX, E. A. and VOORHEES, E. (1985) Advanced feedback methods in information retrieval. *Journal of the American Society for Information Science*, **36**(3), 200–210

SALTON, G. and McGILL, J. (1983) *Introduction to Modern Information Retrieval*, McGraw Hill Inc., New York

SHOVAL, P. (1985) Principles, procedures and rules in an expert system for information retrieval. *Information Processing and Management*, **21**(6), 475–487

SHNEIDERMAN, B. (1986) Designing menu selection systems. *Journal of the American Society for information science*, **37**(2), 57–70

SMEATON, A. F. and VAN RIJSBERGEN, C. J. (1983) The retrieval effects of query expansion on a feedback document retrieval system. *The Computer Journal*, **26**(3), 239–246

TAYLOR, R. (1968) Question-negotiation and information seeking in libraries. *College and Research Libraries*, **29**(3), 178–194

VAN RIJSBERGEN, C. J., HARPER, D. J. and PORTER, M. F. (1981) The selection of good search terms. *Information Processing and Management*, **17**(2), 77–91

VICKERY, A., BROOKS, H. M., ROBINSON, G. and VICKERY, B. C. (1986) *Expert System for Referral*, Final Report, University of London, Central Information Service

VICKERY, A., BROOKS, H. M., ROBINSON, B. and VICKERY, B. C. (1987) A reference and referral system using expert system techniques. *Journal of Documentation*, **4**, 198–203

WADE, S. J. and WILLETT, P. (1988) INSTRUCT: a teaching package for experimental methods in information retrieval. Part II. Browsing, clustering and query expansion. *Program*, **22**(1), 44–61

WILLIAMS, P. W. (1984) A model for an expert system for automated information retrieval. In *8th International Online Information Meeting*, (London, 1984), Learned Information, Oxford, pp. 139–149

WU, H. and SALTON, G. (1981) The estimation of term relevance weights using relevance feedback. *Journal of Documentation*, **37**(4), 194–214

YIP, M. K. (1981) An expert system for document retrieval. *MSC Thesis*, Department of Electrical Engineering and Computer Science, MIT, Cambridge, USA

11 Business information online

Jacqueline Cropley

Introduction

Just as in the 1970s the online industry developed because of the difficulties of accessing scientific information, coupled with the immediacy of the rewards, so the advantages of tracing widely scattered business information have been responsible for the burgeoning growth in the 1980s. This review looks at the provision of online business information from the UK user perspective. Although there are articles covering the US scene fairly comprehensively, (Sieck, 1984; Ojala and Bates, 1986) the picture for the UK and the rest of Europe is much more fragmented.

The history of business information online is relatively recent. It is a sector which has developed and transformed itself over the last eight years. Many significant players were not operational two or three years ago. Others have disappeared or changed substantially. The principal characteristic of business information is its currency. Most users have only intermittent interest in anything more than two years old. It would therefore seem futile to write about business information itself with a longer time frame. This review concentrates on activity since 1985, with the older material being used mainly to document current trends. Most of what is covered can be verified by product announcements. References in the main are restricted to commentary material which adds insight. Many major concerns are common to all online information retrieval, not just the business databases, and so have received wide airing. To retain balance, references in these areas are representative or notable rather than comprehensive.

Market development

The businessperson's need for reliable, timely information that can be readily turned to commercial advantage has been a marketing windfall for the database suppliers and hosts. A scarcely tapped market sector with identifiable needs for both broad brush and niche products has created a surge that has virtually spawned a whole new industry, albeit on the muted scale of the online sector in general. The wide availability of online business information is directly attributable to its commercial success. The top online information providers are primarily business information sources. The vision of their market penetration has brought more services on stream.