ELSEVIER

Interacting with Computers 10 (1998) 249–267

Interacting with Computers

Human–computer interaction in information retrieval: nature and manifestations of feedback

Amanda Spink*, Tefko Saracevic[1]

School of Library and Information Sciences, University of North Texas, PO Box 311068, Denton, TX 76203, USA

School of Communication, Information and Library Studies, Rutgers University, Rutgers, USA

Abstract

This study develops a theoretical framework for expressing the nature of feedback as a critical process in interactive information retrieval (IR). Feedback concepts from cybernetics and social sciences perspectives are used to develop a concept of informational feedback applicable to IR. Models from human–computer interaction and interactive IR are then adapted as a framework for studying the manifestations of feedback in IR. An informational feedback loop is defined as a unit of measure for IR feedback. Results are then presented from an empirical study of real-life interactions between users, professional mediators (information specialists) and an IR system 'computer'. Data are presented involving 885 feedback loops classified in five categories. In conclusion we present a connection between the theoretical framework and empirical observations and provide a number of pragmatic and research suggestions. © 1998 Elsevier Science B.V.

Keywords: Interactive information retrieval; Feedback; Cybernetics; Human–computer interaction; Relevance

1. Introduction

Information retrieval (IR) is a complex process that involves many activities: organization of texts, e.g. images, sounds or multimedia with cognitive content ('texts'); an intellectual representation of such texts, derived by humans directly or indirectly by algorithms; intellectual searching and retrieval by users; and the systems and techniques to accomplish this. IR interaction complexity is also derived from the direct involvement of human generators and users of texts in IR systems, bringing in cognitive, affective, social and situational (problem, task···) variables.

* Corresponding author. Tel.: 001 940 565 2187; fax: 001 940 565 3101; e-mail: spink@lis.admin.unt.edu

[1] E-mail: tefko@scils.rutgers.edu

PII S0953-5438(98)00009-5

IR is also an interactive process, involving feedback as a specific subset of human–computer interaction (HCI). In IR these interactive processes are also viewed as episodes in which a number of different 'things' are attempted and/or accomplished. Feedback is viewed as a specific type of interaction episode and an essential, critical element of IR. Clearly, the study of IR processes, such as indexing or searching algorithms, has not included consideration of feedback. However, many theoretically or experimentally elegant studies are removed from the real world where feedback is always present. As important and omnipresent as feedback is in IR, we still know relatively little about its nature, manifestations, and behavior, and do not have a usable theoretical framework for expressing the nature of IR feedback. This article extends work by Spink [1–8] and reports a pioneering study asking three questions: *What kind of a theoretical framework may be used for IR to depict the nature of IR feedback?* Here we examine the nature of feedback in cybernetics, and social sciences, HCI and IR, and propose an informational feedback perspective for IR and HCI. *What kinds of feedback manifestations are there in IR?* In this section we report the results from a large empirical study of feedback episodes during interactive IR based on a feedback unit of analysis. The third question involves the connection between theory and practice: *How can we reconcile our theoretical framework with the empirical results of our study?*

2. Theoretical framework: the nature of feedback in IR

2.1. Feedback perspectives

Discussions about feedback date back to antiquity. However, the modern concepts related to feedback have evolved into several perspectives and intellectual traditions, among them: engineering, economics, biology, mathematical models of biological and social systems, formal logic, and classical social science [9]. To these we add an informational perspective. Yet all these traditions and perspectives have a basic concept in common: *feedback involves a closed loop of casual influences, a loop of mutual or circular causality*. Feedback can be of different polarity—positive feedback reinforces change in the same direction, negative feedback in the opposite direction and homoeostatic feedback maintains an equilibrium. The loop is central to any and all perspectives of feedback. However, the interpretation of what is involved in the loop and the polarities differs significantly within different perspectives. In a number of perspectives 'information' is that which feeds around the loop, and goes back, eventually, to the beginning point. "Feedback [in cybernetics] is viewed as information transmitted in messages." ([9], p. 128). But the question of what kind of 'information' is left unanswered, i.e. information is treated in the feedback literature and thinking within cybernetics as a primitive concept.

2.2. Cybernetic perspective

The area of cybernetics is mostly associated with and influenced by the works of Norbert Wiener [10–12]. He coined the term and laid out the fundamental ideas that resulted in a large number of inquiries over the following decades. Wiener was concerned with feedback as a mechanism of control. Thus, feedback in cybernetics is defined in terms of an input to a process, the process itself (a 'black box'), an output from the process and an informational link between the input and output to control the input, process and output, e.g., a thermostat. However, the loops are limited to negative polarity, to counteract deviations. The system's stability and the conditions producing instability are of major concern. A list of elements, the imagery, and the formalism in this feedback concept developed by Wiener and many others is of course much more elaborate, including among others: receptors, a control apparatus, effectors, stimulus and response, steady-state process, signals, and even entropy.

The strength of this concept of feedback is in its wide and successful theoretical and pragmatic application in engineering, particularly in servomechanisms, automata, programming, and many other physical processes. However, the ambition to extend these concepts to human activities did not succeed that well. The traditional cybernetic concepts have also been used in depicting feedback in the traditional IR model (discussed later), with the same restrictive results, and even more so on using 'information' to mean signals devoid of any cognitive and situational attributes. Wiener's cybernetics needed an extension.

2.3. Social science perspective

Many researchers in social sciences eagerly embraced the concepts from cybernetic feedback, primarily to counter the inadequacy of depicting social phenomena through cause–effect approaches, exemplified by the ubiquitous dependent–independent variable studies [13]. Such approaches do not depict the reality of the many intervening influences and variables found in social phenomena and processes. In addition, the concept of feedback from cybernetics needed substantial modification to account for the interdependence of a number of mutually intervening variables [9]. As studies in cybernetics tend to emphasize negative feedback, Maruyama [14,15] pioneered the expanded notion of feedback to include positive, reinforcing, or deviation amplifying feedback. These are information-feedback processes that have an important role in mutual casual systems. the concepts were applied to biological and social systems, and Maruyama called this new approach "the second cybernetics". The loop concept was fully retained, and extended with more elements and variables, or even several loops in a complex system, with all polarities possible to economic, political and social processes and loop diagrams [9]. This represents the strength of the second cybernetic approach. The weaknesses are less obvious. First, there is a persistent use of machine behavior and models as analogy to human behavior. But human behavior is very poorly, if at all, depicted by these analogies. Second, is the tendency to use 'input' either in a sensory sense or as 'information' in a most restrictive sense, devoid of any cognitive and situational references, interpretations, and processes. A third weakness is that the emphasis on the loops between intervening variables de-emphasizes the study of the variables themselves. Moreover, the feedback loops remain mostly undefined, raising questions about the necessity for the concept of feedback instead of simply considering these as complex systems with intervening variables. Thus, still another informational perspective is needed to address some of the weaknesses for IR.

2.4. Information feedback perspective

A major objection to cybernetic and social sciences perspective of feedback for IR is the cavalier treatment of the concept of information, mostly used as a primitive undefined term. Yet information is a complex phenomenon that has a number of manifestations, or conceptions, each imposing its own restrictions and conditions on that which follows the use of the given conception. Then, what is 'information'? On a continuum from the narrowest to broadest, information can be depicted in three distinct senses.

On the narrowest end of the continuum information is considered exclusively as a property of a signal or message. This is the most restrictive interpretation. No cognitive or related interpretation is necessary—information is there because the signal or message is there, e.g., Shannon's information theory. In the middle end of the continuum the conception of information is less restrictive and incorporates a cognitive interpretation. It results from the interaction of two cognitive structures, a 'mind' and a 'text'. Information is that which affects or changes the state of a mind. It depends on the conceptualization and interpretation of a human being. At the broadest end of the continuum information is related not only to cognition and cognitive structures, but also to intentionality, affection, and motivation, which connect to an expansive array of contexts—social, situational, and cultural. In IR they connect to a given reason for using the IR system—a task and problem-at-hand—and a users's cognitive structure and affective intentions. Information in this broadest conception involves: (i) a message (text), AND (ii) a cognitive interpretation, AND (iii) a contextual consideration.

So far, the concept of feedback in IR has been largely restricted to the second or cognitive conception of feedback, by incorporating a relevance feedback as originally introduced by Rocchio [16], in order to improve performance of IR systems. This feedback is patterned after the cybernetic perspective of feedback, with no attention to the nature of relevance assessments by human users, but with the advantage that it can be treated as an algorithm, e.g., relevance feedback algorithms have been shown to improve IR performance to some degree [8,17].

However, we suggest that the only appropriate conception of feedback for IR is the third or broadest conception of information, i.e., information that includes not only the textual but also the cognitive and contextual senses. This has important implications for conceptualizing IR feedback. Feedback involves an input and output, a query and a retrieved text, but in addition it involves an *interpreter*, for assessing the text, cognitively, affectively, and situationally, connecting it to a whole array of interactions and mutual causalities, with a possibility of all feedback polarities. The interpreter is a human being, a user or surrogate, and in some distant future it many be an intelligent machine agent, acting and interpreting on behalf of the human user. We can also envision that on the computer side there may be an interpreter, connecting the inputs or texts to an array of interpretive programs. The importance of studying such phenomena is to learn from humans what to put in the machine.

Therefore, as an IR unit of measurement is proposed: *The feedback loop is an interaction which consists of: (i) a query, (ii) a process to obtain a text as a response to a query, (iii) the text of the response, (iv) an interpretation by an interpreter on the appropriateness of the text to whatever contextual (cognitive, affective or situational) variables, and then* *(v) an action to modify in some way the query or the retrieval process.* The important point is: who can initiate the feedback loop? We suggest that only a human interpreter, a user or a surrogate, can be the initiator. In other words, *the loop is user initiated* and not mutual casual like circular loops as depicted in the social feedback perspective. It is user causal, as the act of interpretation begins with the user or the entry of a query. A cognitive and contextual view of information dictates a user causal view. If and when intelligent interpreters may be developed for the computer, we may reconsider this, to provide for a possibility of a computer initiated loop. In the broader context of HCI feedback is not linear, but dynamic and heavily contextual with variables that may be modified, producing still other feedback loops, and different actions and modifications.

3. Human–computer interaction models

HCI research focuses on interaction as "an exchange of information between participating agents through sets of information channels (interfaces) ··· where each has the purpose of using the exchange to change the state of itself or one or more others" ([18], pp. 173, 181). Key concepts are *participant, interaction, purpose, and interface.* The approach provides a model to investigate the type, working and role of participants and interfaces, the nature of exchanges and associated purposes, and the changes, if any, in the states of each participant. Overall, some HCI models include feedback, others do not—it is implied. The models that include feedback have focused primarily on the duality of feedback as a linear transmission: either an input from a user to a computer or an output from a computer to a user. Feedback is not regarded as a complex loop, but as information sent back to a user by a computer resulting from user input to the system [19–21]. A non-loop model of feedback has been incorporated in many HCI models, such as those of Brandes and Wilenski [22], Norman [23] and Resnick [24]. In general, information gained from this type of feedback allows a user to evaluate the systems responses. In a sensory sense, which is obviously very limited but still important, feedback in HCI is considered as visual, tactile, or auditory—a distinction that so far has had limited theoretical or experimental application. However, the area of visualization has become an important area for the consideration of interface design, even with little research into visual feedback.

Interestingly, HCI research in education has produced the most extensive experimental literature on feedback [25]. Here the concept of feedback is more elaborate, including a classification of the types of feedback in HCI. Dempsey and Sales [25] regard feedback as the many procedural messages from system to user: (1) *motivational*, as measured as the rate of accuracy of user responses; (2) *rewarding*, as measured by the number of correct user responses, or (3) *informational*, as measured as user error responses. The classification has not found a larger application beyond the educational testing domain. In general, feedback is still a somewhat ambiguous and even primitive concept in HCI. While recognized, little attention has been paid to cognitive and contextual variables, and mutually casual, complex feedback loops. In HCI, feedback is simply oversimplified. Storrs ([18], p. 174) also suggests that: "HCI is not lacking in theory. It is simply that the theories that exist are vague and weak." A similar statement can be made about related theories in IR.

4. Information retrieval models

The popular traditional IR model, used in most algorithmic approaches, represents IR as a two-pronged set (system and user) of elements and processes converging on matching [26]. The system prong involves texts that are represented in some way and then organized into a file, ready for the matching process. The user prong starts with an information problem or need, represented by a question, which is transformed into a query acceptable to a system, and then a matching process occurs. A feedback function is included, to allow for a modification, and usually it involves modification of the query based on some relevance assessment. In the context of the SMART project led by Gerald Salton, Rocchio [16] proposed to take advantage of an interactive process he called relevance feedback based on a cybernetic perspective on feedback. Feedback was regarded as an automatic process for the modification of a query, based on relevance judgments of the user (or surrogate) on the texts previously retrieved by the system. Texts judged relevant by a user are used to automatically adjust the initial query. Search terms from relevant texts are given preference in subsequent searches, to improve the effectiveness of subsequent retrievals [8,27]. This showed the strength of the traditional model when applied to a limited concept of feedback, namely relevance feedback as narrowly defined.

However, a greater failing of the traditional model is that it does not consider interaction at all or the interactive reality of modern IR. This weakness has recently led to the development of interactive IR models by Ingwersen [28] and Belin et al. [29]. A stratified model of IR interaction was also developed by Saracevic [30] and is applied here as a framework for the elaboration of what is meant by interpretation in the IR feedback loop.

The stratified model starts with the assumption that users interact with IR systems in order to use information, and that the use of information is connected with cognition and then situational application. IR interaction is a dialogue between the participants—user and 'computer' through an interface. Each of them has different levels or strata (Fig. 1) with the main purpose to affect the cognitive state of the user for effective use in connection with an application at hand. The dialogue can be reiterative with various types of feedback, and can exhibit a number of patterns, all of which are topics for study. On the user side we can model IR interaction to appear at least in four strata: surface, cognitive, affective, and situational. On the 'computer' side we can model interaction also in levels: surface, engineering, processing and content.

In this sense, feedback is also represented by an interplay between different levels, giving rise to different types of feedback. In other words, in IR we have a dynamic, interdependent system of relevances (note plural) within types of feedback as investigated in the study detailed below.

5. Empirical study: manifestations of feedback

5.1. Data corpus

Data used in this analysis were collected during a larger study of mediated online searching [31–33]. Forty self-selected academic users (faculty and doctoral students) at Rutgers University with real information problems provided one question each for online

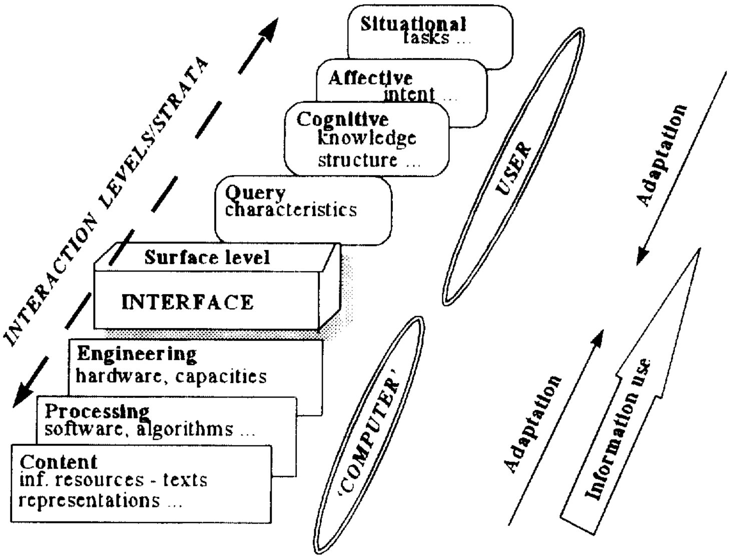

Fig. 1. Elements in the stratified model of IR interaction

searching on DIALOG. Four professional search intermediaries were involved, each doing 10 questions. The intermediaries interviewed the users prior and during online searching and conducted the search. The 40 questions included topics in medicine, social sciences, physical sciences and the humanities, and searching as many DIALOG databases as necessary for a given question. Users filled out prior to the search a standardized form for the question statement, listing the title and description of their question.

The interaction between the users and intermediaries was videotaped during a pre-online search interview and during the actual online search. The transaction logs of the searches were recorded. The discourse between users and intermediaries was transcribed from videos. Later, utterances in the discourse from the transcripts and entries (commands, responses) from the transaction logs were synchronized as to the time of appearance in the total interaction, and provided the basic data for determining feedback loops (see Table 1).

Each user examined each retrieved item and judged as either relevant (R), partially relevant (PR), or not relevant (NR). The users' relevance judgments were collapsed from three to two classes: items judged relevant (R) and partially relevant (PR) were collapsed into one category (R + PR), from now on simply called 'relevant'. We provided users with an instruction sheet defining the meaning of relevance as topical relevance; however, we cannot tell whether users used their own and/or other criteria for judgment. These relevance judgments were used to determine the effectiveness of a number of the variables studied, and were an important part of the identification of certain feedback types.

5.2. Method for feedback analysis

The IR unit of analysis, defined in Section 2, was the base for the analysis. It is a closed loop that encompasses: a user input or query; a process to obtain an output; a system

Table 1

Summary of the data corpus

Questions:	
Number of questions (one per user)	40
Hours of videotapes	46.05 h
Mean time per question	69.08 min
Mean time: presearch interview	13.04 min
Mean time per online search	56.04 min
Search intermediaries:	
No. of search intermediaries	4
Mean experience per intermediary	8.5 y
Items retrieved:	
Total no. of items retrieved	6225
No. of relevant (R) and partially relevant (PR) items retrieved	3565
No. of not relevant items retrieved	2660
Mean precision per question	57%
Min/Max of total items retrieved	13/427
Min/Max of R + PR items retrieved	1/348
Min/Max of not relevant items retrieved	0/180
Databases searched:	
No. of different databases searched	46
Mean searched per question	3
Utterances (from transcripts):	
By users	9714
By intermediary	11318
Total utterances	21032
Commands (from transaction logs):	
No. of commands	1675
Mean commands per question	41.7

output; an interpretation by user or intermediary; and a subsequent action or input to the system. As mentioned, the process is contextual, and dynamic, not linear. Based on the IR feedback unit of analysis, a micro-analysis of the user–intermediary dialogue was conducted concurrently with the online search logs to identify the episodes and types of IR feedback during the 40 mediated online searches. This analysis used an inductive grounded-theory methodology [34] or theory-generating methodology rather than a theory-testing orientation. The resulting categories of interactive feedback were grounded in the data. Transcribed user–search intermediary utterances, videotapes and related online search logs were examined for evidence of a user input, IR system output, user or search intermediary interpretation (evaluation) and judgment related to the system output, and subsequent input or query reformulation. Each user or search intermediary utterance related to an aspect of the system's output was transcribed onto the section of the online search log listing the IR system's output.

A taxonomy of five different types of IR feedback was developed and refined during the iterative coding and categorization process. IR feedback episodes were identified and then classified into one of the five different types that were mutually exclusive and exhaustive. A sample of four (10%) online searches were checked by a second coder for reliability, and the coding results were compared and found to be consistent for all feedback types.

6. Results

Five feedback types were identified:

Content relevance feedback (CRF): user query followed by an IR system output of retrieved items then judged by the user for relevance followed by a query or reformulation.
Term relevance feedback (TRF): user query followed by an IR system output of retrieved items and user selection of a new search term(s) from the retrieved output used in a subsequent query.
Magnitude feedback (MF): user query followed by a judgment based on the size of the output from a query that effects the next query.
Tactical review feedback (TCR): user input followed by a strategy related judgment to display the search strategy history influencing the subsequent query.
Term review feedback (TMR): user input followed by a strategy related judgment to display terms in the inverted file influencing the subsequent query.

A total of 885 feedback loops or episodes were identified during the 40 mediated online searches (mean of 22 feedback loops per search, with a range of six to 65 in a positively skewed distribution with a few outlier scores at 60 and 65). Table 2 presents the distribution of feedback loops within the five categories and subdivided as initiated by users or intermediaries. A description of each type of interactive feedback loop follows, together with an example.

6.1. Content relevance feedback (CRF)

During the 40 online searches, users and search intermediaries made interactive relevance judgments regarding the retrieved items displayed within a total of 354 content relevance feedback episodes. This represented 40% of the total interactive feedback episodes (mean of nine content relevance feedbacks per search). A content relevance interactive feedback consisted of a query followed by one or more relevance judgments, before a modified or reformulated query or another command is entered. Both users' and search intermediaries made relevance judgments during the display of retrieved items (Table 2). Interactive content relevance feedback loops were also found to be either positive and negative.

In IR, the concepts of positive and negative feedback are not used in the cybernetic sense. In cybernetics, negative feedback means deviation-counteracting or correcting feedback. In IR, negative feedback is linked to users' judgments of retrieved items being not relevant. Positive feedback in IR is related to users' judgments of retrieved items being relevant.

Table 2
Summary of number and types of IR feedback

Types of IR feedback	Searches (number)	IR feedback episodes		Totals		
		Number	%	%	Mean	Range
Content relevance feedback (CRF)						
Users	40	252	71			
Intermediaries	24	102	29			
Sub-total		354	100	40	8.8	2–34
Term relevance feedback (TRF)						
Users	15	25	37			
Intermediaries	17	42	63			
Sub-total		67	100	8	3	1–8
Magnitude feedback (MF)						
Users	28	74	19			
Intermediaries	40	322	1			
Sub-total		396	100	45	9.9	3–27
Tactical review feedback (TCR)						
User	0	0	0			
Intermediaries	22	56	100			
Sub-total		56	100	6	1.4	1–6
Term review feedback (TMR)						
User	0	0	0			
Intermediaries	5	12	100			
Sub-total		12	100	1	0.3	1–4
Total		885		100	22	6–65

6.1.1. Positive content relevance feedback

Here is an example of a positive content relevance feedback during question number 16: *Intense Culture of Macrobrachium Rosenbergii: Technical Feasibility*. Search statement number 3 retrieved 91 items. The first five items were displayed in title format only and the user made a positive content relevance judgment.

User: Second yeah looks good. The one with infection is the one I wanted. The development of agriculture might be, because it might be something. This is pretty damn close.

Intermediary: Oh, okay. We can print all 90 abstracts.

The user judged the displayed output to be relevant and all the items were then printed before the entry of the next query.

6.1.2. Negative content relevance feedback

Below is an example of a negative content relevance feedback judgment during question number 6: *Cardiovascular Effects of Methoxamine, DTA Occlusion and PEPs*. Search statement number 12 retrieved 189 items and the first six items were displayed. The user made a negative content relevance feedback judgment.

User: No, nothing really relevant there. They all seem pretty broad still.

This negative relevance judgment was followed by a query reformulation.

6.2. Term relevance feedback (TRF)

Term relevance feedback represents a specific type of interactive feedback when a user or search intermediary identified a term (or terms) within the retrieved items subsequently used to modify the search strategy. There were 67 term relevance feedback episodes identified during the 40 online searches, representing 8% of the total number of feedback episodes. Term relevance feedback episodes were separated from content relevance feedback to allow more detailed analysis of the effectiveness of search terms selected during relevance feedback. The identification and use of new search terms from the retrieved IR system output was not a frequent occurrence. However, it represents a particular type of content relevance judgment (as revealed in additional analysis not reported here) that often led to the retrieval of relevant items [4,5,35].

Here is an example of term relevance feedback during question number 8: *An Analysis of Student Attitudes Toward Writing as Active Knowledge*. During the search statement number 27 retrieved 28 items. All 28 items were then displayed in full format, including abstracts and descriptors, and the user identified the term school role that appeared in the descriptor field of two items, during a later relevance feedback episode.

User: Right. Exactly, uh hum. Let's see we have school role, did you see that one.

The term 'school role' was identified by the user and subsequently used in search statement number 28. Term relevance feedback occurred during 60% (22) of the 40 online searches when new search terms were selected by the user or search intermediary (mean = 3 and range = 1–8 search terms per search). Users selected 25 (37%) and the search intermediaries selected 42 (63%) of the term relevance feedback terms.

6.3. Magnitude feedback (MF)

Users and search intermediaries were found to frequently make magnitude judgments on the number of IR system postings within magnitude feedback. To re-emphasize, this is without ever seeing the actual text of a retrieved item, just seeing the number of items retrieved by a given search statement. Magnitude judgments were either to increase (positive magnitude feedback), decrease (negative judgment) or accept (non-modification judgment) the size of the output. Postings output magnitude was frequently judged during the mediated online searches, with a total of 396 magnitude feedback episodes, a mean of 9.9 magnitude feedback episodes per search, and a range of three to 27 per search. Magnitude feedback represented 45% of the 885 total feedback loops and occurred during each of the 40 online searches. A total of 322 (81%) magnitude judgments were made by search intermediaries and 74 (19%) by users. Search intermediaries seemed more attuned to the problems of large or small numbers of postings, making magnitude judgments in all 40 searches, whereas users made magnitude judgments in only 27 of the 40 database searches.

6.3.1. Positive magnitude feedback

Here is an example of positive magnitude feedback during search number 30: *Separation and Purification of Mabs Using Two-Phase Aqueous Systems*. Search statement 14 retrieves two items. The search intermediary then suggests increasing the small number of postings retrieved by searching another database.

Intermediary: Okay, let's try, now we're not coming up with too much ··· Let's try Medline and do the same thing.

Positive magnitude feedback reflected a concern with the small size of postings output and a desire to increase the number of postings before displaying retrieved items for relevance judgments.

6.3.2. Negative magnitude feedback

An example of negative magnitude feedback occurred during question number 3: *Typical Activity Segments of Elementary/Secondary Classrooms*. Search statement 27 retrieved 2028 items. The search intermediary then suggested restricting the search and decreasing the size of the output due to the large size of the postings output. The user suggests combing the search terms classroom communication and recitation.

Intermediary: Let's try that first ··· classroom communication let's see how big a term it is two thousand. Okay, let's try restricting.

User: Classroom communication and recitation.

The combination of classroom communication and recitation reduced the output to 10 items. Negative magnitude feedback reflects a concern with the large size of the postings and a desire to reduce the number of postings before displaying retrieved items for relevance judgments.

6.3.3. Non-modification magnitude feedback

Non-modification magnitude feedback occurred when the size of the postings' output is accepted without modification. An example appeared during question number 6: *Cardiovascular Effects of Methoxamine, DTA Occlusion and PEPs*. Search statement number retrieves 189 items from the previous 1386 items. The search intermediary then makes a non-modification magnitude feedback judgment and displays the first six titles of the retrieved items before printing the entire 189 items.

Intermediary: That brought it down a lot, it brought it down to a hundred and eighty-nine ··· Yeah, and this is from sixty-six to the present. Let me show you a few titles from the one hundred and eighty-nine.

6.4. Tactical review feedback (TCF)

A total of 68 (7%) feedback loops were not related specifically to the relevance or magnitude of the output, but the strategy implications of the output. Two types of strategy feedback were identified: tactical review feedback and terminological review feedback. The occurrence of strategy feedback extends the user-centered IR feedback model to include relevance, magnitude and strategy judgments within interactive feedback. Tactical review feedback occurred when a search intermediary made a strategy-related judgment by using the display sets command (DS) to display the history of the search strategy. This tactic was used often to locate their position in the online search process. There were 56 tactical review feedback episodes, representing 6% of the total feedback episodes, with a mean of 1.4 per search.

An example of tactical review feedback occurred during search number 10: *Children's Exposure to Lawn Care Pesticides*. After the display of 20 retrieved items during a long and complicated online search, the search intermediary requested a review of the online search history to identify the previous search terms used.

Intermediary: Now, did we do the exposure, residue and persistence while we were in Index medicus?

User: I don't remember.

Intermediary: Okay ··· I think I'll (see) what we have done. It won't search again it (is) simply going to display what we've done.

User: Three and four is not exposure or residue. Three and four is lawn and pesticide. Exposure and persistence is 12.

Although tactical review feedback did not occur frequently during the 40 online searches, it did involve a purposeful judgment on the search output.

6.5. Terminology review feedback (TMR)

Terminology review feedback occurred when a user or search intermediary made a strategy-related judgment by requesting the display of terms in the inverted file. There were 12 terminology review feedback loops, representing 1% of total feedback loops. Below is an example of terminological review feedback during question number 8: *An Analysis of Student Attitudes Toward Writing as Active Knowledge*. During this search, following the display of two retrieved items, the search intermediary initiated an 'expand' command for the term Douglas Barnes, an author mentioned by the user:

Intermediary: Also by expanding what I'm doing (the system) will tell me if there (is) any other form of Douglas F or Douglas (in the inverted file)

Terminological review feedback did not occur frequently. However, TMR formed a separate feedback type not related to either the magnitude or relevance of the output, but to the strategy of the search.

7. Discussion

7.1. Content relevance feedback

Results of this study confirm the work of Saracevic et al. [31], who suggested two types of interactive feedback during IR: relevance and magnitude. Overall, the selection of search tactics during the online interactions was affected by a concern with both the retrieval of relevant material and reducing the size of the output. Clearly, relevance is an integral part of, and is enabled by, the feedback process. Interactive relevance feedback was found to be a major element within the search process. As mentioned, we divided relevance feedback into two categories. Content relevance feedback involved a judgment of retrieved texts and then a subsequent action, such as modification of the strategy, without directly using search terms from those texts. Term relevance feedback is discussed next. Previous research has paid limited attention to interactive relevance judgments [36]. Results of the current study show that both content and term relevance judgments of retrieved texts occur frequently during the search process. These feedback loops play a critical role in determining the final set of retrieved items. As expected, content relevance judgments were made primarily by the users, over two-thirds of such judgments were made by users and one-third by intermediaries. However, even the one-third contribution by intermediaries shows the great influence of a mediated process.

Most significantly, the content relevance feedback is not only positive, as widely assumed in relevance feedback techniques, but also negative. Thus, possibility for both types, positive *and* negative relevance feedback, should be incorporated in the design of relevance feedback algorithms and approaches. Content relevance feedback and the term relevance feedback relate to the user (or even intermediary) cognitive level. We have little evidence that either invoked the situational level.

7.1.1. Term relevance feedback

As mentioned, term relevance feedback is a subset of relevance feedback which concentrated *only* on gaining terms from texts judged relevant by users. It is that type of feedback that is used in relevance feedback techniques and algorithms. Surprisingly, only one-twelfth (8%) of all the feedback loops are of this type. This was the most surprising finding of the study. Feedback which involves the extraction of terms from retrieved texts judged relevance was used very little! If this is a case that can be verified through a number of other studies, it removes justification of putting hope that relevance feedback will significantly improve retrieval performance. In real life it seems to be used rarely. But when used, it was the intermediaries, not users, that identified and incorporating such terms. About two-thirds of the term relevance feedback loops were initiated by intermediaries, and one-third by users. This again illustrates the powerful role of intermediaries or true intelligent agents that match this performance.

7.2. Magnitude feedback

It was magnitude, rather than relevance, that played the largest role in respect to the number of feedback loops. The users and search intermediaries in this study were frequently confronted with magnitude judgments due to number of postings that were to large or to small. An important determinant of the online searching process was the development of strategies to reduce or increase the number of postings. Wildemuth et al. [37] found that end-users made twice as many moves (defined above) to limit set sizes as expand set sizes; Wood et al. [38] also found that end-users primarily used postings data to narrow the search results.

Further analysis of the search cycles indicates the largest number of cycles (defined above) involved a reduction in the number of postings [4]. On many occasions search statements were entered by search intermediaries reducing the number of postings, but with no accompanying discussion by the user and search intermediary. At a certain "futility point" [39], users decide to display a smaller set of items. Saracevic et al. [31] suggested that in mediated database searching, 150 items should be the threshold number of postings leading to the display of retrieved items. The results of the current study show a mean of 123 items preceded the first type command. This finding also differs from other studies investigating postings' information [38], reporting lower thresholds for end-user searches. This was possibly due to both the mediated nature of the online searches conducted for faculty and graduate students and the often complex subjects searched.

Intermediaries were more aware and concerned with the magnitude of the output and users with the relevance of the output: intermediaries initiated four-fifths of the magnitude feedback loops and users only one-fifth. This is consistent with the search intermediary's experience with searching large databases. A greater proportion of relevance feedback episodes by users are consistent with the user's role as the domain knowledge expert during the online searching process. This is not to say that during mediated searching users were not concerned with the magnitude of the output. However, users were less familiar with the potential for large and small outputs and the necessity to be concerned with magnitude issues and were not entirely unconcerned with the relevance of the retrieved documents. Search intermediaries were also more likely to prompt the user to determine the relevance of the retrieved documents than directly examine the retrieved documents for relevance. Such mediation should be a very important part of search engines and intelligent agents.

7.3. Strategy feedback

We included two strategy feedback types: tactical review feedback and term review feedback. The first type accounted for 6% of feedback loops and the second for only 1%. Both strategy feedbacks represented a minor element within the search process. However, they cannot be ignored and neglected. The identification of strategy feedback extends the account of feedback types in IR interaction beyond relevance and magnitude feedback, and the specific focus on query reformulation. Potentially strategy feedbacks could have a large influence, particularly when browsing; navigational and browsing capabilities of IR are extended. Clearly, further research is required to explore the relationship between navigation, browsing and visualization on the one hand, and strategy feedback on the other, as related concepts representing non-query focused activities during interactive search processes.

7.4. Feedback and the search process

In this section we propose a model of the relationship between feedback and the elements in the IR search process. As illustrated in Fig. 2, an interactive IR search process may consist of one or more search strategies or optimizing plans for a search, e.g., pearl growing, building block, etc. [40].

Each search strategy may consist of one or more cycles (one or more search commands ending in the display of retrieved items [41]). Each cycle may consist of one or more interactive feedback loops (human user or intermediary input—IR system output—human interpretation and judgment—human input). An input may also represent a move within the search strategy [42] and may be regarded as a search tactic to further the search [43].

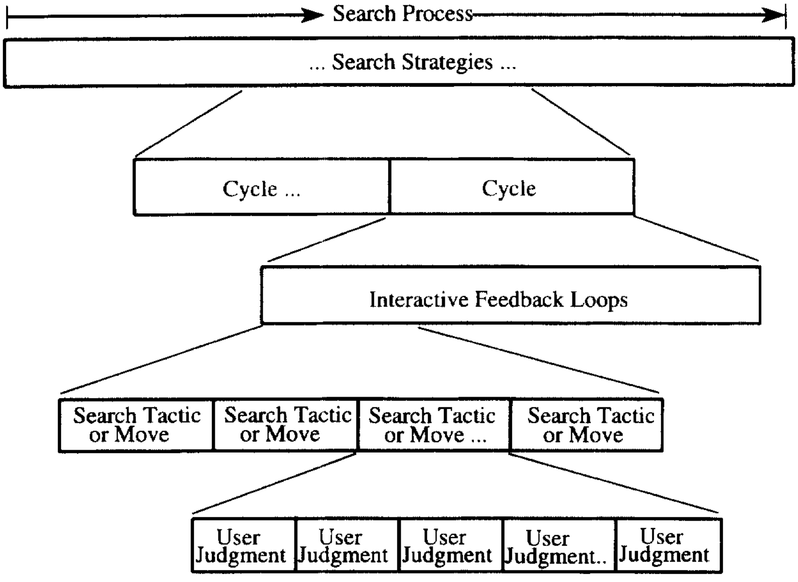

Fig. 2. Elements of the interactive search process

Each move consists of a user input or query requesting a system's output. An interactive search process may consist of a series of search strategies made up of one or more cycles, and one or more interactive feedback loops, of the type described above, within each cycle. An IR feedback may include one or more moves or search tactics, and user interpretations or judgments of the systems output. In order words, IR feedback, which includes search strategies, facilitates communication between user and IR system, and is related to the cognitive, affective, and/or situational state (levels) of the user. Thus, decomposition of the process, illustration of types of cycles and moves associated with types of feedback may further our understanding of the whole process.

8. Conclusions

We believe that this is the first study about the multitude of types of feedback that exist in interactive IR, that also includes the connection between feedback and IR interaction, in a broader framework. We suggest an informational feedback perspective that includes mutual causality; however, it has an interpreter, who treats information in its broadest sense, incorporating a text, a cognitive structure, and a context. Context involves an expanding set of elements from task and problem-at-hand to organization and even culture. Thus, an informational feedback loop for IR is defined as an interaction that involves a query, a process to obtain a text from a system, the text(s), an interpretation by an interpreter, and a subsequent action to affect the output or process. Such a loop, as yet, can be initiated only by a human, user or surrogate. Machines are not at the present time capable of being the initiator. But observations of what is going on in these processes where humans are involved can serve as powerful criteria for the design of intelligent agents and interfaces.

We also borrowed general definitions and concepts from HCI, primarily the identification of key components: participants, interaction, purpose, change, and interface. Interaction is that where the exchange of information has a purpose to change the state of one or more participants. We also suggest a stratified IR model that defines user and 'computer' strata, to encompass the feedback process at least four levels or strata (surface, cognitive, affective, and situational), and on the 'computer' side there are surface, engineering, processing, and content levels. Interaction is an interplay between these levels. IR feedback is an interplay between different levels giving rise to different types of feedback. The theoretical framework for explication of feedback in IR provides the basis for empirical study of informational feedback.

Five types of feedback loops were identified. Magnitude feedback loops were used most—showing the huge concern of users and intermediaries with sheer size, above and beyond the content. Relevance feedback was used less than magnitude feedback. Another surprise was the relatively low occurrence of term relevance feedback, where modifications in the query were made directly from examination of retrieved texts assessed as relevant by users. Such relevance feedback, which is a major effort in algorithmic IR research, played, at least in this study, a very minor role compared to all the other types of feedback. Clearly, all types of feedback are not created equal, and relevance feedback involving terms imported from relevant texts is not used that much at all. It seems that the users are saying, controlling how much is the largest preoccupation. This is a most interesting research question.

We also call for an enlargement of the IR feedback concept beyond relevance feedback so commonly taken as the only feedback in IR and propose the following challenges: (1) adaption of general feedback concepts and models to requirements of and situations in IR; (2) connection of IR models to HCI models; (3) integration of the traditional and interactive models of IR; (4) development of a theoretical framework for IR interaction including feedback, and to connect interactive feedback models to general feedback models; (5) integration of the traditional and interactive concepts of IR feedback; and (6) modeling the IR feedback loop, including the relationship between IR feedback episodes and search performance and tasks, and the dynamic and situational aspects of the IR search process. We believe that the results also give clues and should be considered in the design of IR systems and interfaces.

Acknowledgements

The data used in this study were from a large study funded by grants from the Library Research and Demonstration Program Grant, US Department of Education (ref. no. R039A80026) and DIALOG Information Services. We also thank the users and intermediaries, Professors Nicholas Belkin and Hartmut Mokros from Rutgers University, and the reviewers.

References

[1] A. Spink, Feedback in information retrieval. Unpublished doctoral dissertation, School of Communication, Information and Library Studies, Rutgers University, 1993.

[2] A. Spink, Interaction with information retrieval systems: reflections on feedback, Proceedings of the 56th Annual Meeting of the American Society for Information Science 30 (1993) 115–121.
[3] A. Spink, Magnitude feedback during mediated online searching, Proceedings of the 22nd Meeting of the Canadian Association for Information Science 22 (1994) 264–286.
[4] A. Spink, Term relevance feedback and mediated database searching: implications for information retrieval practice and design, Information Processing and Management 31 (2) (1994) 161–172.
[5] A. Spink, Term relevance feedback and query expansion: relation to design, in: Proceedings of the 17th Annual ACM Conference on Research and Development in Information Retrieval, 1995, pp. 75–81.
[6] A. Spink, Information science: a third feedback framework, Journal of the American Society for Information Science 48 (8) (1997) 741–760.
[7] A. Spink, A study of interactive feedback during mediated information retrieval, Journal of the American Society for Information Science 48 (5) (1997) 382–394.
[8] A. Spink, R.M. Losee, Feedback in information retrieval, Annual Review of Information Science and Technology 31 (1996) 33–78.
[9] G.P. Richardson, Feedback Thought in Social Science and Systems Theory, University of Pennsylvania Press, Philadelphia, PA, 1991.
[10] N. Wiener, Cybernetics: or Control and Communication in the Animal and the Machine, The MIT Press, Cambridge, MA, 1948.
[11] N. Wiener, The Human Use of Human Beings: Cybernetics and Society, Houghton Mifflin, New York, 1950.
[12] N. Wiener, What is Cybernetics? in: P.P. Wiener (Ed.), Readings in the Philosophy of Science: Introduction to the Foundations and Cultural Aspects of the Science, Charles Scribner's & Sons, New York, 1953, pp. 100–110.
[13] S.J. Heims, Constructing a Social Science for Postwar America: The Cybernetics Group 1946–1953, The MIT Press, Cambridge, CA, 1993.
[14] M. Maruyama, The second cybernetics: deviation-amplifying mutual causal processes, American Scientist 51 (1963) 164–179.
[15] M. Maruyama, Mutual causality in general systems, in: J.H. Milsum (Ed.), Positive Feedback: a General Systems Approach to Positive/Negative Feedback and Mutual Causality, Pergamon Press, Oxford, 1968, pp. 80–100.
[16] J.J. Rocchio, Relevance feedback in information retrieval, in: G. Salton (Ed.), The SMART Retrieval System: Experiments in Automatic Document Processing, Prentice-Hall, Englewood Cliffs, NJ, 1971, pp. 313–323.
[17] K. Sparck Jones, Reflections on TREC, Information Processing and Management 31 (3) (1995) 291–314.
[18] G. Storrs, A conceptualization of multiparty interaction, Interacting with Computers 6 (2) (1994) 173–189.
[19] S.E. Brennan, E.A. Hulteen, Interaction and feedback in a spoken language system: a theoretical framework, Knowledge-based Systems 8 (2-3) (1995) 143–151.
[20] M.A. Perez-Quinones, J.L. Silbert, A collaborative model of feedback in human–computer interaction, in: CHI' 96: Proceedings of the Association for Computing Machinery Special Interest Group on Computer–Human Interaction (ACM/SIGCHI), Conference on Human Factors in Computing Systems, Vancouver, British Columbia, Canada, April 13–18 1996, pp. 316–323.
[21] B. Shneiderman, Designing the User Interface: Strategies for Effective Human–Computer Interaction, Addison-Wesley, Reading, MA, 1978.
[22] A. Brandes, U. Wilensky, Treasureworld: a computer environment for the study and exploration of feedback, in: Constructionism, Ablex Press, Norwood, NJ, pp. 391–415, 1993.
[23] D.A. Norman, The Psychology of Everyday Things, Basic Books, New York, 1988.
[24] M. Resnick, Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds, The MIT Press, Cambridge, MA, 1994.
[25] J.V. Dempsey, G.C. Sales, Interactive Instruction and Feedback, Educational Technology Publications, Englewood Cliffs, NJ, 1993.
[26] W.B. Belkin, W.B. Croft, Retrieval techniques, Annual Review of Information Science and Technology 22 (1987) 109–145.
[27] D. Harman, Relevance feedback revisited, in: Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in information Retrieval, Copenhagen, Denmark, June 21–24 1992, pp. 1–10.

[28] P. Ingwersen, Cognitive perspectives of information retrieval interaction, Journal of Documentation 52 (1) (1996) 3–50.
[29] N.J. Belkin, C. Cool, A. Stein, U. Theil, Cases, scripts, and information seeking strategies: on the design of interactive information retrieval systems, Expert Systems with Applications 9 (3) (1995) 379–395.
[30] T. Saracevic, Modeling interaction in information retrieval (IR): a review and proposal, in: Proceedings of the 59th Annual Meeting of the American Society for Information Science, 1996.
[31] T. Saracevic, H. Mokros, L. Su, Nature of interaction between users and intermediaries in online searching: a qualitative analysis, Proceedings of the 53rd Annual Meeting of the American Society for Information Science 27 (1990) 47–54.
[32] T. Saracevic, H. Mokros, L. Su, A. Spink, Interaction between users and intermediaries in online searching, Proceedings of the 12th Annual National Online Meeting 12 (1991) 329–341.
[33] T. Saracevic, L. Su, Nature of interaction between users and intermediaries in online searching: a qualitative analysis, Proceedings of the 52nd Annual Meeting of the American Society for Information Science 26 (1989) 75–80.
[34] A. Strauss, J. Corbin, Basics of Qualitative Research: Grounded Theory Procedures and Techniques, Sage Publications, Newbury Park, CA, 1990.
[35] A. Spink, T. Saracevic, Interactive information retrieval: sources and effectiveness of search terms, Journal of the American Society for Information Science 48 (8) (1997) 728–740.
[36] L. Schamber, Relevance and information behavior, Annual Review of Information Science and Technology 29 (1994) 3–48.
[37] B.M. Wildemuth, E.K. Jacob, A. Fullington, R. deBliek, C.P. Friedman, A detailed analysis of end-user behaviors, Proceedings of the 54th American Society for Information Science Annual Meeting 28 (1991) 302–312.
[38] F. Wood, N. Ford, C. Walsh, The effect of postings information on searching behavior, Journal of Information Science 20 (1) (1994) 60–71.
[39] D.C. Blair, Language and Representation in Information Retrieval, Elsevier Science, New York, 1990.
[40] S.P. Harter, Online Information Retrieval, Academic Press, San Diego, CA, 1986.
[41] C.H. Fenichel, Online searching: measures that discriminate among users with different types of experience, Journal of the American Society for Information Science 32 (1981) 23–32.
[42] R. Fidel, Moves in online searching, Online Review 9 (1) (1985) 61–74.
[43] M.J. Bates, Search tactics, Annual Review of Information Science and Technology 16 (1981) 139–169.