



<http://researchoutput.csu.edu.au>

This is the Author's version of the paper published as:

Author: P. Hider

Author Address: phider@csu.edu.au

Title: Search goal redefinition through user-system interaction

Year: 2007

Journal: Journal of Documentation

Volume: 63

Issue: 2

Pages: 188-203

ISSN: 0022-0418

DOI: <http://dx.doi.org/10.1108/00220410710737178>

Keywords: search goal redefinition user system interaction

Abstract: Purpose of this paper: A study examining search goal redefinition during users' interaction with information retrieval systems is reported. Methodology: The study utilised transaction logs from the OCLC FirstSearch service. Within each search session, logged queries were coded chronologically, according to their conceptual relationships, and indices of goal redefinition were constructed. Redefinition levels for different databases were compared, and certain features were isolated for examination as possible redefinition factors. Findings: The transaction log analysis showed that different databases induced goal redefinition to different extents and identified several factors which can contribute to goal redefinition, including the presence of abstracts and hyperlinking descriptors. On the other hand, no evidence was found to indicate that abstract length has much effect on redefinition, nor hit rate or retrievability of records. Research implications: The research needs to be followed up using other methodologies and other information retrieval systems. A range of other possible factors affecting goal redefinition should be investigated. Value: The paper shows that system feedback affects not only strategy, but also higher levels of information seeking behaviour interaction. This aspect of user-system interaction has rarely been researched. An index of goal redefinition and an interpretative form of transactional log analysis are put forward as means by which it may be investigated.

Search Goal Redefinition through User-System Interaction

ABSTRACT

Purpose of this paper

A study examining search goal redefinition during users' interaction with information retrieval systems is reported.

Methodology

The study utilized transaction logs from the OCLC FirstSearch service. Within each search session, logged queries were coded chronologically, according to their conceptual relationships, and indices of goal redefinition were constructed. Redefinition levels for different databases were compared, and certain features were isolated for examination as possible redefinition factors.

Findings

The transaction log analysis showed that different databases induced goal redefinition to different extents and identified several factors which can contribute to goal redefinition, including the presence of abstracts and hyperlinking descriptors. On the other hand, no evidence was found to indicate that abstract length has much effect on redefinition, nor hit rate or retrievability of records.

Research implications

The research needs to be followed up using other methodologies and other information retrieval systems. A range of other possible factors affecting goal redefinition should be investigated.

Value

The paper shows that system feedback affects not only strategy, but also higher levels of information seeking behaviour interaction. This aspect of user-system interaction has rarely been researched. An index of goal redefinition and an interpretative form of transactional log analysis are put forward as means by which it may be investigated.

Keywords: search goal redefinition user system interaction

Paper classification: research paper

Introduction

Search goals of users of information retrieval systems have commonly been assumed by experimenters to be static and well-formed. However, real-life experience suggests that users may sometimes redefine their search goals as they encounter feedback from a system, even during a single search session. This paper forms the second part of a report on a study examining the extent and some possible reasons why such goal redefinition occurs when users interact with a document retrieval system. A fuller account of the methodological aspects of the study has published as the first part of this report (Hider, 2006).

For the purposes of this paper, *search goals* are defined as goals that a user has in mind when they enter a query into an IR system, representing a particular *information want*, which it hoped will be satisfied by one or more documents retrieved, either directly or indirectly, as a result of the query. A particular information want which forms the basis of an episode of purposeful information seeking, is also described in this paper as an *information goal*. We are defining search goal here in terms of a direct consequence of an information goal, thereby making a distinction between goals and *strategies*. A strategy is employed in order to attempt to fulfill a search goal, and this of course may also be revised as the user encounters results or new functionality during a search session, but the goal on which it is based, may remain the same.

Further, the concept of search goal redefinition should be distinguished from search goal *change*. When a user 'changes' their goal during a search session, they may introduce a new goal or revise their existing one. New goals may be derived from system feedback, but may also be premeditated – the user had more than one search goal in mind when commencing

the search session. It may be assumed that revised, or to use the term employed in this paper, *redefined* goals, on the other hand, would normally be due to interaction with the system, and would not be premeditated. We use the term 'redefined' instead of 'revised' to emphasize this unpremeditated quality.

When users' redefined search goals are met, or partly met, then information acquisition has taken place. In the case of a bibliographic retrieval system, users have acquired citations to documents which they believe may lead to the satisfaction of this goal –bibliographic information acquisition which may lead to information acquisition of a more general kind. The acquisition of information based on redefined search goals would not normally be purely 'accidental,' that is, there would be a degree of causality between the initial information goals of the user at the start of the search session, and the result of goal redefinition during it. However, the result would not have been fully intended at the start of the session, and so in this sense we might consider the phenomenon of goal redefinition as the basis for a form of *incidental* information acquisition.

Incidental information acquisition has been shown to play a major part in everyday information behaviour (Erdelez, 1997; Williamson, 1998), and it may well play a significant part in the real-life use of IR systems. Indeed, Erdelez (1997) calls for a fresh approach to IR systems design, one which more fully acknowledges other kinds of information acquisition apart from the purposeful, matching search. Similarly, Toms (2005) argues that serendipitous discovery as well as well-defined searching and partially-defined searching all need to be properly supported by digital libraries.

It has been observed that systems which facilitate browsing tend to foster greater levels of serendipity, and it may be speculated that they also tend to foster greater levels of goal redefinition. Toms' study (2002) of users browsing a digital newspaper, indicates that content itself can make a difference, with browsers identifying 'landmarks or cues' in the text capable of shifting their 'search focus.' Since matching-type searching also allows for the browsing and evaluation of results, it is possible that elements of bibliographic records might act as similar cues. Evaluations of citations may be carried out according to a fixed conceptualization of an information goal, but they do not need to be. Instead, they may give rise to reviews and revisions of the information (and search) goal. Matching-type searching is commonly characterized as very directed, but the phenomenon of goal redefinition still needs to be investigated.

Literature review

Search or information goal redefinition has not been studied very extensively, particularly not in the context of matching-type searching on information retrieval systems. As was mentioned above, a long tradition of experimental IR research has preferred to assume a static information goal. However, several authors have pointed out the importance of the phenomenon, while a substantial body of research has grown up around related topics, particularly that of relevance.

Some researchers of real-life information retrieval have investigated the dynamism and situatedness of searching. Wang and White (1995), for example, describe how relevance

judgments and criteria changed as new information was encountered during a longitudinal study. Others, such as Hert (1995) and Ng (2002), have examined changes in subgoals during information seeking projects. Xie's research (1997, 2000) investigated 'shifts of interactive intention,' and while many of the shifts analysed were lower-level and strategic, she also noticed 'shifts of current search goal, especially those shifts related to "keeping up to date"' (Xie, 1997). Further, Xie (1997) notes that the importance of 'situation' means that 'not only planning but also feedback from the interaction influences the information seeking process.'

Similarly, Robins (1997, 2000) has examined 'shifts of focus' on an information problem during dialogue between end-user and search intermediary. Robins found many shifts taking place, particularly during online interaction. Although many were strategic or operational in nature, a small proportion appears to pertain to information goals (including those coded 'SCOPE' and 'TOPIC'). Robins (2000) goes on to examine the results of a survey carried out at the time of the dialogue and finds 'moderate evidence' for changes of information problem. It should be noted that the information needs represented by the searches were characterized by Robins (1997) as 'well-defined and stable.' Robins (2000) thus points out that situational factors may play an important role in determining the extent of information problem change, and notes that Kuhlthau's subjects, for example, 'were much younger and perhaps less skilled at formulating research questions' (Kuhlthau, 1991). Robins (1997) concludes that 'one of the remaining questions in information behaviour research is whether users' conceptions of their information problem changes during information retrieval interaction. Further research should, without question, attempt to identify such changes.'

Another study that coded search reformulations is described by Dalrymple (1990). Instead of dialogue, Dalrymple's coding is based on the search terms themselves, according to a protocol analysis (monologue). She compares initial statements of information need with those used in the search session. Dalrymple's study is based on a theory of remembering from psychology, a cognitive process also called 'reformulation,' and investigates how search and pre-search reformulations may vary according to other variables, such as the nature of the IR system. In particular, she compares the frequency of search reformulations produced by card and online formats of a library catalogue. Dalrymple concluded that the card catalogue produced significantly fewer reformulations than did the OPAC, thus demonstrating that the nature of a system can have an effect on the amount of reformulation that takes place. Although Dalrymple does not distinguish between reformulation induced by strategic considerations and reformulation caused by a genuine shift of information want, it is still worth noting two factors she suggests might account for varying amounts of reformulation: the user's frustration with negative feedback, and the 'speed' of a system.

While traditional measures of system effectiveness have generally assumed well-formed relevance judgments from the outset, it has long been recognized that in real life the initial search goal may be imperfectly conceived, and revised upon further reflection. Although expert systems have been developed with the aim of assisting users to formulate their queries as accurately as possible (sometimes with a view to replacing the human expert intermediary), Borgman (1996) stresses that the users' 'conceptual knowledge' of how to

translate an information need into a searchable query involves, at least in part, an evolving consciousness of the information need itself. However, IR researchers have rarely attempted to directly examine this growing consciousness of information want.

A paper which does examine the searcher's consciousness was published by Harmon and Ballesteros (1997). They found that greater elicitation of 'unconscious cognition,' through 'information counseling' and a 'programmed relaxation device,' made for a 'positive impact on the formulations and representation of research problems for the inquirer and the intermediary.' 'Unconscious cognition' is defined by Harmon and Ballesteros as 'that set of cognitive contents and processes in the cognitive system that is unavailable to awareness but nevertheless affects thought and action.' Harmon and Ballesteros were inspired by Taylor's classic theory (1968) of information need development, which pointed to the importance of emerging, but still unexpressed information needs in the search process.

Related to search goal is 'relevance,' a concept which has always been key to IR research. Again, it has been pointed out that while it may be convenient to use fixed, binary relevance judgments for the purposes of laboratory testing, such judgments do not necessarily reflect real-life information seeking. Schamber (1994) thus distinguishes between the 'system view,' the 'information view,' and the 'situation view' of relevance. In the situation view, relevance depends on the situation of the judge, so that judgments may be affected by feedback from the system.

Approaches which recognize how relevance judgments change are listed by Schamber (1994) and include Kuhlthau's model of the information search process (1991). Since then, Tang and Solomon

(1998, 2001) have investigated an end-user's relevance judgments made at both citation and full-text stages of an information retrieval process. Tang and Solomon (2001) show that not only do judgments change in relation to particular documents, but that the criteria for judgments change. Similarly, Bruce (1994) observed that certain metadata elements may become more or less important as clues to relevance across different stages of the information seeking process.

Spink and Greisdorf (1997) have examined 'partial relevance' and found that searchers were more likely to encounter documents they deemed partially relevant if they were to change their information problem definition; or if their interaction with the search system had improved their knowledge of the topic in question; or if they were to revise their criteria of relevance. Spink and Greisdorf (1997) argued that 'partially relevant' documents might also be useful during the early stages of the process, and that the more vague the information problem, the more items were judged partially relevant and the more chance that the user's problem definition was revised.

Spink, Greisdorf and Bateman (1998) analysed users' relevance judgments in terms of both criteria and extent. Differences between users' criteria for highly, partially and non-relevant items were identified. The authors proposed 'a useful concept of relevance as a relationship and an effect on the movement of a user through the iterative stages of their information seeking process.'

Such recent relevance studies are not necessarily based on the notion of changing information goals, however. While users may change their method for making a relevance

judgment, the basis for making the judgment may stay the same. A growing clarity of information want is also not necessarily demonstrated through different degrees of relevance. Indeed, Greisdorf and Spink (2000) themselves point out that degrees of relevance are not the same as ambiguity; in their analysis, it is not that users do not know what they want, they are just not 100% confident that their information want will be met by a particular document.

A relevance judgment is, in effect, a reflection of the user's information want, conceived in terms of a search goal. If goals and judgments change, then so presumably do wants. Few researchers have discussed the dynamism of information wants. Harter (1992) is an important exception. In his paper on the psychological nature of information retrieval, he argues how information needs (or wants) are dependent on context, which changes as new information becomes available and old information re-enters consciousness as it takes on new meaning in the light of new information. We shall be exploring exactly this phenomenon, in terms of the way in which content encountered as feedback from an IR system affects users' perception of their search goal, thus changing relevance judgments.

Methodology

The study attempts to address questions about whether particular system-related variables could be identified as factors affecting the frequency of search goal redefinition. Given the subtle nature of the subject of this research and the impact of many variables involved in real-life search behaviour, a large quantity of data was considered necessary. The transaction logs from five days of OCLC's

FirstSearch service were acquired; approximately two thousand sessions represented on the logs were analyzed, representing many thousands of queries.

Transaction log analysis (TLA) has generally been used to answer basic questions about system or user performance (Peters, 1993). However, the analysis performed on the FirstSearch logs attempted to interpret the use of FirstSearch at a deeper level. To this end, the queries recorded in the logs were coded in a way akin to a content, discourse or protocol analysis. A transaction log may not present us with a full picture of the searcher's thought processes, but it is assumed in this study that there is a significant correlation between queries input into an IR system, and thoughts about that IR system and its output. Ultimately, all analyses of thought processes require interpretation of the data derived from those thought processes. This particular TLA technique has been discussed at greater length in a separate paper (Hider, 2006).

Possible redefinition factors

There are in fact over 70 databases which can be accessed (usually through the Web) in the FirstSearch service, including WorldCat, the OCLC Union Catalogue. Most of the databases contain records of journal articles and other documents, some of which include links to the full text (according to the user's authorization). Importantly for the purposes of this study, the FirstSearch interface is broadly the same, irrespective of which databases are being searched. The service is designed for end-users, although some libraries may offer users a mediated search service. We are assuming here that the amount of mediated searching does not vary significantly across databases, given that this amount is probably quite small (Tenopir, 2001).

However, other aspects of the FirstSearch databases do vary, and the study involved testing some of these variables as potential factors of search goal redefinition, through the comparison of goal redefinition levels for sessions on different databases. For instance, levels of goal redefinition on databases with abstracts could thus be compared with levels on databases without abstracts.

Two particular metadata elements were considered likely candidates for goal redefinition factors, at least in relation to subject searching: abstracts and subject keywords (descriptors and identifiers). Both elements were displayed on full records, but not at the citation level. Unfortunately, the FirstSearch logs did not detail which full records (if any) were displayed, so we need to assume that users were exposed to metadata from full records that typified the metadata available on the various databases.

As well as frequency of abstracts and subject keywords, another variable identified for testing was abstract length. Other variables examined in the study were: searching on bibliographic-only databases versus searching on databases with full-text access; item-level versus article-level searching (i.e. the difference between searching for whole items such as monographs, serials, etc., and searching for articles and papers in journals, proceedings, etc.); hit rate; retrievability (the potential for records to be retrieved through more index entries and/or less precise searching strategies); and subject/discipline area.

The main datasets, derived from the logs for analysis, were based on sessions on different FirstSearch databases, as listed below. The content of some of the datasets may overlap a little.

1. WorldCat-only
2. Non-WorldCat databases, without any full-text downloading
3. Non-WorldCat databases, with full-text downloading
4. Databases with no abstracts or identifiers (descriptors only), with full-text downloading¹
5. Databases with no abstracts, with full-text downloading²
6. OCLC ArticleFirst
7. ABI/INFORM
8. Business Management
9. Education Abstracts
10. Health & Wellness
11. Medline
12. Periodical Abstracts
13. Sociological Abstracts

Coding

Indices of search goal redefinition were compiled through a rigorous coding of the queries represented in the above datasets. This was possible because the sessions were demarcated in the logs by means of a session code, and within sessions, queries were logged in

chronological order. The coding aimed to identify goal redefinition through comparison of the expressed queries. In particular, it focused on subject searching: what were interpreted as non-subject searches (such as known-item searches) were excluded from analysis.

Of course, certain elements of the logged search expression may not be related to search goal, but may instead be purely strategic. Query reformulations which were considered likely to represent purely strategic changes were discounted. This left query reformulations which *might* have been due to goal revision (or at least in part). It was assumed that the proportion of queries coded as such, but which were in fact purely strategic, was similar across datasets.

A detailed coding system was developed, covering the different elements of syntax found in the logs, such as the various index labels and Boolean operators, as well as the semantic content. There were guidelines on the interpretation of word strings, and on the extent to which the context of other queries could be used.

The coding proceeded through each session's queries chronologically. The first query identified as a subject search, was regarded as an *initial subject query*. Subsequent queries were coded as either a query based on the same goal, a query based on a revised goal, or a query based on a completely different goal. The frequency of redefinition based on an initial subject query could thus be compared.

The coding scheme consisted of a total of twelve codes for different categories of query. One code was assigned for what were deemed non-subject queries, one code for codes which did not substantially modify a previous query, and one code for an initial subject query:

K known-item query

S initial subject query

X repeat query

The other nine codes represented modified queries, thus dependent on an initial subject query:

NS narrowing of initial query in terms of subject

NT narrowing of initial query in non-subject terms

NC narrowing of initial query in both subject and non-subject terms

BS broadening of initial query in subject terms

BT broadening of initial query in non-subject terms

BC broadening of initial query in both subject and non-subject terms

MS mixed modification of subject terms

MT mixed modification of non-subject terms

MC mixed modification of both subject and non-subject terms

The author carried out all the coding on the datasets, but the coding system's reliability was evaluated by means of preliminary parallel coding performed on two specimen log sections, by the author and a colleague. The percentage of queries coded the same, out of the dozen codes available (at least after an initial search), was 91% and 96% for the two sets. The Kappa coefficient (for inter-rater agreement) was 0.75 and 0.79, indicating a 'good' level of agreement.

An index of search goal redefinition based on all nine codes of query modification was supplemented by another index based on just the *NS* and *MS* codes, taking account of only those modifications with substituted or additional subject search terms, which were considered the most likely to represent search goal

redefinition, as opposed to strategic reformulations. The two indices were defined, in the first instance, as the median number of modified queries stemming from each initial subject query; and in the second, as the interquartile mean number of modified queries from each initial subject query. The index embracing all nine modification codes was labeled the *raw index*, while that using the two codes, the *refined index*.

It is assumed that both raw and refined indices would reflect search redefinition uniformly across databases, but it is hypothesised that the refined index would be a purer measure of redefinition, and as such would show similar results, but ones more pronounced, in the sense that the degree of significance or non-significance would be larger than in the case of the raw index, which would be blunted by more strategic query modification.

Control tests

Distributions of the indices derived from two sections of the WorldCat dataset were compared using the Kolmogorov-Smirnov (K-S) 2-sample test. The two sections represented two different days of search sessions. A similar test was carried out using two sections of the Medline dataset.

Results

Redefinition indices

The control tests provided no evidence of significant differences between different sections of the same dataset, in terms of either the raw or refined index. The results are shown in Table I.

Database	Raw index		Refined index	
	<i>D-statistic</i>	<i>p-value</i>	<i>D-statistic</i>	<i>p-value</i>
WorldCat	0.0811	0.781	0.1105	0.398
Medline	0.1566	0.876	0.1756	0.520

Table I

K-S test for control distributions

With mid-high p -values, the null hypothesis of no difference between the pairs of distributions was not rejected. Assuming that search behaviour on different days, but on the same database, is likely to be similar, the result indicates that tests using these redefinition indices are unlikely to produce a false positive, that is, they would not show significant differences between datasets that are not actually present.

For both raw and refined indices, the number of modified queries derived from each initial query was tabulated to produce frequency distributions for each of the thirteen datasets listed above. The median and interquartile mean of these distributions are shown in Table II.

Dataset	Sample size		Raw index		Refined index	
	<i>raw</i>	<i>refined</i>	<i>median</i>	<i>iqmean</i>	<i>median</i>	<i>iqmean</i>
WorldCat	433	433	0	0.36	0	0.03
Non-full-text	627	627	0	0.41	0	0.27
Full-text	287	287	1	1.03	1	0.81
Group B	12	12	1	1.83	0.5	1
Group Z	65	65	1	1.12	0	0.7
OCLC ArticleFirst	22	20	0	0.5	0	0.4
ABI/INFORM	26	26	1	1.29	1	1.07
Business & Mgt	54	45	2	1.86	2	1.65
Education Abstracts	36	28	1	1.11	0.5	0.71
Health & Wellness	216	190	1	1.37	1	1
Medline	161	142	1	1.09	1	0.74
Periodical Abstracts	40	38	1	1.2	0.5	0.75
Sociological Abstracts	32	27	1	1.5	1	1.27

Table II
Central tendencies for distributions

We may observe a reasonable correlation between the two measures of central tendency. The median was considered the theoretically most suitable measure, but in the event of ties, the interquartile means were used as a secondary measure.

Metadata effects

A K-S test was performed on the raw and refined index distributions for dataset #6 in the list above, which represented searching on the ArticleFirst database, which has no abstracts or descriptors, and dataset #3, which represented searches on the full-text databases as a whole, most of which include abstracts and/or subject keywords. Another K-S test was carried out on the difference between index distributions for dataset #3 and for dataset #4, which represented searching on databases with descriptors, but no abstracts. The index distributions of dataset #4 and the ArticleFirst dataset were also compared by means of the K-S test. In all three tests, a significant difference was detected ($p < 0.001$), suggesting that the presence of certain metadata elements has an effect on search redefinition, and in particular, that the abstracts and descriptors in FirstSearch have a bearing on the amount of search redefinition taking place.

Evidence of a more marked 'abstract effect' was sought by performing a Kendall's *tau b* correlation test between rankings of datasets 4-13, in terms of abstract frequency and goal redefinition. However, in this case, the p -values showed no particular evidence of a relationship between abstract frequency and the indices of redefinition. This result does not necessarily contradict the results of the K-S tests, given that the refined index was positively correlated with abstract frequency, producing a p -value of 0.178, suggesting that abstract frequency might nevertheless be a factor on search redefinition, though perhaps not a strong one. (This p -value is reduced to a value verging on the 0.1 level of significance if a different measure of central tendency and/or a dichotomous frequency scale is used.)

The six datasets with a high level of abstract frequency were ranked according to their abstracts' length, and by the redefinition indices. The degree of correlation between the index rankings and the abstract length rankings was again tested using Kendall's *tau b* correlation coefficient. The *p*-values show no particular evidence of a relationship between abstract length and the indices of redefinition. While abstract presence may be a factor on redefinition, the length of an abstract may be of little significance.

Bibliographic versus full-text searching

The difference between searching on databases which offered full text and searching on those that did not (at least not to the user) was tested by comparing the redefinition indices for dataset #2 (no full-text downloading) and dataset #3 (full-text downloading). Again, the K-S test was used. The results showed a significant difference in redefinition ($p < 0.001$), with subject searching on full-text databases likely to result in more redefinition than on those where full text is not available.

Item versus article searching

The K-S test was used to compare redefinition in the WorldCat dataset with that found in dataset #2, which represented searching on other databases without full-text downloading. Dataset #2 databases mostly consist of records for articles; the WorldCat database mostly contains records for item-level documents, such as books and serials. The results show a significant difference in redefinition ($p < 0.001$), with subject searches on WorldCat less likely to produce redefinition than subject searches on periodical indexes.

Hit rate

The proportions of initial queries producing zero hits, fewer than six and fewer than eleven hits were calculated for each of the thirteen datasets; they are shown in Table III as percentages and ranks.

The table also shows the rankings for redefinition.

Dataset	0 hits		<6 hits		<11 hits		redefinition (raw)	redefinition (refined)
	%	rank	%	rank	%	rank		
WorldCat	15.5	4	36.8	5	44	6	13	13
Non-full-text	29.6	12	48.5	12	54.1	12	12	12
Full-text	24	8	43.2	8	51.3	9	10	5
Descriptors only	25	9	37.5	6	37.5	3	2	7
Non-abstract	22.5	7	38.5	7	47.7	7	7	10
OCLC ArticleFirst	0	1.5	0	1	0	1	11	11
ABI/INFORM	8.7	3	30.4	3	39.1	4	5	3
Business & Mgt	15.6	5	34.4	4	40.6	5	1	1
Education Abstracts	0	1.5	0.1	2	0.2	2	8	9
Health & Wellness	20.6	6	44.7	9	50.9	8	4	4
Medline	29.2	11	45.3	10	52.1	10	9	6
Periodical Abstracts	37.8	13	54.1	13	56.8	13	6	8
Sociological Abstracts	28.2	10	46.5	11	53.5	11	3	2

Table III
Hit rate and redefinition

Correlation between the hit rates and the redefinition indices was tested using Kendall's *tau b* correlation coefficient. The results show no evidence whatsoever of a relationship between hit rate and redefinition.

Another examination of the possible hit rate factor was performed using a sample from the Medline dataset, in which the initial queries with ten or fewer hits were identified. The number of hits for each initial query was compared with their number of modified queries, according to raw and refined index definitions. Correlation was tested using Kendall's *tau b* coefficient, with results again indicating no significant correlation between low hit rate and redefinition.

Retrievability

The levels of search redefinition (in raw and refined index terms) occurring immediately following general-keyword and subject-keyword queries in the WorldCat dataset were compared and the proportions tested for association using Fisher's Exact test. The null hypothesis of non-association was not rejected. If there is a relationship between the two variables, it would appear to be weak. That is, the difference in retrievability does not appear to much affect redefinition. Moreover, in this case, the subject-keyword queries, rather than the general-keyword queries, appeared to generate more redefinition, so if this were significantly more, it would invert the theory of retrievability engendering redefinition.

Discipline area

Six of the thirteen datasets pertained to specific discipline areas, aimed at particular audiences. They could be grouped into three pairs: economics, medicine, and sociology. Direct observation suggested that there *might* be some association between discipline group and redefinition index. More investigation is needed.

Conclusions and discussion

Redefinition indices

The indices employed in this study are not measures of the amount of redefinition taking place *per se*; rather they are intended as reasonably reliable measures of the *relative* amounts of 'strong' search redefinition occurring across the FirstSearch databases, where strong redefinition is defined as that which results in follow-up queries during the same session. This is not the only kind of goal redefinition that may take place during search sessions: 'weak' redefinition may also occur, whereby a goal is modified, but no further queries are performed, either because sufficient information is derived from the existing results, or because the user does not consider it worthwhile to attempt another query.

The redefinition indices are blunted, it is assumed, by a good deal of noise. That is, some of the queries registered in the indices, especially in the raw index, are actually not the product of redefinition, but instead the product of a modification of search *strategy*. Conversely, the refined index in particular is likely to have missed queries that were in fact the product of redefinition (at least in part). In these ways, the indices may show smaller differences between amounts of strong redefinition than is in fact the case. Given this possible bluntness, we should be more confident about a statistical test which shows a real difference.

The two tests performed on the control sample indices provided evidence that the indices do not produce false positives, while some of the tests performed on the indices for other datasets showed statistically significant differences. This does not in itself mean that the

indices accurately represent differences in the amount of search redefinition. However, query modification, on which the indices are based, can only otherwise be explained in terms of changing search strategy. The key assumption here, then, is that strategy plays a reasonably uniform role in query modification across databases and datasets. If the statistically significant differences cannot be put down to differences of strategic behaviour, then they must be attributed to different amounts of goal redefinition. As such, the indices may be accepted as indices of redefinition. Further discussion as to the indices' validity and reliability is provided in the first part of this report (Hider, 2006).

Comparing the two indices of redefinition, we find that the refined index consistently shows greater differences between redefinition than does the raw index. This consistency suggests that the refined index is a more sensitive one.

Impact of system

The K-S tests show statistically significant differences in the redefinition indices across various FirstSearch databases. For the purposes of this study, different FirstSearch databases represent different IR systems (they can be interrogated independently). Assuming that other aspects of users' search behaviour are uniform for at least some of the dataset comparisons, we may conclude that the nature of the information retrieval system can have an impact on search goal redefinition. Further, by examining the shapes and central tendencies of the index distributions, we observe that on the average some systems produce more search redefinition than do others.

Impact of metadata

Comparison between indices for datasets representing different types of metadata indicated a 'metadata factor.' The lack of evidence to support a retrievability factor would point to the lack of *encountered* metadata (as opposed to *indexed* metadata) being a factor working against redefinition in FirstSearch subject searching.

A metadata element which appears to be a particular factor, when encountered, is the abstract, though more research needs to be done on the impact of retrievability of abstracts specifically, before we can firmly conclude that it is the displayed abstract, rather than the indexed abstract, that is the larger factor. In any case, the correlation test on abstract frequency and redefinition index did not indicate that either the indexed or displayed abstract had a huge effect. The abstract factor may be much stronger if users display more abstracts (e.g. by opening up more full records).

Comparison of indices also pointed to a 'descriptor effect.' It could well be that descriptors provide cues for search goal revision, but it is speculated that it is the *hyperlinked* nature of the descriptors in the FirstSearch system which is the largest factor here. Users following the hyperlinks would enter new queries which would have often been registered in the redefinition indices. The ease with which hyperlinks produce fresh result sets probably encourages more exploration of the databases, and this is also likely to increase search redefinition.

Other redefinition factors

It would appear that metadata factors are by no means the only ones affecting goal redefinition. Redefinition on full-text databases occurred more frequently than it did on bibliographic-only databases; indeed, the non-full-text average is clearly lower than any of those for the full-text datasets. This suggests that the presence of full text has a major effect on redefinition, probably

more so than metadata differences. It might be that searching is more 'ambitious' when the user is presented with the possibility of full text, which may lead to more critical evaluation of results and review of the information goal.

Within the range of non-full-text databases, WorldCat would appear to induce significantly less redefinition than others. This could well be due to the difference of document representation, that is, WorldCat represents whole items (such as monographs), the other databases represent articles. It may be that this difference translates into broader goals on the part of WorldCat users which need less refinement.

Not all variables considered possible redefinition factors turned out to be so. Although a preliminary study (Hider, 2004, pp. 56-76) had indicated that redefinition might sometimes be due to search failure, the hit rate analyses in this study showed no particular association. Similarly, no evidence was produced to support the hypothesis that greater retrievability leads to more redefinition; indeed, if there is any relationship between retrievability and redefinition, the analysis suggested that more retrievability makes for less redefinition, even though more retrievability would generally mean less precision. It may be that a certain mid-level of precision is optimal for redefinition, with the user encountering citations which are *partially* relevant, as suggested by the relevance research carried out by Spink and Greisdorf (1997). It may also be that, notwithstanding an optimal level, precision is not a particularly important factor, such that no measure of relevance would be a reliable indicator of redefinition potential.

This study did not really show whether or not there is a relationship between discipline area and redefinition. It is possible that discipline is a factor and, as such, might reduce association between redefinition and other variables.

Implications

It would be exceptionally difficult to perform a cost-benefit analysis on search redefinition – very hard to estimate the additional time a user might spend, following redefinition, and even harder to gauge the additional information gain that might result.

Furthermore, it is important to note that an IR system which induces more search goal redefinition may do so because it facilitates less opportunity for users to clarify their goals at the start of the session.

Nevertheless, we may assume that the system should facilitate search redefinition as much as possible, all other aspects of user-system interaction being equal. According to this study's findings, it would thus seem reasonable to call for abstracts to be included in all bibliographic records. However, there are costs involved in doing this, and it may be that some abstracts are almost as effective as many abstracts. What we may propose, instead, is that bibliographic system producers should consider including abstracts in their records if they wish to provide a quality service and can obtain the abstracts without significant delay or overly burdensome financial outlay. Abstracts need not necessarily appear in all records, nor do they need to be very long, for the purposes of inducing redefinition. The more frequently abstracts are encountered through the system – for instance, they might be encountered much more frequently if displayed at the citation level – the more important is this recommendation.

What became clear from this research was that the advent of hyperlinks has very likely had a large impact on the amount of 'strong' redefinition that takes place on modern document retrieval systems. When a user realizes that a different term would more accurately represent their search goal, instead of having to start over and type in the new term, they can simply click on the highlighted term within a citation and new search results appear. System designers have rightly observed that even two or three additional mouse and/or keyboard actions, and two or three additional screen redraws, can deter the user from implementing a new query.

More hyperlinks, from more words in the record display, would probably increase still further the amount of follow-up on redefinition. Most contemporary systems do not present hyperlinks on words in titles or subtitles, or abstracts. It is feasible for a title or abstract to be initially displayed without individual words highlighted as hyperlinks, but for these to appear as a cursor is moved over them. This would minimize distraction while giving the user a great deal more hyperlinking scope. A 'synthetic hyperlinking' feature might also be attractive to more sophisticated users, whereby various words or phrases in a display, could be combined (in Boolean fashion), perhaps with right clicks, and then searched on. Admittedly, fairly sophisticated programming would be required to achieve this.

While developing such new features, we must not forget the importance of user education in the pursuit of search goal redefinition. Users can be educated to conceptualize a more accurate 'mental

model' not only of the system, but of the search process itself, and thus appreciate the difference greater thought about their information wants can make to their information gains. Users can also be taught to review their information goals through action – for example, by taking the time to read more of the metadata, including abstracts.

The author's preliminary research (Hider, 2004, pp. 56-76) on search goal redefinition, suggested that, at least in certain situations, redefinition can take place frequently, even at the level of the search session, while the study described in this paper has indicated that system feedback can shift the direction of a user's search in terms of goal as well as strategy. However, models of information retrieval have left little room for this phenomenon, and this is something which needs to be addressed.

This study has produced many more questions than answers. Much more research on the phenomenon of search goal redefinition is required. A range of other user and system

variables need to be investigated as possible redefinition factors, while results need to be triangulated using other methodologies and measures.

References

- Borgman, C.L. (1996), "Why are online catalogs still hard to use?", *Journal of the American Society for Information Science*, Vol. 47 No. 7, pp. 493-503.
- Bruce, H.W. (1994), "A cognitive view of the situational dynamism of user-centered relevance estimation", *Journal of the American Society for Information Science*, Vol. 45 No. 3, pp. 142-148.
- Dalrymple, P.W. (1990), "Retrieval by reformulation in two library catalogs: toward a cognitive model of searching behavior", *Journal of the American Society for Information Science*, Vol. 41, pp. 272-281.
- Erdelez, S. (1997), "Information encountering: a conceptual framework for accidental information discovery", in Vakkari, P., Savolainen, R. and Dervin, B. (Eds.), *Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, Taylor Graham, London, pp. 412-421.
- Greisdorf, H. and Spink, A. (2000), "Recent relevance research: implications for information professionals", *Online Information Review*, Vol. 24 No. 5, pp. 389-395.
- Harmon, E.G. and Ballesteros, E.R. (1997), "Unconscious cognition: the elicitation of deeply embedded information needs", in Vakkari, P., Savolainen, R. and Dervin, B. (Eds.), *Information Seeking in Context: Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, Taylor Graham, London, pp. 422-433.
- Harter, S.P. (1992), "Psychological relevance and information science", *Journal of the American Society for Information Science*, Vol. 43 No. 9, pp. 602-615.

Hert, C.A. (1995), "Information retrieval as situated action", in Kinney, T. (Ed.), *ASIS '95: Proceedings of the 58th ASIS Annual Meeting*, Information Today, Medford, pp. 172-180.

Hider, P.M. (2004), "User redefinition of search goals through interaction with an information system", PhD thesis, City University, London.

Hider, P. (2006), "Constructing an index of search goal redefinition through transaction log analysis", *Journal of Documentation*.

Kuhlthau, C.C. (1991), "Inside the search process: information seeking from the user's perspective", *Journal of the American Society for Information Science*, Vol. 42, No. 5, pp. 361-371.

Ng, K.B. (2002), "Toward a theoretical framework for understanding the relationship between situated action and planned action models of behavior in information retrieval contexts: contributions from phenomenology", *Information Processing and Management*, Vol. 38 No. 5, pp. 613-626.

Peters, T.A. (1993), "The history and development of transaction log analysis", *Library Hi Tech*, Vol. 110, pp. 41-66.

Robins, D. (1997), "Shifts of focus in information retrieval interaction", in Schwartz, C. and Rorvig, M. (Eds.), *ASIS '97: Proceedings of the 60th ASIS Annual Meeting*, Information Today, Medford, pp. 123-134.

Robins, D. (2000), "Shifts of focus on various aspects of user information problems during interactive information retrieval", *Journal of the American Society for Information Science*, Vol. 51 No. 10, pp. 913-928.

Schamber, L. (1994), "Relevance and information behavior", *Annual Review of Information Science & Technology*, Vol. 29, pp. 3-48.

Spink, A. and Greisdorf, H. (1997), "Partial relevance judgments and changes in users' information problems during online searching", in Williams, M.E. (Ed.), *Proceedings of the 18th National Online Meeting*, Information Today, Medford, pp. 323-334.

Spink, A., Greisdorf, H. and Bateman, J. (1998), "From highly relevant to not relevant:

examining different regions of relevance”, *Information Processing and Management*, Vol. 34 No. 5, pp. 599-621.

Tang, R. and Solomon, P. (1998), “Toward an understanding of the dynamics of relevance judgment: an analysis of one person’s search behavior”, *Information Processing and Management*, Vol. 34 No. 2-3, pp. 237-256.

Tang, R. and Solomon, P. (2001), “Use of relevance criteria across stages of document evaluation: on the complementarity of experimental and naturalistic studies”, *Journal of American Society for Information Science and Technology*, Vol. 52 No. 8, pp. 676-685.

Taylor, R.S. (1968), “Question-negotiation and information seeking in libraries”, *College and Research Libraries*, Vol. 29, pp. 178-194.

Tenopir, C. (2001), “Online databases: database and online system usage”, *Library Journal*, Vol. 126 No. 16, pp. 41-42, 45.

Toms, E.G. (2002), “Information interaction: providing a framework for information architecture”, *Journal of the American Society for Information Science and Technology*, Vol. 53 No. 10, pp. 855-862.

Toms, E.G. (2005), “Serendipitous information retrieval”, available at:
http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf.

Wang, P. and White, M.D. (1995), “Document use during a research project: a longitudinal study”, in Kinney, T. (Ed.), *ASIS '95: Proceedings of the 58th ASIS Annual Meeting*, Information Today, Medford, pp. 181-188.

Williamson, K. (1998), “Discovered by chance: the role of incidental information acquisition in an ecological model of information use”, *Library and Information Science Research*, Vol. 20 No. 1, pp. 23-40.

Xie, H. (1997), “Planned and situated aspects in interactive IR: patterns of user interactive intentions and information seeking strategies”, in Schwartz, C. and Rorvig, M. (Eds.), *ASIS '97: Proceedings of the 60th ASIS Annual Meeting*, Information Today, Medford, pp. 101-110.

Xie, H. (2000), "Shifts of interactive intentions and information-seeking strategies in interactive information retrieval", *Journal of the American Society for Information Science*, Vol. 51 No. 9, pp. 841-857.

Notes

1 Databases:

Applied Science & Technology Index

Art Index

Biological & Agricultural Index

Education Index

General Science Index

Humanities Index

Library Literature

Social Sciences Index

2 Databases:

Applied Science & Technology Index

Art Index

Biological & Agricultural Index

Biography Index

Business Dateline

Education Index

Electronic Collections Online

General Science Index

Humanities Index

Library Literature

MLA International Bibliography

Social Sciences Index