

Known-item searching: variations on a concept

Abstract

The concept of known-item searching has long been central to research and application in library and information science. It is surprising then that this concept has received practically no systematic discussion. We survey the various conceptual and operational characterizations of known-item searching in the LIS literature in order to determine exactly how the concept is being understood by its users. We demonstrate that this apparently simple notion is actually quite complex and varied, and moreover, that there is hardly a single feature ordinarily associated with it that can confidently be said to be an essential part of the concept.

1. Introduction

Researchers and theorists in the information science community have upon many occasions called for more rigor in the basic concepts and terminology of our field. Although interest in increased rigor may seem especially intense at present (Hjørland, 2005a), the concern is not new:

Information retrieval as a discipline stands in need of a conceptual and theoretical framework, of an articulation of its fundamental principles, and identification of key issues and problems. It is particularly curious, in the light of its importance in bringing order and structure to other disciplines, that it itself has been so deficient in these same qualities.

Don Swanson, Foreword to *Principles of Information Retrieval*
(Kochen, 1974)

This paper is the start of an exercise toward that end, focusing on the concept of a *known-item search*. The importance of this concept to the field is manifest. It is not only widely used in explanations, hypotheses, and theories of information seeking, but is a stock category in the literature of bibliographic instruction. What at first appears to be a straightforward concept is really quite complex. Sources of complexity include (i) differences in orientation: cataloging, information retrieval, reference, information seeking behavior studies; (ii) differences in level of discussion – the concept vs. its operationalization in a particular system; and (iii) greater variation as the technology supporting information retrieval has evolved.

We review the use of the concept in Library and Information Science (LIS) research and survey previous efforts to articulate definitions in order to demonstrate the complexity and inconsistent use of this concept. The organizing framework for the survey is an investigation of assumptions underlying these definitions. We consider counterexamples, boundary cases, and variations on the theme. We show that certain familiar characterizations are problematic and explore some alternatives. In the process the conceptual contours of this notion are identified and the options for formalization become clearer.

2. Significance of the Study

“Known-item search” is one of the most widely deployed concepts in the field of library and information science. While the dichotomous view of categorizing searches into “known-item search” and “subject search” has been criticized by authors like Buckland (1979), and a number of other additional types of searches have been suggested (e.g., “bibliographic searches” (Lipetz, 1972), “location searches” (Specht, 1980), “existence searching”, “exploratory searching”, “comprehensive searching” (Rosenfeld & Morville, 1998), “negative searching” (Stielow & Tibbo, 1988), “unknown item searches”, “area searches” (Slone, 2000)), the concept of a known-item search has continuously appeared in a wide range of LIS literature for the last several decades. In catalog studies, especially those exploring OPAC (Online Public Access Catalog) use, known-item search is often listed as one of the two major types of searches, along with subject search (Matthews et al., 1983). Known-item search commonly appears as a component or in an explanation of the search modes in information seeking and retrieval models (Ellis, 1989; McKenzie, 2003), and is a common task characterization in the evaluation of information retrieval systems and techniques (TREC: Text REtrieval Conference) (Voorhees, 2004). Finally, in the numerous bibliographic instructions (e.g., searching tutorials) provided for library users on various library websites, known-item searches are typically distinguished from subject searching or browsing, and given a specific search strategy suitable for that type of search.

Most researchers articulate their own conceptual and operational definitions of a known-item search, making little effort to explicitly connect these to the general concept and rarely providing citations to sources or authorities. For the most part researchers seem to follow their intuitions or common sense on

what a known-item search is, and then develop operational definitions that seem, again, on intuitive grounds, to be roughly coextensive with known-item searching in the particular research context they are exploring. The result of course is that definitions and assumptions vary, and the significance and comparability of the findings are undermined.

3. Method

For this study, we searched for books, articles, and web resources in various areas of LIS (e.g., catalog use studies, information retrieval, information seeking behavior, web searches, reference service) that specifically deal with a known-item search in an empirical way (e.g., comparison of known-item searches with subject searches, exploration of what information elements are typically known by users in known-item searches) and also for the resources in which researchers attempt to provide conceptual definitions of a known-item search. All the works indexed under the subject descriptor “known-item search” in major LIS article indexes and digital libraries were retrieved and examined. Additional resources were obtained by following the cited references in the bibliographies of the papers already acquired. Some additional resources were found on the Web by using common search engines.

The relevant passages were extracted by examining all the collected resources and analyzed by the authors in a series of discussions. The examples that are presented in this paper are selected (i) to show typical representations of definitions of a known-item search in certain sub-areas of the LIS domain, (ii) to exemplify conflicting characterizations and variations in defining the concept, and (iii) to help us effectively present boundary cases and counterexamples.

4. Operational and Conceptual Definitions of a Known-Item Search

We will now look at various characterizations of known-item searching in the literature. In this section, we provide some exemplary definitions to illustrate problems that will be explored in more detail in section 5. In what follows we assume that there are some relatively conceptual understandings of known-item searching that are independent of the specific operational definitions designed to be used to identify known-item searches in a

particular context. We specifically want to draw the readers' attention to the issues caused by mixed uses of operational and conceptual definitions of the concept in the literature.

In the context of card catalogs, one particular operational definition of known-item searching seemed particularly natural and was commonly used: a search for an item for which the author or title is known. However, the introduction of an OPAC and the Internet resulted in the addition of numerous new access points and this definition is now inappropriately restrictive. This is partly due to the fact that originally a known-item search had been defined operationally with little attention to whether the operational definition in fact matched the concept, and there was usually almost no effort to clarify the concept itself.

A catalog is the single, most important key to a library's collections. Its major function is to show whether the library owns a particular bibliographic item whose author and/or title are known (known items) and, if so, where it is located. (Lancaster & Joncich, 1977; p.19)

Here Lancaster and Joncich are probably best interpreted as saying that the catalog has a major function supporting a search where author and title are known, and those are known-item searches, without intending that to be a definition of known-item searching. But it is obvious that the distinction can be difficult to maintain in the context of observing and discussing catalog use.

Searching for known items by subject is very inefficient, but can be successful when other approaches fail. (Lewis, 1987; p. 153)

Here we see a recognition that known-item searches can be carried out by subject, reflecting a conceptual understanding that is wider than the typical operationalization.

Most searches are of two types: a search for some item for which either the author or title is known (called a known-item search) or a search for some item on a particular subject (American Library Association, 1958; Hafter, 1979). (Baker & Lancaster, 1991; p.200)

This is a typical characterization that nicely illustrates the difficulty in distinguishing operational and conceptual definitions. The description of subject search is at a fairly conceptual level, but the correlative characterization of known-item searching appears operational. One could understand the authors to be saying that "a search for some item for which

either the author or title is known” simply *is* “a known-item search” which would exclude many searches that are commonly thought of as searches for known items, or we could more charitably interpret this as an informative operational definition even though the corresponding definition of subject searching is conceptual. When using a subject card catalog, this searching based on a particular subject could possibly be understood as an operational definition as well. However, outside of the context of using the card catalog, “a search for some item on a particular subject” can easily come across as keyword searching, not searching under subject headings, thus understood as a conceptual definition rather than an operational one.

Some people who approach a library catalog have a particular item in mind, and they want to determine whether the library holds that item and where in the library it is located. Such a person would conduct a known-item search. A known-item search may include the author, the title, the subject, or a combination of these and other pieces of information to identify the item in the catalog. (Wildemuth & O’Neill, 1995; p.265)

Phrases such as “have a particular item in mind” are also common characterizations of the general concept of “known-item searching”. When Wildemuth and O’Neill go on to say “Such a person would conduct a known-item search. A known-item search may include the author, the title, the subject, or a combination of these and other pieces of information”, it seems unclear whether they intend to say that some particular search strategy implements the “known-item search” or whether the strategy is the search itself. The ambiguity is perhaps unresolvable in virtue of the tight focus on the context of using a card catalog or OPAC. But one may still ask: would Wildemuth and O’Neill consider it a known-item search if the patron used some unanticipated search strategy to find the entity in mind?

5. Exploring the Assumptions

We now attempt to identify the specific conditions that seem to be understood by LIS researchers as constituent of known-item searching. Here our principal focus is on the characterization of the concept, rather than operationalizations of that characterization, or descriptions of how known-item searches are usually carried out. However we attend to both and note that they are often blended or conflated in ways difficult to untangle. At this time we do not assume any particular theory of what a conceptual

characterization is, or should be. Rather we simply explore what information scientists have said and assumed about known-item searching, clarifying and comparing their characterizations and testing them against intuitions about cases.

5.1. Is a Special Epistemic Relationship Required?

The most prominent feature of definitions of known-item searching is the requirement that the user is searching for a “known” object or an object “known to exist”. In the next section we will take up this requirement in its most general form. In this section we discuss a more restrictive variant: that the user has a distinctly close relationship to the object sought, one closer than simply knowing it exists.

This requirement is suggested by the use of words like “exact” or “specific” or “particular” in phrases like “seeking a specific known item”. These may be redundant adjectives of course, but we think that their prevalence and similarity suggest that they are not redundant, but that they are being used to express, however vaguely, an additional restriction. Moreover there are also other phrases that suggest a particular close connection with the object, an acquaintance that goes beyond mere knowledge of existence:

...a specific work which he knows to exist – possibly one *with which he has had previous contact*. (Swanson et al., 1968; p.1)

...a situation in which a user is trying to find an item previously read, and consequently in which *the user’s memory of the item is of primary importance*. (Allen, 1989; p.247)

...to reach a particular site that the user has in mind, *either because they visited it in the past* or... (Broder, 2002; p.5)

[Italics added]

Here we see specific references to the kinds of close contact the user may have had with the item (“previously read”, “visited it in the past”), and the kind of current cognitive/epistemic relationship he may have (the “user’s memory of the item”).

Is something more than mere knowledge required and if so what would that be? We present two cases that suggest competing intuitions.

Case 1: In support of the claim that a special relationship is required, consider the situation where someone goes looking for a textbook on

differential equations, justifiably sure (as we all are) that there are some, but the user has no specific memory of ever having actually seen such a textbook, or seen a citation to one, or heard a particular edition mentioned. In such a situation we would be reluctant to say that the user has a “specific known entity in mind” (an exemplary characterization of known-item searching), even though we would probably say that they are looking for something that they know exists.

Case 2: Against the claim that a special close epistemic relationship to the object is required, consider the common situation where someone obtains information about an item from article citations, or from a class reading list, or from a colleague and on that basis begins looking for the item. These would seem to be unexceptional known-item searches, and yet the user’s relationship to the object is hardly close.

5.2. Is Knowledge Required?

Now we turn to whether knowledge of the item sought is required *at all*, in any sense. This may seem an odd question, since references to knowledge of the item sought are so prominently mentioned in characterizations of known-item searching, but it is nevertheless quite possible that literal interpretations of these informal characterizations do not accurately capture the concept we have in mind. And it is also possible that the psychological effect of these phrases has prevented us from seeing just how little knowledge has to do with known-item searching.

Traditionally in epistemology there are three conditions for knowledge: (i) the knower has the appropriate belief; (ii) the knower is justified in this belief; (iii) the belief is true. We take up each one in turn, considering them as possible individual requirements for known-item searching. If any one of them is not a requirement, then knowledge is not a requirement.

5.3. Is Existence of the Sought Item Required?

Here we have thrown doubt on whether the existence of the object is required and therefore whether knowledge is required. But we may also wonder whether existence is required even if knowledge is not. It might be that knowledge is not required because we do not require the user to have that high level of justification for their belief, but nevertheless we do expect them to be searching for something that exists.

Consider the case of a citation that despite appearances does not refer to an actual article. To avoid difficult cases, imagine that the citation is deliberately concocted, but located in a scientific article in a respected journal and with no hints of playfulness or any other reason to believe the citation is not authentic. Consider now the innocent (duped) user searching for this item. Doesn't that appear to be a known-item search? But because there is no item to be found, there is no item "known to exist".

One possible explanation for why our intuitions are indifferent to the existence of the item would be that our concept of known-item search is wholly about *the state of mind of the user*, and has no dependency on the external world. This is in fact suggested, indirectly, by Bates when she writes:

The user's experience is phenomenologically different from the indexer's experience. The user's task is to describe something that, by definition, he or she does not know (cf. Belkin, 1982). (Knowledge specifically of what is wanted would lead to a "known-item" search.) (Bates, 1998; p.1186)

Just as Bates spoke of a "phenomenological" *difference* between the indexer's and user's experience, we might say that there is phenomenological *similarity* (in all relevant respects) between the experience of someone looking for the referent of a genuine veridical citation and someone looking for the referent of a concocted citation. Seeing known-item searching as being only about the cognitive state of the user is consistent with its actual role in research in information seeking where its explanatory function is always vis-à-vis the user's beliefs and never turns on external facts about the search environment except insofar as they are manifest themselves in the user's experience.

We now have competing intuitions: (i) known-item searching requires a known item; and (ii) known-item searching is a characteristic of the user's cognitive state, of the user's phenomenologically identified experience, and does not depend on any external facts about the world. There are tradeoffs in each case. If we prefer the first then we seem to allow that at least one intrinsic characteristic of the concept is actually irrelevant to any explanatory role that it has in cognition and behavior, and we will in fact lose some of the explanatory power of the concept of known-item searching: it will not collect cases that are behaviorally indistinguishable from others. But if we prefer the second then we must revise our terminology and most

importantly, give some account of what it is that *is* required and which we attempted to capture with the (misconceived) knowledge requirement.

The natural move of course is to reduce the burden of the knowledge requirement to specifying only that the user *believes* they know that the object exists, but not require that they actually do know. This reflects both the sense that something related to knowledge of the object is required, as well as the intuition that only the phenomenologically identified cognitive state is relevant.

5.4. Is a *Justified* Belief that the Sought Item Exists Required?

We examine this weaker requirement, that only belief is required, in two versions. The weakest version of this weak requirement that only belief, not knowledge, is necessary is that only *mere* belief is necessary. We take that version up in the next section. The stronger version of the weak requirement that only belief, not knowledge, is necessary is the claim that a *justified* belief is necessary. We take that up here.

In the counterexample of the previous section the user's belief that the referent of the citation existed was, given the evidence, quite reasonable. We might say that it was justified, warranted, appropriate, etc. It may be thought that this is in part what supported our intuition that despite the belief's being false the case was still a case of known-item seeking.

But is that so? Suppose our user is, at least on this one occasion, being just a bit careless and overly optimistic. He forms a belief that is not quite warranted given the evidence (although perhaps almost warranted) to the effect that a bibliographic item exists — and sets about attempting to find it. Would this be a known-item search? Here again it seems that, at least in a sense, the user's cognitive state is phenomenologically similar to that of a user who is indeed performing an authentic known-item search. Can simply a slight misjudgment of the weight of the evidence make the difference between whether a search is a known-item search or not?

5.5. Is Any Belief Required?

If the preceding consideration is persuasive we might retreat to the claim that some belief in the object's existence is necessary, but not insist on justified belief. We now examine this weak version of the weak knowledge requirement.

Consider a case when a user hears about an unlikely entity from a colleague. He conducts a search for that entity although he does not believe that it can possibly exist, and he wants to prove that his colleague is wrong. In such a case, is he still conducting a known-item search even though he does not even believe it exists? At least some researchers would say so, and regardless of whether the item exists.

For "known item search" is in particular so-called descriptive cataloging data useful. Bibliographical verification is a process in which libraries search known items. (Documents to be verified sometimes turn out to be phantoms, cf., Dubin, 2004). (Hjørland, 2005b)

Apparently the possibility of known-item searching where the user has no belief whatsoever that the object sought exists is consistent with the intuition of some LIS researchers.

5.6. Nature of Object Sought

Now we turn to another sort of issue: exactly what sort of object is being sought? As a source of useful terminology for our discussion we draw on IFLA's *Functional Requirements for Bibliographic Records* (FRBR), and, specifically, the FRBR Group 1 entities: *work*, *expression*, *manifestation*, and *item*. Works are characterized by FRBR as "a distinct intellectual or artistic creation", expressions as "the intellectual or artistic realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms" (a rough synonym might be text), manifestation as "the physical embodiment of an expression" (an edition is a manifestation), and items as "a single exemplar of a manifestation" (items are individual physical copies) (IFLA, 1998). On this account works, expressions, and manifestations are abstract entities, and items are concrete physical entities. Works are *realized* by expressions, expressions *embodied* in manifestations; and manifestations *exemplified* by items — and works are typically realized by multiple expressions, expressions typically embodied in multiple manifestations, and manifestations typically exemplified by multiple items.

Since the term "item" already appears in the phrases associated with our concept of interest (known-*item* searching), we will to avoid confusion often use the term "copy" to refer to the FRBR entity "item", and we will more

frequently than elsewhere use the term “object”, rather than “item” to refer generally to the thing sought.

For now, we assume a user really intends to come into actual possession of an object, and not simply learn that the object exists in a particular place, such as the library where the search is being conducted. If we further assume the strict identity of the object that is sought and the object that comes to be possessed in the case of a successfully concluded search, it might seem that only a copy (a FRBR “item”) will do as the object sought — as only the copy is a physical object and could be actually possessed by a user.

But this cannot be the whole story. Users in a rare book library may indeed sometimes wish to find a particular copy, but others are more likely looking for a particular edition, or a particular translation, or a particular work. Relevant is Yee and Layne’s revision of “known-item search” to “known-work search”:

Known-Work Search

Research has shown that one of the most common searches done by our users is a *known-item search* (actually, a search for a particular work, or a *known-work search*). (Yee & Layne, 1998; p.74)

Indeed it is a commonplace of library science that users sometimes wish to possess a particular edition, sometimes an edition with a particular translation, and sometimes any copy of a work. How do we describe these cases? The obvious start is to observe that although phrases like “looking for a particular edition” are natural enough, “looking for a copy of a particular edition” is equally natural and better reflects our intuition that only coming into possession of a physical copy would be a successful search. Now we may feel we can describe our different cases with somewhat more ontological clarity:

- (a) ...searching for a particular copy
(e.g., one desired for its scribal marginalia, provenance, or the passport used as a bookmark and forgotten).
- (b) ...searching for a copy which exemplifies a particular manifestation
(e.g., the 1851 NY Scribner’s edition).
- (c) ...searching for a copy which exemplifies any manifestation that embodies a particular expression
(e.g., say the emended text of the 1851 edition).

- (d) ...searching for a copy which exemplifies any manifestation that embodies any expression of a particular work
(e.g., *Moby Dick*).

(We use the FRBR categories only to show how a solution might go, remaining agnostic, here, as to whether or not FRBR has the framework just right or not.)

There may be additional ambiguity beyond the FRBR distinctions in the sought object of a known-item search – is it a specific object or the object as the source of a specific answer? The assumption that a single, correct answer/result exists for a known-item search is found in a number of articles, especially related to evaluation of information retrieval systems. This blurs the level of granularity: (i) item as answer, (ii) item that contains an answer.

...information needs are clearly defined and have a single, correct answer. (Rosenfeld and Morville, 1998; p.102)

...usually only one "right" result (up to syntactic/semantic aliases). (Broder, 2002; p.5)

...a search for a known item may carry with it the expectation that there is a "correct answer" and as soon as the searcher finds an item that appears "correct," the search is terminated. (Dalrymple, 1990; p.279)

5.7. The De Re Attribution Assumption

However the similarity of phrasing in each case above conceals an interesting difference. In case (a) it is natural to say about the copy in question that *it* is being sought. However in case (b) it does not seem natural to say about any one of the many copies that *it* is being sought. After all, which one would it be? And it can't be all of them since our user wants just one copy. If this is right, then that someone "seeks a copy which exemplifies a particular manifestation" does not imply that there actually is a copy which is such that it is being sought.

Does that mean that in cases (b) through (d) there is something else other than a copy such that *it* is being sought? For instance, the manifestation, expression, and work, respectively for (b)-(d)? Not given our assumptions that (i) what is sought is what is identical with what is found (when successful) and (ii) what is found is always something concrete and physical. We must now either revisit those assumptions or give up the intuition that in

at least some known-item searches of types (b)-(d) there is something about which we can say plausibly “it is sought”.

It might be thought that this is not a problem if we agree that for the reasons argued earlier existence is not a requirement. However it seems nevertheless plausible that even if existence is not a general requirement, still, in at least those cases of (b)-(d) searching where full knowledge of the sought object actually does obtain, there must then be something about which we can say: it is sought.

In favor of holding that in no case of known-item searching (b)-(d) is there anything such that it is sought, it will inevitably be adduced that Diogenes sought an honest man, Schliemann the site of Troy, and Ponce de Leon the Fountain of Youth — and in none of these cases is there something about which we could say “*it is sought...*”. But the obvious rejoinder is that even if those are cases of known-item searching (which is doubtful), they are not relevant cases as they are not cases where something *known* to exist was sought. Rather each searcher only thought that the thing they sought existed. If we wish to retain any notion that in some cases of (b)-(d) known-item searching there is truly something about which we can say “*it is sought*” then we need to solve this problem. If we don’t it will not be clear just what it means to say even in the case of a known-item search where the knowledge condition obtains that “a known item is sought”.

Any complete treatment of “known-item” searching will need to deal with this difficult problem. However we will not discuss it further here. It is a familiar problem in philosophical logic and our examples of seeking nonexistent objects are taken from that literature.

5.8. Location of the Sought Object

Another assumption which is occasionally mentioned in some characterizations of a known-item search is that the user knows the location of the sought object. An expression such as “a known haystack” implies that the user knows *where* to look in order to find the sought object.

...you know exactly what you want, that the information exists, and where it can be found. (Rosenfeld & Morville, 1998; p.102)

A close analogy of the known item search is Koll’s “A known needle in a known haystack” (Koll, 2000). (Chu, 2003; pp.82-83)

It is interesting though, that we found the exact opposite characterization, in other words, not knowing where the sought object is as well in some literature.

...the user knows of a particular document but does not know where it is. (Ogilvie & Callan, 2003; p.143)

This is an issue more apparent in literature related to web searches or information retrieval, rather than catalog or OPAC use studies.

5.9. The Information Known and Used for the Search

5.9.1. Types of Information

In the 1876 edition of Cutter's *Rules for a Printed Dictionary Catalogue* (p.10), we find that the first objective of a catalog is:

1. To enable a person to find a book of which either
 - (A) the author
 - (B) the title
 - (C) the subject } is known.

It is interesting that an interpretation of this allows subject as an access point for known-item searching. However in the later literature, "title" and "author" dominate as the major attributes used for conducting a known-item search, and the mention of "subject" as an access point becomes harder to find. In some cases, known-item searches are even considered to be equal to the aggregation of "title" and "author" searches (Cooper & Chen, 2001).

...a search for some item for which either the author or title is known (called a known-item search) (American Library Association, 1958; Hafter, 1979). (Baker & Lancaster, 1991; p.200)

Known item search is generally conducted based on a known document attribute such as author and title (Koll, 2000). (Chu, 2003; p.83)

A few authors do consider other attributes such as publisher, series, or subject as the types of information used for known-item searches.

...generally accompanied by descriptive bibliographic information such as author (personal, corporate, or institutional), title, publisher, etc. (Swanson, 1972; p.302)

A known-item search may include the author, the title, the subject, or a combination of these and other pieces of information to identify the item in the catalog. (Wildemuth & O'Neill, 1995; p.265)

...searching for documents that are known to the user (known item search) or documents for which he or she already has some information, for instance, author, title, publisher, or series, which has nothing to do with the content of the document. (Hjørland, 1997; p.21)

The last characterization by Hjørland raises a question though: does the title of the document have nothing to do with the content of the document?

5.9.2. Completeness of Information

The amount of information that the user needs to know is often not specified, but it seems that partial information also counts as well as complete citation information.

...expects the user to be searching for a specific item of which he or she knows at least a part of the name. (Brinkley & Burke, 1995; p.4)

...searching for one specific site when at least part of the title is known. (DeZelar-Tiedman, 1997; p.8)

Then again, we find interesting characterizations such as:

It was found that 56% of the “original known-item searches” (searches in which the user had not obtained any information from the card catalog) were successful. (Specht, 1980; p.335)

If such is the case, mustn't we assume that the user needs to have complete title and author information before the search?

5.9.3. Accuracy of Information

Several studies on catalog uses found that the user's information was often incorrect and incomplete (Jackson & American Library Association, 1958; Swanson, 1972; Lewis, 1987; Dwyer et al., 1991). Thus, it seems reasonable to say that the complete accuracy of information that the user brings to the search should not be a necessary condition for a known-item search, in fact, for any kind of search. If so, in an extreme case, if the user is looking for a book he previously read, but *all* the information he remembers about the item is incorrect, can we say that the user really *knows* the item? The user may not be able to easily initiate the search in the existing system, but he

may be able to recognize the item and verify it to be the correct item when he sees it. Except the fact that he does not really know any correct information about the item to use as access points, other conditions (i.e., particular item in mind, awareness of the existence of the item due to previous contact with it) seem to suggest that it is most likely to be a known-item search. Does this mean, then, that one can conduct a known-item search without really knowing the attributes of the item?

The second most frequent cause of failure to find an item owned by the University Libraries was the use of an incorrect citation (Dwyer, Gossen, & Martin, 1991; p.231)

The information patrons bring to catalogs, regardless of form, is often incomplete or partially incorrect. (Lewis, 1987; p.156)

A known-item search occurs when the user has a limited but correct description of an existing document. The user is sure of the fact that the document exists, that its title and author are explicitly stated somewhere in the document, and these assumptions are true to the actual state of the docuverse. This situation is fairly different from when a user searches for unknown documents. (Dahlstrom & Gunnarsson, 2000)

Dahlstrom and Gunnarsson's characterization of "the user has a limited but correct description of an existing document" lets us avoid having an unreasonable assumption that all of the user's information must be correct. It is still questionable though, what we can say about the case in which the user is looking for a book he previously read, but *all* the information he remembers about the item is incorrect. Can we still say that this is a known-item search?

5.10. Relationship with Other Types of Searches

Traditionally, there is a strong dichotomous view of distinguishing between known-item searches and subject searches (Buckland, 1979). However, the blurriness of this boundary is mentioned by several authors and noted in various studies.

...even searches which began as known-item searches became subject searches, the known-item being the user's access point to that subject (Hancock-Beaulieu, 1990). (Brinkley & Burke, 1995; p.4)

In other words a “known item” search may, in fact, be an indirect and disguised “subject” search for specific information not necessarily unique to the document used. (Buckland, 1979; p.145)

...however, some known-item searches are subject searches with the known item used as an entry into a subject area. (Lewis, 1987; p.153)

Buckland’s comment that a known-item search “may, in fact be an indirect and disguised ‘subject’ search” is now a common observation (also found in Lewis, 1987; Brinkley & Burke, 1995) alluding to situations such as these: (i) the user looks for an item known to exist in order to get the subject descriptors; (ii) the user searches on keywords in the title field to locate items on a subject; or (iii) the user searches on an author anticipating, but not necessarily knowing that that author has written books on a particular subject. Given our agenda however we need to ask what the statement “may in fact, be an indirect and disguised ‘subject’ search” is intended to claim. The rhetoric of “disguise” suggests that something is not what it seems, but this does not necessarily mean that there is no known-item search in the cases in question. In case (i) for instance it might be best to say that there is a known-item search taking place which is part of a subject search. Cases (ii) and (iii) nicely illustrate the problem of managing conceptual and operational definitions. If we understand known-item searching operationally as a search using author or title fields then we cannot say that there is any disguise: the searches are known-items searches — the definition however is a bad one because it is not in the relevant context coextensive with our conceptual understanding. Cited reference searching is also an interesting example where the search for a known item (a particular cited reference) is often interpreted as a form of subject searching because the contents of citing articles are likely to be related to the cited article in some way.

A user searches by subject for the following purposes: (1) to identify or retrieve specific items of which the subject is known (sometimes referred to as known-item searching) and (2) to search for information on a given subject. (Chan, 1994; p.155)

Searching for known items by subject is very inefficient, but can be successful when other approaches fail. (Lewis, 1987; p. 153)

Here, we again see a recognition that some known-item searching may be carried out by searching under subject. We might say, in fact, that just as some known-item searching disguises subject searching, *some subject*

searching disguises known-item searching. Buckland (1979) also presents a scenario of using the subject catalog when the author and title are not remembered by the user to illustrate such a case. What is important for our discussion here though is that in neither case does the search, or at least part of it, fail to be a known-item search.

In some articles related to information retrieval system evaluation, authors started to consider navigational queries as closely related to (i.e., part of, equal to, an example of) known-item searches.

Navigational queries.

This type of search is sometimes referred as "known item" search in classical IR, but is mostly used in the evaluation of various systems. (Broder, 2002; p.5)

Lastly, various OPAC studies concluded that known-item searches are generally more successful than subject searches. This may explain why known-item searches have not received as much attention as subject searches in LIS research.

Known-item searches tended to be more successful than subject searches. (Lewis, 1987; p. 153)

Moore found that known-item searches generally were more successful than subject searches in online catalogs. (Dwyer, Gossen, & Martin, 1991; p.229)

It seems necessary to be cautious in accepting these statements, or any kind of assertion/conclusion drawn from empirical studies on known-item searches as the operational definitions of the concept vary considerably, as we saw in the above examples.

6. Discussion and Areas for Future Research

Through this paper, we attempted to demonstrate the complexity of the issues in defining a concept of "known-item search" and explicitly spell out the assumptions in various characterizations of the concept in the LIS literature. We see this paper as an effort to initiate a discussion rather than to settle a question, and we hope to have successfully raised the readers' awareness of this issue.

We note that this paper is only the first step in a larger project. To truly make substantial progress on the agenda of increased rigor in the foundation of information science, the results that we express informally in this paper must be eventually re-expressed in the common coin of foundational agendas: a simple prose that could uncontroversially be translated into a formally defined logical notation. That final development will be carried out in further work.

We also plan to explore more empirical studies in different contexts (web, OPAC, information retrieval) and look into the case of searching for non-textual items such as audio/music objects. The nature of encountering music objects is different from textual documents, and therefore it is not difficult to imagine a situation where the user gets to know the existence of a specific piece of music by hearing it but not the properties that can help them initiate the search. The user may recall other kinds of information about the sought object (e.g., has lyrics “so and so”, fast beat, sad and moody, used in a car commercial, sounds like artist A), but many of these clues are not represented in bibliographic records, thus cannot be used in an OPAC for searching a music object. In this case, the users simply do not have *enough* or the *right kind* of information to initiate the search in a typical OPAC, thus those queries, although they exist, may never appear in the search log. In our future work, we hope to elaborate the notion of a known-item search such that we can explain such cases in our framework.

7. References

- Allen, B. (1989). Recall cues in known-item retrieval. *Journal of the American Society for Information Science*, 40(4), 246-252.
- Baker, S. L., & Lancaster, F. W. (1991). *The measurement and evaluation of library services* (2nd ed.). Arlington, VA: Information Resources Press.
- Bates, M. J. (1998). Indexing and access for digital libraries and the Internet: human, database, and domain factors. *Journal of the American Society for Information Science*, 49(13), 1185-1205.
- Brinkley, M., & Burke, M. (1995). Information retrieval from the Internet: an evaluation of the tools. *Internet Research*, 5(3), 3-10.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2), 3-10.

- Buckland, M. K. (1979). On types of search and the allocation of library resources. *Journal of the American Society for Information Science*, 30(3), 143-147.
- Chan, L. M. (1994). *Cataloging and classification : an introduction* (2nd ed.). New York: McGraw-Hill.
- Chu, H. (2003). *Information representation and retrieval in the digital age*. Medford, NJ: Published for the American Society for Information Science and Technology by Information Today.
- Cooper, M. D., & Chen, H.-M. (2001). Predicting the relevance of a library catalog search. *Journal of the American Society for Information Science and Technology*, 52(10), 813-827.
- Cutter, C. A. (1876). *Rules for a Printed Dictionary Catalogue*. Washington: G.P.O.
- Dahlstrom, M., & Gunnarsson, M. (2000). Document architecture draws a circle: on document architecture and its relation to library and information science education and research. *Information Research*, 5(2).
- 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*, 41(4), 272-281.
- DeZelar-Tiedman, C. (1997). Known-item searching on the World Wide Web. *Internet Reference Services Quarterly*, 2(1), 5-14.
- Dwyer, C. M., Gossen, E. A., & Martin, L. M. (1991). Known-item search failure in an OPAC. *RQ*, 31(2), 228-236.
- Ellis, D. (1989). A behavioral approach to information retrieval system design. *Journal of Documentation*, 45(3), 171-212.
- Hjørland, B. (1997). *Information seeking and subject representation : an activity-theoretical approach to information science*. Westport, CT: Greenwood Press.
- Hjørland, B. (2005a). Library and information science and the philosophy of science. *Journal of Documentation*, 61(1), 5-10.
- Hjørland, B. (2005b). Core Concepts in Library and Information Science. Retrieved January 20, 2006, from <http://www.db.dk/bh/Core Concepts>

[in LIS/home.htm](#)

- IFLA Study Group on the Functional Requirements for Bibliographic Records., & International Federation of Library Associations and Institutions. Section on Cataloguing. Standing Committee. (1998). *Functional requirements for bibliographic records : final report*. München: K.G. Saur.
- Jackson, S. L., & American Library Association. Cataloging and Classification Section. Policy and Research Committee. (1958). *Catalog use study; director's report*. Chicago: American Library Association.
- Kochen, M. (1974). *Principles of information retrieval*. Los Angeles: Melville Pub. Co.
- Koll, M. B. (2000). Track 3: information retrieval [report from the 1999 ASIS Annual Meeting]. *Bulletin of the American Society for Information Science*, 26(2), 16-18.
- Lancaster, F. W., & Joncich, M. J. (1977). *The measurement and evaluation of library services*. Washington: Information Resources Press.
- Lewis, D. W. (1987). Research on the use of online catalogs and its implications for library practice. *Journal of Academic Librarianship*, 13(3), 152-156.
- Lipetz, B. A. (1972). Catalog use in a large research library. *Library Quarterly*, 42(1), 129-139.
- Matthews, J. R., Lawrence, G. S., Ferguson, D. K., & Council on Library Resources. (1983). *Using online catalogs : a nationwide survey : a report of a study sponsored by the Council on Library Resources*. New York, NY: Neal-Schuman.
- McKenzie, P. J. (2003). A model of information practices in accounts of everyday-life information seeking. *Journal of Documentation*, 59(1), 19-40.
- Ogilvie, P., & Callan, J. (2003). Combining document representations for known-item search. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 143-150). Toronto, Canada: ACM Press.
- Rosenfeld, L., & Morville, P. (1998). *Information architecture for the World*

- Wide Web* (1st ed.). Cambridge ; Sebastopol, CA: O'Reilly.
- Slone, D. J. (2000). Encounters with the OPAC: on-line searching in public libraries. *Journal of the American Society for Information Science*, 51(8), 757-773.
- Specht, J. (1980). Patron use of an online circulation system in known item searching. *Journal of the American Society for Information Science*, 31(5), 335-346.
- Stielow, F., & Tibbo, H. (1988). The negative search, online reference and the humanities: a critical essay in library literature. *RQ*, 27(3), 358-365.
- Swanson, D. R., Vaughan, D. K., & Cooper, W. S. (1968). *Requirements study for future catalogs*. Progress report no. 2. Chicago: University of Chicago.
- Swanson, D. R. (1972). Requirements study for future catalogs. *Library Quarterly*, 42(3), 302-315.
- Voorhees, E. M. (2004). Overview of TREC 2004. In E. M. Voorhees & L. P. Buckland (Eds.), *NIST Special Publication 500-261: The Thirteenth Text REtrieval Conference Proceedings (TREC 2004)* (pp. 1-12). Gaithersburg, MD: National Institute of Standards and Technology (NIST).
- Wildemuth, B. M., & O'Neill, A. L. (1995). The 'known' in known-item searches: empirical support for user-centered design. *College and Research Libraries*, 56(3), 265-281.
- Yee, M. M., & Layne, S. S. (1998). *Improving online public access catalogs*. Chicago: American Library Association.