FEATURE

Known Item Search

Theoretical and Practical Considerations

Birger Hjørland

This article looks at the concept "known item search" (KIS) and considers it in relation to library practices. The author critically examines previous research on KIS and argues that the concept is important because it is categorically different from "subject search" and because it is assumed in processes such as bibliographic verification and descriptive cataloging. The article further discusses which kinds of metadata best serve KIS and argues that the traditional distinction between descriptive cataloging and subject cataloging is a fruitful point of departure for describing the metadata needed for respectively KIS and subject searches.

Introduction

Known item search (KIS), the search for a particular thing or document, is a common activity performed in library catalogs, in bibliographies, databases, and search engines, as well as in everyday life (e.g., finding a song in which just a fragment is remembered).[1] It is a concept that seems easy to understand, and it is often regarded as a rather trivial problem in library and information science (LIS), where the main focus has been subject searching (also termed "topic searching").[2]

Because KIS puts other demands on both search systems (including document descriptions) and on search strategies, however, it is an important concept in its own right. Min-Yen Kan and Danny C. C. Poo wrote:

> "How important is known item search? In the setting of an OPAC [online public access catalog], it is very important. Larson [3] points at the long term decline of subject searching in OPACs, in which known item search accounts for a growing proportion of library catalog searches, up to 50%. However, supporting these types of searches has largely been ignored by the information retrieval community, whose focus has been on topical search (e.g., TREC bakeoff competitions [4]). While these efforts have improved the state of the art for topical search, we see a need to support better known item query detection and retrieval."[5]

This quote is not meant to claim that KIS is more important than subject searches but rather to illustrate that KIS is important enough to deserve special attention in LIS.

In library science, the purposes of library catalogs have been discussed since Charles A. Cutter, who in 1876 presented the following "objects":[6]

1. To enable a person to find a book of which either:
   (a) the author is known
   (b) the title is known
   (c) the subject is known

Birger Hjørland (birger.hjorland@hum.ku.dk) is Professor Emeritus in the Department of Communication at the University of Copenhagen, http://orcid.org/0000-0001-9535-1732.

Library Resources & Technical Services | June 2025
https://doi.org/10.5860/lrts.69n3.8497

2. To show what the library has:
    (d) by a given author
    (e) on a given subject
    (f) in a given kind of literature
3. To assist in the choice of a book:
    (g) as to its edition (bibliographically)
    (h) as to its character (literary or topical)

The first of Cutter's objectives is about enabling KIS, while 2 (e) is about subject searching. Jin Ha Lee, Allen Renear, and Linda C. Smith discussed these rules from the perspective of KIS, writing:

> 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 search. 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).[7] However, a few authors do consider other attributes such as publisher (Swanson, 1972;[8] Hjørland, 1997),[9] series (Hjørland, 1997), subject (Wildemuth & O'Neill, 1995)[10] as the types of information used for known-item searches.[11]

In spite of Cutter's explicit mention of subject searches, library and information scientist Pauline A. Cochrane described what she considered a paradigm shift in library science:

> "Common wisdom since Cutter's time has been that most users of the library want a catalog where they can find a particular item, a known item."[12]

Although this example may misinterpret Cutter, it is reflective of a tendency in library cataloging theory and practice to prioritize descriptive cataloging. The opposite tends to be practiced in scientific subject databases such as Chemical Abstracts, MEDLINE, PsycINFO, and Web of Science, which tend to have inferior descriptive data—for example, to have less author authority data, or, as with the Web of Science, not to provide original titles for non-English articles. Such descriptive elements are better developed in library cataloging, but, as argued by Birger Hjørland, this represents a problematic tendency in library cataloging practice to prioritize KIS, confirming Cochrane's view.[13]

The priorities between KIS and subject searching have therefore not been uniform in LIS, and different kinds of searches have not always been clearly distinguished in objectives for document representation. Important research has been carried out in relation to KIS, and the main purpose of this article is to address the main conceptual and theoretical issues of this topic.

Definitions and Meaning of "KIS"

Lee, Renear, and Smith found it surprising that despite its central status in LIS over a long period, the concept "KIS" has received practically no systematic discussion.[14] The authors further demonstrated "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." The article pointed out that it is not required that the searcher "know" the searched document to exist or that it really exists, because a bibliographical reference can be to a nonexistent document (a so-called "bibliographical ghost"). A famous example of such a "ghost" is a highly cited, non-existing paper by information scientist Gerald Salton, which he never wrote.[15] The term "known item" has therefore misleading associations. A person may be searching for a document, on which very little is remembered or known with a reasonable degree of certainty, and which may even have a doubtful existence. Therefore, when information scientist Michael Buckland wrote, "Known items generally have names, addresses, or distinguishing physical features," this is only true when the searcher knows some of these characteristics.[16] There are easy as well as difficult cases of KISs, and in the difficult cases, such characteristics may be unknown.

Dictionary for Library and Information Science defined:

> "[K]nown-item search: A search in a library for a specific work, as opposed to a search for any work by a known author or for works on a particular subject."[17]

ISO 5127 defined:

> "[K]nown item retrieval: *search and retrieval* (3.10.2.01) for a specific item present in the searcher's mind from the start on" [18]

Buckland suggested:

> "Known item search is ordinarily understood to mean a search where the searcher has a specific item in mind and either has an address for it or else believes (or hopes) that sufficient clues, such as author surname and/or title words, will enable that particular document to be found. It is, in effect, a citation search with, commonly, an incomplete or uncertain citation."[19]

Buckland also wrote that the opposite of "known item search" is a "not-known item search," writing:

> Known item search is traditionally distinguished from subject search. Strictly, this is an incomplete view because the logical complement of a known item search has to be a not-known item search.[20] Subject search is a common kind of not-known item search in a library context, but it is not the only kind. The many other possibilities include, for example, searches by genre, bestsellers, and banned books.[21]

In a library or other bibliographic context known item searches are often not, in fact, for a particular known item but, more loosely, for any instance of a particular known edition or of an instance of

any edition of a particular known title. This is a departure from the pure case of search for a unique, particular document.[22]

In his article, Buckland elaborated on the distinction between "particulars" and "specimens." A particular document may be an exemplar (i.e., an individual copy) of a book with notes or marked passages, whereas a specimen may be any copy of a given edition of a book.[23] Normally just specimens are meant in relation to KIS, but in some instances, the user may want a particular, unique document. In this connection, Lee, Renear, and Smith brought the attention to IFLA's Functional Requirements for Bibliographic Records (FRBR), which later developed to IFLA *Library Reference Model* (LRM),[24] which provides a model describing the relations between "work," "expression," "manifestation," and "item."[25] FRBR/LRM thus enable users to distinguish four kinds of known items. In the example of Shakespeare's *Hamlet*, the known item may mean (1) work: the tragedy by that title written by Shakespeare; (2) Expression: a Danish translation of *Hamlet*; (3) Manifestation: a particular Danish edition of Hamlet; (4) Item: a specific exemplar of a particular Danish edition of Hamlet (in libraries that have multiple copies of the same book, "items" are often given an individual barcode or RFID [radio frequency identification] tag).

What then, is a KIS? We saw above that Buckland suggested it to mean a search where the searcher has a specific item in mind and has either an address or clues expected to be sufficient to find it.[26] If having "a specific item in mind" includes having a bibliographic reference (whether sufficient or insufficient, and whether a ghost or not), then this part of Buckland's definition seems okay. The other part of the definition—"has either an address or clues expected to be sufficient to find it"—excludes searching for something without any idea of whether one's clues are sufficient to find it, which, however, is probably often the case (e.g., when a name is on the "tip of the tongue" [ToT], as with example number 5 below).[27]

Examples and Characteristics of KIS

(1) A researcher looks up a bibliographical reference in an OPAC to order and to obtain a copy of the document referred to. Normally a basic set of essential bibliographic data is considered essential in both bibliographical references and in OPACs—for example, author's last name, publication year, title of book or journal, publisher's name (in the case of books), and volume, first page, and article title (in the case of journal articles) and, when available, the digital object identifier (DOI).[28] Looking up a combination of a few of these essential data will, if the library has or provides access to the document, in the overwhelming number of cases lead to an unproblematic match, and the document can be ordered. Frederic G. Kilgour, for example, prescribes how informed library users can issue effective known item queries by including the author's surname and specific words from the title of the item.[29] Normally, the redundancy in the bibliographical data (both the user's reference and the library's OPAC) is high, meaning that if the first attempt fails, then another combination of such essential data will probably succeed.

(2) If the document is not found as described in point (1), it may be (a) because there are errors in the user's bibliographical data, (b) because there are errors in the library's bibliographical data, (c) because the library does not have the document, or (d) because the reference is a bibliographical ghost.[30]

Normally the best procedure is first to verify the user's data (e.g., by searching in Google, WorldCat, or other more comprehensive bibliographical databases). If verification succeeds, the document can be ordered, or if it is not in the library, an interlibrary loan may be requested. A number of special problems exist, including:

a. A common problem in KIS is spelling errors (in user queries or in databases, e.g., when input is made by optical character recognition) or spelling variations (e.g., "color" and "colour"). Now that search engines support fuzzy searching and approximate string matching, such errors have been reduced in search engines and some databases. Unfortunately, many OPACs have not yet successfully addressed spelling search errors.

b. Another example is due to ambiguities in printing years. Publishers have a tendency to provide the wrong publication date for their book (e.g., books published in the fall are given the following year's publication date).[31]

(3) In some cases, the user's reference is very problematic. This author got the following reference from ChatGPT-4o and has not been able to verify it: "Sayre, Kenneth M. 'Cybernetics and the Philosophy of Mind: The Neglected MacKay-McCulloch Exchange.' *Kybernetes*, vol. 38, no. 9, 2009, pp. 1539-1555." The journal *Kybernetes* exists, the vol. issue and year match each other in the existing journal, but the article is not on the specified pages. There exists a book by Sayre, *Cybernetics and the Philosophy of Mind*, but without the subtitle and not containing the demanded information about MacKey. Google searches of both "The Neglected MacKay-McCulloch Exchange" and "MacKay-McCulloch Exchange" returned zero hits. A return to ChatGPT-4o, indicating the error and asking for a correction, provided the same reference, but now in issue 7/8, which is also wrong.

This example shows the user's options in difficult cases: systematically use every bit of information in the given reference (and systematically exclude every other bit of information) in order to verify the reference and try to obtain further information about the sought item.[32] If the reference contains a reasonable amount of bibliographical data, and it cannot be verified, then it must be considered a bibliographical ghost; if it does not contain a reasonable amount of bibliographical data, then it must be considered lost (at the least for now).

(4) KISs are not always initiated by bibliographical references, but also user memory about documents, or about some contents of a document. For example, one may have heard about "the 20 percent rule" (or was it "the 25% rule"?) in library classification and expect it to be part of a library's written guidelines and therefore search for it. Or, one may have a more or less vague memory from prior reading about some information that could be relevant for a present argument and try to recall where this was read and then to retrieve that document. In such cases, the KIS resembles a subject search, in which the remembered information is used as input and search criteria, and there is the possibility that more than one document fulfils the user's need.

(5) The phenomenon known as "tip of the tongue" (ToT) is relevant to many cases of KIS: The searcher cannot remember the relevant term for something but has a partial memory of it and a feeling that it is

likely to be remembered soon. ToT is studied in many fields, particularly in psychology.[33] There are very few studies relating this term to document searches in databases, however, and these seem not clearly to distinguish the general failing to retrieve documents from the cases with the feeling to be *almost* able to remember the term needed to retrieve a document.[34] ToT has also been used to discuss recall of non-textual items (e.g., music).

(6) Some examples are due to library users' lack of knowledge about the library catalog. Catherine M. Dwyer et al. found that more problems were associated with periodical articles than with monographs.[35] For example, many requests were based on article titles when journal titles should be used. This issue is related to the problem that some documents (e.g., *Educational Resources Information Center* documents) are not cataloged in the OPAC, but rather identified in a separate database, even if the library holds them.[36]

Based on a small set of queries, Kan and Poo provided some general characteristics of KISs (as opposed to subject searches):

- They are longer and often copied from a syllabus or a web search.
- They contain determiners: In English titles, determiners (such as "the," "an," and "a") are often parts of book titles and are thus also prevalent in known item queries. In contrast, most area or unknown item searchers do not type determiners into search boxes as many know that such words are often ignored by OPACs.
- They contain proper nouns, including names of authors and editors and names of things that may appear in document titles.
- They contain mixed case—for example, exactly matching a title's orthographic case (whether or not the OPAC is case insensitive).
- They contain certain advanced operators, such as specifying terms for the author and the title fields.
- They contain keywords such as "journal," "course," and "textbook." These usually connote the desired type of resource, rather than a keyword search for the word. Similarly, many titles in libraries but few subject headings consist of these words.[37]

These characteristics of KIS are, as already noted, based on a small sample of requests. However, even if studies are performed on large samples, such characteristics will only be indicative: some KISs may not conform to certain rules or statistical patterns. Nonetheless, they are important because, as suggested by Kan and Poo, they may provide a basis for improved search interfaces that may be helpful for users.

The Functions of the KIS Concept

Bibliographic Verification (or "Bibliographical Validation")

One function of the term "known item search" relates to the concept "bibliographical verification." In libraries, bookstores, and databases, many requests for documents contain errors and therefore cannot easily be found. Bibliographic verification is admittedly easier in the online catalog compared to the

card catalog, but this is just a question of degree, not of a categorical difference. If, for example, the author or title in a search or request is wrong or misspelled, a first conclusion may be that the required document is not in the library. Rather than providing this answer to the users, the library may start a verification process examining the request for errors (or examine if the reference is a "bibliographic ghost"), correcting the errors, and obtaining the document (if not from the library's own stacks, then potentially from an interlibrary loan). Bibliographical verification is the process of confirming the accuracy and completeness of bibliographic information for a given source. This involves checking details of a basic set of essential bibliographic data—such as author name, title, and publication date—and thereby verifying or falsifying the existence of a document about which such data have been given. The staff working with this task in large libraries used to be trained in bibliographical verification (often in relation to interlibrary loan), and a textbook has been written on this (in Danish).[38] The verification process was often an algorithmic procedure based on national bibliographies, catalogs from large libraries such as Library of Congress, and other bibliographical tools. The point here is that such verification processes are KISs and that they are very different from subject searches performed in libraries, such as helping students and researchers find books, articles, and other documents for their theses and papers.

The above is written in the past tense because today libraries no longer tend to perform verification processes in the same formalized ways.[39] This does not make KISs and bibliographical verification needless concepts, however, because it is still important to distinguish them from subject searches in order to optimize both kinds of search processes.

An important implication of this issue of verification is the need for researchers and students to know about essential bibliographic data. These data are required for readers to obtain the documents to which the references refer. This is often done by teaching a specific referencing style or standard, for example, the *Chicago Manual of Style* or the "ANSI/NISO Z39.29-2005 standard.[40] Such styles develop over time. For example, today it is mostly required that references to journal articles include the article's DOI, which has contributed to facilitating KIS (also in OPACs when these are integrated with discovery services that support DOI searching.)

"Descriptive Cataloging" Versus "Subject Cataloging"

The dichotomy between KIS and subject search is related to the dichotomy between descriptive cataloging and subject cataloging. Joan M. Reitz emphasizes the difference between the two last processes in the following definition:

> Descriptive cataloging: The part of the library cataloging process concerned with identifying and describing the physical and bibliographic characteristics of the item, and with determining the name(s) and title(s) to be used as access points in the catalog, but not with the assignment of subject headings and genre/form terms. In the United States, Great Britain, and Canada, descriptive cataloging is governed by Anglo-American Cataloguing Rules (AACR2) [and its successor Resource Description and Access, RDA].[41]

In relation to the part of the library cataloging process concerned with classification and indexing, Reitz defined:

> Subject analysis: Examination of a bibliographic item by a trained subject specialist to determine the most specific subject heading(s) or descriptor(s) that fully describe its content, to serve in the bibliographic record as access points in a subject search of a library catalog, index, abstracting service, or bibliographic database.[42]

One reason for the differentiation between descriptive and subject cataloging is that generalist librarians in major libraries trained in the standards mentioned by Reitz and typically performed the former, while subject specialists typically performed the latter.

Therefore, as reported by Hjørland, large libraries used to have separate departments for descriptive and subject cataloging, staffed with general librarians and subject librarians.[43] A similar separation can also be found in subject bibliographical databases such as MEDLINE, and these two library processes have their parallels in the field of bibliography, where a distinction exists between "descriptive bibliography," which describes documents as physical objects, and "subject bibliography," which compiles and characterizes documents, emphasizing their subject.[44] Descriptive bibliography is primarily based on knowledge about techniques of book production, whereas subject bibliography requires subject knowledge.[45]

It is too simple to say that descriptive cataloging serves KISs while subject cataloging serves subject searches, although overall this is the case. Whereas a subject assignment to a document is generally a bad tool for verification (further described below), many descriptive data are often useful for subject searches (e.g., searches using words from document titles). Nonetheless, KISs and subject searches make different demands regarding the prioritization of metadata, and this implies that KIS is a concept that requires its own aim to be considered in developing bibliographic databases.

Kinds of Metadata Suited for KIS

The author has already presented the concept of essential bibliographical data for KIS in the first of the examples of KIS. The present section focuses on discussing three dichotomies suggested by Michael Buckland for understanding KIS, and it ends with an overall conclusion about metadata suited for KIS.[46]

Terms for "Individual Concepts" Versus "General Concepts"

Buckland discusses the relation between KIS versus subject search on the one hand and individual concepts versus general concepts on the other.[47] Concerning individual concepts, Buckland, citing indexing theorist Robert Fugmann, wrote: "'individual concepts' are persons, institutions, and towns, all with proper names, and which occur in single or very few instances."[48] Buckland seems to suggest that individual concepts somehow correspond to, or are appropriate for, supporting KISs. Before we discuss this, it can be mentioned that individual concepts (e.g., the name of a person), may be indexed by general terms such as "biography,"[49] "anamnesis" (medical history of an individual person), "case

reports," and so forth. These examples demonstrate that general concepts are also developed in order to facilitate communication and retrieval of information about individual concepts considered from different perspectives and interests.

Concerning the use of individual concepts for KISs, Buckland wrote:

> "Fugmann rightly stresses the use of proper names to refer to individual concepts, but proper names may also be used to describe (dispositively). Authors' names are ordinarily associated with known item searches for particular books . . ."[50]

Although it is often true that author names are known when items are sought, this need not be the case, nor is it always the case that other proper names (e.g., journal titles) are known, or any other individual concept for that matter. Many kinds of KIS occurs when authors have a vague memory of a relevant quote they have formerly read and are now trying to retrieve. In such cases, general concepts often are the only available clues.[51]

Referring Versus Describing

Buckland suggested that KIS corresponds to the process of referring, whereas subject search corresponds to the process of describing:

> "The difference between naming what is wanted in a known item search and specifying what is desired in a not-known item search corresponds to the distinction between referring and describing.[52] Referring indicates directly; describing indicates indirectly by specifying characteristics which may in turn indicate appropriate targets. In a traditional digital database one looks up the name of a record of interest in the appropriate table, with possibly a data dictionary to resolve any ambiguity. In a full-text search one searches using descriptors, closely related terms, and vocabulary control which, one hopes, will indicate a small enough set to allow selection of any one or more suitable items without missing other, more suitable items."[53]

Let us exemplify Buckland's claims. In a KIS, author names, journal titles, or specific (combination of) terms may be looked up in order to see if the item searched for can be recognized, possibly after further specifications, and the task thus solved. In a subject search, a combination of terms or other subject access points are looked up to see if the set of items thus retrieved seems relevant and satisfactory in relation to recall and precision.[54] If not, the process continues with modified concepts using so-called "recall devices" and "precision devices" until the task is considered solved. In both cases, what is done is to look up what a certain combination of subject access points are *referring* to. It is difficult to describe information searching as a descriptive process, because the relevant documents are unknown and therefore impossible to describe. It is better to say that the searcher lists a set of terms describing criteria, which the documents must fulfill in order to be relevant.

Kinds of Properties

In KISs versus subject searches, there are no differences in the properties of the documents sought for; in principle, these documents, and therefore their properties, are the same.[55] Differences in properties are not specific to the items sought, but rather in the way the search processes are performed and occasionally in the databases used. In relation to the present article, an important issue to clarify is the nature of the data most relevant for KISs in databases as distinct from those most relevant for subject searches.

Buckland discussed the distinction between material and non-material properties:

> Material properties are the physical attributes, the "brute facts" of a document, such as a title as printed, the author's name as given, and its literal text as well as physical features such as its height, pages, binding, and other objective characteristics. Its non-material properties are any imaginable characteristics other than its material properties, including ownership, topics discussed, point of view, copyright status, genre, and the language of the text.[56]

In this quote, "the author's name as given" is considered a material property of a book, but in table 2 (p. 4), exemplifying the book Bodin's *République* (Paris, 1580), the property that it is authored by Jean Bodin is considered a non-material property. This is somewhat confusing, and here it is suggested instead to distinguish the kinds of data obtained by respectively descriptive and subject cataloging, as described by Reitz above.

- *Data obtained by descriptive cataloging*: The physical and bibliographic characteristics of the item, and the name(s) and title(s) to be used as access points. Other points can be added, such as tables of contents, and, in citation indexes, the reference lists of the documents catalogued.
- *Data obtained by subject analysis*: Assigned classification notations, subject headings, genre/form terms, and notes about the contents.

Although both categories might in some circumstances serve KISs, I shall here argue that data obtained by subject analysis is relatively unhelpful because of the nature of subject analysis. A given subject analysis (and the resulting metadata) represent one individual's view of what the document is about, and we know from inter-indexer consistency studies that inconsistency is an inherent feature of subject indexing, rather than a sporadic anomaly.[57] Whereas there is a fair chance that a person remembers some of a document's physical or bibliographic characteristics, or (parts of) its title or the author's name, the same is not the case with a classification code or a subject heading, which is not a part of the document itself, but is something that somebody has assigned to a bibliographical record. This corresponds to the finding by David W. Lewis: "Searching for known items by subject is very inefficient, but can be successful when other approaches fail."[58]

Our conclusion is that although a basic set of descriptive data (as provided by recognized reference style guides) is often fully adequate, there may be difficult cases for which a broader set of descriptive data are needed, even including subject metadata. We can say, through a modification of a quote by Buckland, "We conclude that we are unable to say confidently of any bibliographical data that

it could not be relevant for KIS."[59] This does not mean, however, that it is impossible to prioritize bibliographical metadata for KIS.

Kinds of Metadata Best Suited for KIS

We have seen that Kilgour prescribed how informed library users can issue effective known item queries by including the author's surname and specific words from the title of the item. Such a simple procedure resolves very large parts of identifying KIS, but not all. We have also considered how scholarly norms of bibliographical referencing—for example, the *Chicago Manual of Style*—prescribe essential sets of metadata, which are meant to guarantee findability of the documents referred to, and we have seen that such norms develop over time and today include the DOI for journal articles. This may be considered the essential knowledge about metadata for KIS. Still, however, there are difficult problems that cannot be solved by such essential sets of metadata. We may fear that these problems will increase because of problems with hallucinations in systems like ChatGPT, as have been exemplified above.

Although we have concluded above "that we are unable to say confidently of any bibliographical data that it could not be relevant for KIS," we have also claimed that this does not mean that it is impossible to prioritize bibliographical metadata for KIS. This becomes, however, much more difficult beyond what is considered the essential set prescribed by referencing norms. It has been argued above that, contrary to Buckland's suggestions, the dichotomies between "individual/general concepts," "referring/describing," and "material/non-material" properties may not be important. The further development of metadata for this purpose may be based on studies of different kinds of documents in a way related to the ways in which documents are studied in the field of descriptive bibliography.[60]

Conclusion

KIS is generally considered the easiest and the most successful kind of document searching in OPACs. Debra J. Slone, for example, wrote that query formulations for KIS seems a natural state for searchers and that 88 percent of searchers were successful.[61] KIS is, however, also a very frequently used kind of search, and some databases, such as WorldCat, are primarily used for KISs.[62] We have claimed that library cataloging—in contrast to scientific bibliographical databases—have prioritized KIS higher than subject searches. However, KIS often encounters greater problems when performed on the web.[63]

Which strategies can be used by the library community to improve KIS?

One point is to reconsider the metadata in library catalogs. Seymour Lubetzky provided the important principle of functional library cataloging in which the purposes, functions, and values of the different kinds of metadata need to be carefully explored.[64] There is a need for updated investigations and considerations for cataloging of all kinds of information resources. More obviously, there is a need to provide techniques such as fuzzy spelling/spell-check techniques, already common in search engines.[65] It seems obvious to focus such efforts on databases such as WorldCat, which are mostly intended and used for KIS.

Kan and Poo provided a set of characteristics that distinguish KIS from subject searches.[66] Based on such characteristics, machine learning, language modeling, and machine translation evaluation techniques were used to automatically identify KIS among other online enquiries. The authors found that this approach has the potential to streamline the interfaces of both OPACs and digital libraries in support of KIS. This too seems to be a way forward.

Acknowledgments

Thanks to the editors, Rachel Scott and Michael Fernandez, and to three anonymous reviewers for careful and detailed suggestions, which have been very helpful for improving the article. Also, thanks to the two anonymous reviewers in the second round of review.

References

1. KIS is also known as KIR (known-item retrieval). The term "navigational queries" for specific documents (as opposed to "informational queries") was used by Madian Khabsa, Zhaohui Wu and C. Lee Giles, "Towards Better Understanding of Academic Search," *JCDL '16: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (2016): 111–114, http://doi.org/10.1145/2910896.2910922.
2. A subject search can be defined as the search for documents, which potentially are relevant for a certain task—for example, writing an academic paper on a given subject/topic, the existence and identity of which are unknown for the searcher. (Concerning the concept of relevance, see Birger Hjørland, "The Foundation of the Concept of Relevance," *Journal of the American Society for Information Science and Technology* 61, no. 2 (2010): 217–37, https://doi.org/10.1002/asi.21261.)
3. Ray R. Larson, "The Decline of Subject Searching: Long-Term Trends and Patterns of Index Use in an Online Catalog," *Journal of the American Society for Information Science* 42, no. 3 (1991): 197–215, https://doi.org/10.1002/(SICI)1097-4571(199104)42:3<197::AID-ASI6>3.0.CO;2-T.
4. Larson here has an endnote 4 referring to Donna K. Harman (ed.), *The Fourth Text REtrieval Conference (TREC-4),* US Government Printing Office, 1996 (NIST Special Publication 500-236), Available at https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication500-236.pdf.
5. Min-Yen Kan and Danny C. C. Poo, "Detecting and Supporting Known Item Queries in Online Public Access Catalogs," In *Proceedings of the 5th ACM/IEEE Joint Conference on Digital Libraries. June 7 to June 11 2005* (2005): 91–99, https://doi.org/10.1145/1065385.10654.
6. Charles A. Cutter, *Rules for a Printed Dictionary Catalogue* (Government Printing Office, 1876): 10, available at https://babel.hathitrust.org/cgi/pt?id=wu.89101448959&seq=6). Cutter here uses the term "object" in the sense described in the *Oxford English Dictionary*: "I.2.a. A goal, purpose, or aim; the end to which effort is directed; the thing sought, aimed at, or striven for."
7. Michael D. Cooper and Hui-Min Chen, "Predicting the Relevance of a Library Catalog Search," *Journal of the American Society for Information Science and Technology* 52, no. 10 (2001): 813–27, https://doi.org/10.1002/asi.1140.
8. Don R. Swanson, "Requirements Study for Future Catalogs," *Library Quarterly* 42, no. 3 (1972): 302–15.

9. Birger Hjørland, *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science* (Greenwood Press, 1997).
10. Barbara M. Wildemuth and Ann L. O'Neill, "The 'Known' in Known-Item Searches: Empirical Support for User-Centered Design," *College and Research Libraries* 56, no. 3 (1995): 265–81, https://doi.org/10.5860/crl_56_03_265.
11. Jin Ha Lee, Allen Renear, and Linda C. Smith, "Known-Item Search: Variations on a Concept," *Proceedings of the American Society for Information Science and Technology* 43, no.1 (2007): 1–17, http://dx.doi.org/10.1002/meet.14504301126.
12. Pauline A. Cochrane, "A Paradigm Shift in Library Science" (guest editorial), *Information Technology and Libraries* 2, no. 1 (1983): 3–4.
13. Birger Hjørland, "Description: Its Meaning, Epistemology, and Use with Emphasis on Information Science," *Journal of the Association for Information Science and Technology* 74, no. 13 (2023): 1532–49, https://doi.org/10.1002/asi.24834.
14. Lee, Renear, and Smith, "Known-Item Search: Variations on a Concept," 1.
15. "Bibliographic ghosts" (or "phantoms") are references that refer to non-existing documents. About the example by Salton (1975) mentioned in the text, see Davin Dubin, "The Most Influential Paper Gerard Salton Never Wrote," *Library Trends* 52, no. 4 (2004): 748–64.) The extremely hyped generative AI system ChatGPT-4o provides many bibliographical ghosts and other kinds of what are often called "hallucinations," but should rather be called "fabrications" and "falsifications" (cf., Robin Emsley, "ChatGPT: These Are Not Hallucinations – They're Fabrications and Falsifications," *Schizophrenia* 9, 52 (2023), https://doi.org/10.1038/s41537-023-00379-4).
16. Michael Buckland, "Known Item Search and Subject Search," *Library Resources & Technical Services* 68, no. 3 (2024): 1–9, quote p. 4, https://doi.org/10.5860/lrts.68n3.8132.
17. Joan M. Reitz, *Dictionary for Library and Information Science* (Libraries Unlimited, Western Connecticut State University, 2004; Digital edition: *ODLIS. Online Dictionary for Library and Information Science*, http://www.abc-clio.com/ODLIS/odlis%5Fa.aspx). Entry "known-item search." Reitz's "definition" continues by including advice on how to perform KIS.
18. ISO 5127: 2017(E), *Information and Documentation: Foundation and Vocabulary*, 2nd ed. (International Organization for Standardization).
19. Buckland, "Known Item Search and Subject Search." (Already Michael Buckland, "On types of Search and the Allocation of Library Resources," *Journal of the American Society for Information Science* 30, no. 3 (1979), 143–47, https://doi.org/10.1002/asi.4630300305, criticized the dichotomy between KIS and subject searches. On p. 145 he wrote "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.")
20. Buckland here has an endnote 3: "For example, Michael Buckland, *Information and Information Systems* (New York: Greenwood, 1991), 105 no. 3; Birger Hjørland, *Information Seeking and Subject Representation* (Westport, CT: Greenwood [1997]), 14, 20."
21. Buckland has here an endnote 4: "Keyword searching is commonly used in subject searches, but not always, and keyword searches are not always subject searches. So the distinction in process between keyword search and other forms of search is different from the distinction in purpose between known item search and subject search and its examination would require a different paper."

22. Other categories of search include "fact retrieval," the search for specific information rather than for documents containing this information, and "area search [which] is one in which a person uses the on-line system to identify the area of the library where a group of subject or author related books are located," cf., Debra J. Slone, "Encounters with the OPAC: On-Line Searching in Public Libraries," *Journal of the American Society for Information Science* 51, no 8 (2000): 757–73 (quote p. 762), https://doi.org/10.1002/(SICI)1097-4571(2000)51:8%3C757::AID-ASI80%3E3.0.CO;2-T.

23. *Oxford English Dictionary* provides, among others, the following definition of the noun specimen: "4.a. A single thing selected or regarded as typical of its class; a part or piece of something taken as representative of the whole." (In this article, "specimen" is used, for example, on a single copy of a book representing all books published in the same edition of that book.) In relation to Buckland's distinction between particular document and specimens, different kinds of metadata are required for each of these categories. Individual documents are not normally considered in library services, but in special cases, such as rare book collections, detailed descriptions as known from the field of descriptive bibliography are needed. For an introduction to the different fields of bibliography, including descriptive bibliography, see Birger Hjørland, "Bibliographical Foundations of Information Science: A Review Essay," *Journal of Documentation* 81, no. 1: 128–46, https://doi.org/10.1108/JD-06-2024-0126. Concerning the description of individual copies of books, see Marcia Reed, "Provenance of Rare Books," in *Encyclopedia of Library and Information Sciences*, 4th ed., ed. John D. McDonald and Michael Levine-Clark (CRC Press, 2017), 3766–773.

24. See, for example, Maja Žumer, "IFLA Library Reference Model (IFLA LRM): Harmonisation of the FRBR Family," *Knowledge Organization* 45, no. 4 (2018): 310–18, https://www.nomos-elibrary.de/10.5771/0943-7444-2018-4-310.

25. Lee, Renear, and Smith, "Known-Item Search: Variations on a Concept," 9.

26. Buckland, "Known Item Search," 7.

27. Lee, Renear, and Smith, "Known-Item Search: Variations on a Concept," p. 3 made a distinction between operational and conceptual definitions of a KIS, where operational definitions confuses the conceptual issue with the issue about how to perform KIS.

28. ISBN, on the other hand, is not normally considered an essential element in reference styles for academic writing, but it is a standard element in library catalogs. The fact that hardback and paperback versions get different ISBNs is important for the book trade, but not for scientific communication and library users. The researcher identification ORCID (Open Researcher and Contributor ID) was established in 2009 as a collaborative effort by publishers of scholarly research in order to resolve the author name ambiguity problem in scholarly communication. Many journals now require an ORCID, and it is now used by databases such as Web of Science, but it seems not (yet) to be demanded in academic reference manuals or to be used by library catalogs.

29. Frederick G. Kilgour, "Known-Item Online Searches Employed by Scholars Using Surname plus First, or Last, or First and Last Title Words," *Journal of the American Society for Information Science and Technology* 52, no. 14 (2001): 1203–209, https://doi.org/10.1002/asi.1186.

30. Christiane Behnert and Dirk Lewandowski, "Known-Item Searches Resulting in Zero Hits: Considerations for Discovery Systems," *Journal of Academic Librarianship* 43, no. 2 (2017): 128–134, https://doi.org/10.1016/j.acalib.2016.12.002. The authors investigated the reasons that KISs in discovery systems resulted in zero hits and identified the following reasons: (1) item in stock, but query incorrect (e.g.,

containing spelling errors), (2) item not in stock, (3) item in stock, but incomplete or erroneous metadata, (4) query is ambiguous or not understandable.

31. For example, in the colophon of Rick Szostak, *Integrating the Human Sciences: Enhancing Cohesion and Progress Across the Social Sciences and Humanities*. Routledge, the publication year is 2023, but the book was out, catalogued by the Royal Library in Denmark and borrowed by me in 2022.

32. Marcia J. Bates, "Information Search Tactics," *Journal of the American Society for Information Science* 30, no. 4 (1979): 205–14, https://doi.org/10.1002/asi.4630300406, provided an overview of twenty-nine "information search tactics." She did not explicitly discuss strategies for KISs, but Jeppe Nicolaisen, *Sådan finder du videnskabelig litteratur: databaser og informationssøgning* [How to Search Scientific Literature: Databases and information Searching] (Hans Reitzels Forlag, 2023), found that the combination of two of Bates' tactics, "EXHAUST" (extension of a query) and "REDUCE" (shortening of a query), provide the best results for KISs. Nicolaisen's suggestion is to expand the search request with information that is supposed to match the document sought, and if that document is not found, then reduce the search elements in order to remove potentially defective items and thereby increase the probability of success. In this process, it is helpful to have knowledge about which kinds of errors are common in bibliographical records and an understanding of why such errors occur (e.g., the confusion of family names and given names in documents by Chinese authors).

33. Alan S. Brown, "A Review of the Tip-of-the-Tongue Experience," *Psychological Bulletin* 109, no. 2 (1991): 204–23, https://doi.org/10.1037/0033-2909.109.2.204.

34. Jaime Arguello et al., "Tip of the Tongue Known-Item Retrieval A Case Study in Movie Identification," *CHIIR '21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, 2021: 5–14, https://doi.org/10.1145/3406522.3446021; Samarth Bhargav, Georgios Sidiropoulos, and Evangelos Kanoulas, "'It's on the Tip of My Tongue': A New Dataset for Known-Item Retrieval," In *WSDM '22: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022: 48–56, https://dl.acm.org/doi/10.1145/3488560.3498421.

35. Catherine M. Dwyer, Eleanor A. Gossen, and Lynne M. Martin, "Known-Item Search Failure in an OPAC," *RQ* 31, no. 2 (1991): 228–36.

36. Dwyer et al. (p. 235) wrote: "While it is tempting to suggest that more and better bibliographic instruction would increase the accuracy and efficiency of patrons' searching, it may be futile to try to provide instruction to the entire student body and/or faculty at an institution if many of them will be such infrequent users of the catalog that they will forget what they were taught before they come into the library again. It would probably be more effective to target instruction in the use of the online catalog and periodical printout to faculty and students who are just preparing to embark on research."

37. Kan and Poo, "Detecting and Supporting Known Item Queries," 93–94.

38. Svend Bruhns, *Bibliografisk Verifikation*, 2nd ed. (Center for Bibliographical Studies, 1999).

39. Possible reasons for the diminishing role of verification in libraries may be: (1) that verification has become easier in the digital environment so that few such requests are received by the libraries, (2) that the absence of direct requests (which formerly often were forms completed in writing) in the online context implies that the users' needs in this respect is not effectively communicated to the libraries, (3) that library administrators have downgraded this service because they believe it has become unnecessary, or (4) that the general library policy has changed toward making such tasks the users' own responsibility.

40. ANSI/NISO Z39.29-2005 (R2010), Bibliographical References, National Information Standards Information, available at: https://groups.niso.org/higherlogic/ws/public/download/12969/Z39_29_2005_R2010.pdf.
41. Reitz, *Dictionary for Library and Information Science*, entry "Descriptive cataloging."
42. Reitz, entry "Subject analysis."
43. Hjørland, "Description." A strong tendency since about 2000 has been to save libraries' own descriptive as well as subject cataloging and replace these with imported data. Therefore, departments for descriptive cataloging and subject classification have mostly disappeared today.
44. National Library of Medicine, for example, distinguishes descriptive and subject cataloging processes: "MMP [*Metadata Management Program*, formerly the *Cataloging and Metadata Management Section*] is responsible for review and development of cataloging policies for descriptive and subject cataloging and classification of all print, audiovisual, and electronic resources and applying them to resources acquired for the NLM collection." (Retrieved January 25, 2025, from https://www.nlm.nih.gov/tsd/tsdhome.html.) *National Library of Medicine* wrote: "A prospective indexer must have no less than a bachelor's degree in a biomedical science" (National Library of Medicine, 2018): "Frequently asked questions about indexing for MEDLINE: Who are the indexers, and what are their qualifications?" http://web.archive.org/web/20180415005151/https://www.nlm.nih.gov/bsd/indexfaq.html. However, "As of April 2022, all journals indexed for MEDLINE are done by automated indexing, with human review and curation of results as appropriate. MeSH indexing for MEDLINE was done completely by human indexers until 2011," https://www.nlm.nih.gov/bsd/indexfaq.html.
45. Concerning descriptive bibliography and subject bibliography, see Hjørland, "Bibliographical Foundations of Information Science."
46. Buckland, "Known Item Search."
47. Buckland, "Known Item Search," 4.
48. Robert Fugmann, "The Complementarity of Natural and Indexing Languages," *International Classification* 9, no 3 (1982), 140–44. Reprinted in *Theory of Subject Analysis: A Sourcebook*, ed. Lois M. Chan, Phyllis A. Richmond, and Elaine Svenonius (Libraries Unlimited, 1985), 392–402.
49. The term "biography" is also used about non-human entities such as libraries and towns, for example, Matthew Battles, *Widener. Biography of a Library* (Harvard College Library, 2004) and Simon Sebag Montefiore, *Jerusalem: the Biography* (Weidenfeld & Nicolson, 2011).
50. Buckland, "Known Item Search," 5.
51. If literature is sought about an individual concept, such as "Copenhagen," many documents may exist, and this is therefore a subject search. If a user is seeking for a particular document about Copenhagen, the term "Copenhagen" therefore is insufficient and other concepts, whether individual or general, have to be included in the search.
52. Buckland here has an endnote 15: "Peter F. Strawson, "On Referring," *Mind* 59, no. 235 (July 1950) 320–44.
53. Buckland, "Known Item Search," 5.
54. See Birger Hjørland and Lykke Kyllesbech Nielsen, "Subject Access Points in Electronic Retrieval," *Annual Review of Information Science and Technology* 35 (2001), 249–98. On pages 251–52 it is explained that "[h]ypothetically, it may be relevant to limit a subject search according to the name of a publisher, a

journal, or even a language code. Subject data are not strictly limited to specific kinds of data; under specific circumstances any kind of data may serve to identify documents about a Subject."

55. If a subject search is performed, and a number of potential relevant documents have been selected, these documents may subsequently be looked up in a library catalog, which is a KIS process. If there are errors in some of their bibliographical descriptions, this may make a further verification process necessary. The point here is that there are no differences in the properties of documents found in subject searching and documents found in known item searching: it is by principle the very same documents.
56. Buckland, "Known Item Search," 3–4.
57. On inter-indexer consistency studies see pp. 614–15 in Birger Hjørland, "Indexing: Concepts and Theory," *Knowledge Organization* 45, no. 7 (2018): 609–39, https://doi.org/10.5771/0943-7444-2018-7-609.
58. David W. Lewis, "Research on the Use of Online Catalogs and Its Implications for Library Practice," *Journal of Academic Librarianship* 13, no. 3 (1987): 152–56.
59. The original quote is in Michael K. Buckland, *Information and Information Systems* (Greenwood Press, 1991), 50; italics in original: "*We conclude that we are unable to say confidently of anything that it could not be information.*"
60. About bibliographical traditions, including descriptive bibliography, see Hjørland, "Bibliographical Foundations of Information Science," and Birger Hjørland, "Bibliography (Field of Study)" In *ISKO Encyclopedia of Knowledge Organization*, eds. Birger Hjørland and Claudio Gnoli, https://www.isko.org/cyclo/bibliography.
61. Slone, "Encounters with the OPAC," 763.
62. See Simon Wakeling et al., "Users and Uses of a Global Union Catalog: A Mixed-Methods Study of WorldCat.org," *Journal of the Association for Information Science and Technology* 68, no. 9 (2017): 2166–81, https://doi.org/10.1002/asi.23708.
63. See, for example, Lydia Dixon et al., "Finding Articles and Journals via Google Scholar, Journal Portals, and Link Resolvers Usability Study Results," *Reference & User Services Quarterly* 50, no. 2 (2010): 170–81, https://doi.org/10.5860/rusq.50n2.170.
64. Elaine Svenonius and Dorothy McGarry, eds., *Seymour Lubetzky. Writings on the Classical Art of Cataloging* (Libraries Unlimited, 2001). On p. 48, the editors wrote "Studies was a landmark in the history of Anglo-American cataloging. To begin with, it was notable for the approach it took. This was a systematic approach, which took its departure from the assumption that before describing a book it is necessary first to be aware of the objectives that description is to serve. Only then it is clear what is and what is not to be included in a bibliographic record. Only with an awareness of the objectives is it possible to evaluate existing rules and to make proposals for change."
65. See further in Rebekah Willson and Lisa M. Given, "The Effect of Spelling and Retrieval System Familiarity on Search Behavior in Online Public Access Catalogs: A Mixed Methods Study," *Journal of the American Society for Information Science and Technology* 61, no. 12 (2010):2461–76, https://doi.org/10.1002/asi.21433.
66. Kan and Poo, "Detecting and Supporting Known Item Queries."