FEATURE

Known Item Search and Subject Search

Michael K. Buckland

The traditional distinction between known item search and subject search is analyzed drawing on Robert Pagès' distinction between particulars and specimens, the distinction between objects and their properties, Robert Fugmann's distinction between individual concepts and general concepts, and the difference between referring and describing. The resulting analysis coincides with E. J. Lowe's four-category ontology.

There has long been much interest in different aspects of search: search strategies, search interfaces, search engine optimization, and more. There is a large applied literature on this and, notably, a study of web searches by Andrei Broder, who categorized web queries as informational, navigational, or transactional.[1] Many studies have been concerned with categorizing searches based on log files or user behavior. Much of this work is empirical. Terminology is not standardized and theoretical distinctions are not always clear.

My purpose is different in context and approach. First, I am concerned with searches in library catalogs and comparable bibliographic sources. Second, my approach is conceptual and theoretical. Third, my approach is independent of technology. Good practice is technology-specific; good theory is technology-independent, equally applicable in principle to printed, digital, or any other technology. Fourth, my point of departure is the traditional library distinction between known item searches and subject searches, a distinction which appears to have had little attention beyond passing mention in introductory texts and rather meager commentary.[2] In this article I offer a fuller analysis.

Known Item Search

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.

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.[3] 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.[4]

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. Relaxing the focus to any instance of a known edition or a known title reduces the practical difficulty but it does not change the underlying analysis. Two other distinctions provide a basis

Michael K. Buckland (buckland@berkeley.edu) is Professor Emeritus, University of California, Berkeley.

Library Resources & Technical Services | July 2024
https://doi.org/10.5860/lrts.68n3.8132

for fuller analysis and will be described below: (1) The difference between a particular and a specimen; and (2) the difference between material objects and non-material properties.

Particulars and Specimens

I use particular in a strict sense: Every actual, individual physical document, whether paper, microform, electronic, or in any other form, is a material (i.e., physical) object and, as such, it is unique. No other physical document, even if digital, no matter how similar, can occupy the exact same point in space and time. Each is an individual. We may say that nothing else can be the same, but the word *same* is unreliable because of its ambiguity. Same can mean acceptably similar for some practical purpose (as in the phrase, "It is all the same to me") or it can mean not different (as in "We were on the same flight"). German avoids this ambiguity by the use of separate terms *gleich* and *derselbe* respectively.[5]

A specimen, in contrast, denotes any member of a set of examples sharing some defining attribute, such as plants of the same species or printed books of the same edition. This is very similar to the type-token distinction, where items deemed the same are referred to as tokens of some type and also to the individual-kind distinction by which items judged to be in some sense to be the same are said to be individuals of some *kind*. The particular-specimen distinction has a small but important difference: specimens are defined by the sharing of some attribute or property but may be quite different in other ways. The type-token and individual-kind distinctions are tidy and useful in formal logic. The particular-specimen distinction is characteristic of an untidy world. Different feline species (e.g., lion, tiger, domestic cat) are all specimens of the mammalian family *Felidae* but the species are otherwise quite different. And the specimens of each species have individual personal differences. The status of being a *specimen*, therefore, is defined by the sharing of some property or attribute by particulars that in other ways may differ.

Documentary Particulars and Specimens

The particular-specimen distinction seems rarely mentioned in bibliographic or library literature, but a brief and colorful discussion was provided by Robert Pagès in 1948. Pagès shared the European documentalists' expansive definition of document which included any material object from which one might learn.[6] In an essay on the cultural role of documents in society Pagès distinguished particulars (which he called "auto-documents," as they only represent themselves) and specimens as follows:

> Specimens and Auto-Documents
>
> A "unique object" or a member of a collection of similars perhaps kept and used as a "document" about the domain to which it belongs, or even about its own objective properties as 1) an individual or 2) of a type.
>
> In the second case it is matter of a specimen: an anonymous Egyptian mummy, a gorilla in a cage, a piece of rock crystal. In the first case the document transmits information about itself (Napoleon's hat, a unique meteorite). It is, therefore, an "auto-document."[7]

Pagès does not appear to have published any development of this analysis. If he had, he could have added that every one of his examples could be regarded either as a particular or as a specimen. Pagès' specimens could also be regarded as particulars. An individual gorilla in a cage can indeed be considered as a specimen of the genus *gorilla* or of primates or of zoo favorites or any other imaginable kind, but gorillas nevertheless also exist as individuals. Each one has had a mother, a unique life history, a distinct personality, and its own pattern of behavior. Napoleon's hat may be a unique artifact, but it could also be regarded as a specimen of the headwear of its period or of Napoleon's possessions, for example. So every document and, indeed, every material object is always necessarily a particular but may also be regarded as a specimen in an unlimited number of ways. The choice of how it is regarded in any given circumstances depends on the nature of a subject's interest in it. Ways in which a particular may be a specimen appear to be limited only by one's imagination, so it is hard to imagine how any particular could not also be a specimen. Philosopher E. J. Lowe goes further and declares that no particular is not also a specimen: "Individuals are only individuable *at all* qua instances of some species or sort. There are no 'bare' particulars, only individual exemplars or samples of this or that *kind*."[8] Lowe does not appear to explain why this is so, but all material objects, hence all documents, exist in space and time, so, at the very least, they must have spatiotemporal properties and it is properties that make individuals "of a kind." Also, presumably, an individual without properties would not be discernable.

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. Material and non-material properties are different in kind, but that does not mean that they cannot be associated. A book's physical format and any topic discussed in it are different in nature, although they are sometimes associated closely enough for the topic (or at least the genre) to be inferred from the physical form.

These two distinctions—(1) particulars and specimens (i.e., individuals and kinds) and (2) material and non-material (abstract) properties—can be used to form four combinations as shown in table 1.

These four combinations and the differences between them can be illustrated with a worked example. For this purpose we use a book that was in Thomas Jefferson's personal library.

Example: Jefferson's *Bodin*.

American colonists wanting to justify independence from the British monarchy were attracted by the idea that sovereignty was somehow conferred on a ruler by the population and that this consent could be retracted. But, awkwardly, there was no known historical precedent for this convenient contractual theory. Thomas Jefferson, however, found one in his copy of Jean Bodin's *Les six livres de la république* [*The Six Books of the Republic*] (Paris, 1580) in which Bodin described the ritual whereby Slovene peasants interrogated each prospective Duke of Carinthia before allowing him to assume power.[9]

Jefferson marked the passage on page 129 of his copy.[10] Thus we have a historically significant instance among the many copies of one of the numerous editions of this title. Among the many possibilities, we can illustrate this situation using table 1 to distinguish four cases:

Table 1. Four combinations: Objects/Properties and Specimens/Particulars.

	Material Objects	**Non-material Properties**
Specimens	A set of material objects.	Shared properties.
Particulars	A particular object	The particular's properties.

Table 2. Jefferson's annotated copy of Bodin's *Republic* as an example of the four combinations in Table 1.

	Material Objects	**Non-material Properties**
Specimens	Kind: Bodin's *République* (Paris, 1580).	Shared property: Authored by Jean Bodin.
Particulars	Individual object: Jefferson's annotated copy of Bodin's *République* (Paris, 1580).	The particular's property: Owned by Jefferson.

1. Specimens: Copies of Bodin's *République* (Paris, 1580).
2. A shared non-material property of the specimens: e.g., authorship is attributed to Jean Bodin.
3. A particular material object: Jefferson's annotated copy.
4. Non-material properties of the particular: Jefferson's copy, e.g., owned by Jefferson.

Table 2 shows these four views displayed according to the arrangement in table 1.

Differences in Searching

"Searching for specific citations is unlike subject searching, much like hunting is unlike farming," wrote Donald W. Krummel.[11] The spatial difference is apt. The location of the farm is known to the farmer, but the location of prey is not known to the hunter. A closer parallel would be the difference between having an address when looking for a house compared with having only a description of the building.

Known items generally have names, addresses, or distinguishing physical features. Further, the known item target is quite specific, even if the details are incompletely known. In contrast, the resources responsive to a subject search may be numerous, varied, and widely dispersed even for a narrowly defined topic. Indexing theorist Robert Fugmann described the difference between searches for highly specific "individual concepts" and searches for "general concepts," which are broad and may subsume many concepts that are more specific.[12] Fugmann does not use the terms *known item* search and *not-known item*, but his arguments appear sufficiently similar to be applicable and his examples of "individual concepts" are persons, institutions, and towns, all with proper names, and which occur in single or very few instances. For these, he argues, natural language searches are generally sufficient. "It is typical of individual concepts … that they are represented with good predictability and perfect fidelity even in the natural language of the author."[13] As examples of "general concepts" he cites classes of organisms, properties, substances, and processes, e.g., the eradication of malaria transmitting insects by new pesticides. In these cases there are many possible named targets and a tendency toward a multiplicity of diverse narrative descriptions using varied terminology. In this situation reliance on natural language searching is inefficient and vocabulary control is desirable.

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, but in the humanities, unlike the sciences, searchers also make very heavy use of proper names for topic searching, especially for the difficult task of finding writings about a person's work or impact.[14] That is an example of how search practice can vary between different domains.

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.[15] 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. Strawson's distinction between referring and describing corresponds to Lowe's distinction between "individual terms," referring to singular items, and "sortal terms," referring to divisible sets (sorts, kinds) of items.[16]

Search Sequences

Searches do not, of course, occur only in isolation. One subject search is likely to lead to another and a known item search may well lead to another, different known item search. The two different kinds of search are sometimes used in conjunction. Here I note just two of the many possibilities:

1. Via set to known item. This can be an efficient process, especially if the details of a known item search are complex and/or incomplete, to start with a subject search that yields a retrieved set small enough that the known item becomes easily recognizable. For example, a title word search may be a convenient path if the author's name is hard to spell. The needle in the proverbial haystack has been recontextualized as a needle in a pincushion.
2. Via known item to set. Any particular book can lead to others like it. A nice example I used when teaching at Berkeley some years ago was an exercise that instructed students to do a subject search for books about the Vietnam War and to report the number found. At that time the online catalog in use supported search by Library of Congress Subject Headings but lacked cross-references between subject headings. A subject heading search using "Vietnam War" found no books at all, an implausible result. However, a title word search using "Vietnam War" would lead to a catalog record for a book about that war and inspection of the catalog record revealed that the Library of Congress Subject Heading was the phrase *Vietnamese conflict*, which retrieved hundreds of books.

Figure 1 shows these four kinds of search in relation to the same four-category table:

1. Known item search.
2. Subject search or other not-known item search.
3. Via known item to set, and
4. Via set to known item.

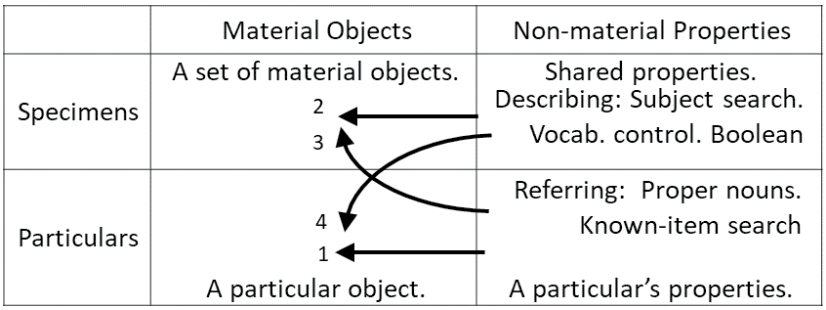

Figure 1. Searches: 1. Known item search; 2. Subject search and mixed strategies; 3. Via known item to set; and 4. Via set to known item.

Lowe's Four-Category Ontology

The tables above were derived by combining Pagès' distinction between specimens and particulars with Fugmann's distinction between individual concepts and general concepts. The resulting four combinations were found to coincide with the four-category structure of philosopher E. J. Lowe's metaphysical analysis of the natural world.[17] Lowe's terminology differs from mine but there are kinds of objects (substantial universals) characterized by properties and relations (non-substantial universals) and these are instantiated by particular objects (individual substances) and particular properties (tropes, "modes") respectively as shown in figure 2.

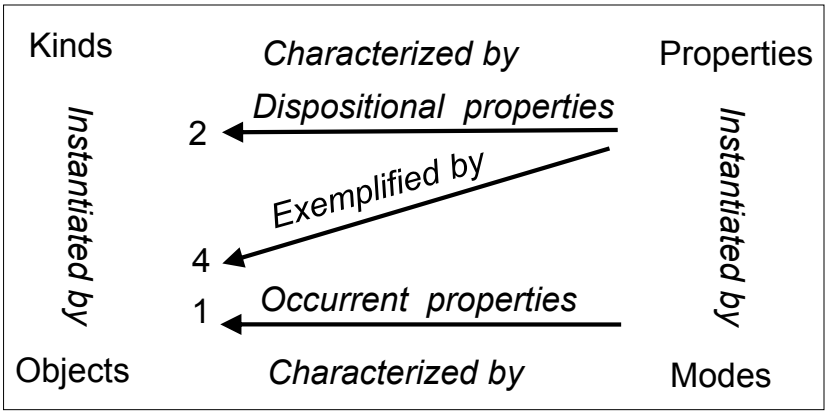

Figure 2. Lowe's four category ontology with search types from figure 1 added. Adapted from E. J. Lowe, *The Four-Category Ontology: A Metaphysical Foundation for Natural Science* (Oxford: Clarendon, 2006), fig. 7.1, p. 111.

Lowe's immanent, realist approach is concerned with theorizing the existence and properties of objects. In Lowe's view, any individual particular object is characterized by properties in two different ways: directly by any property specifically attributed to it (Lowe's "mode") and indirectly because any property attributed to a kind is also, by implication, exemplified more or less by the individual objects of that kind, as can be seen in figure 2.

Lowe notes a difference between these direct and indirect attributions. A direct ("occurrent") attribution might be a statement such as "This stuff is dissolving in water," an actually occurring characteristic of a particular object. The indirect ("dispositional, exemplified") attribution arising from the individual's status as a specimen of a kind is illustrated by a general statement such as "This stuff dissolves in water," which Lowe calls "dispositive," meaning a general tendency which might not obtain under all circumstances, for example at temperatures below freezing.

Lowe does not mention search, but, as a practical matter, search is the use of properties to find objects, so the relevance is clear. Of the four types of search enumerated in figure 1, three are shown in figure 2: Search type 1 (Known item search) is shown by the arrow labeled "Occurrent properties." Type 4 (Via set to known item) by the arrow labeled "Exemplified by." Type 2 (Subject search) can be considered implicit in the "Dispositional properties" arrow. Lowe makes no mention of type 3 (Via known item to set).

Summary and Conclusion

The traditional distinction between known item search and subject search is commonly mentioned but has received little examination. This paper provides clarification and a formal analysis. Relevant work was found to use quite varied terminology.

Known item search is understood 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. Strictly, a known item search is for a unique document, but the term is also used more loosely for searches for any copy of a specific edition or any copy of any edition of a specific title. Logically, the complement of a known-item search is a not-known item search, but the common type, subject search, has remained the customary term.

This analysis uses two basic distinctions. The first distinction is between a unique object (a "particular") and a specimen (i.e., an object with properties shared by other objects). Any given document can be regarded as either a particular or as a specimen depending on the purpose. The second basic distinction is between material objects themselves and properties associated with them. All searches use properties as criteria. Common examples are authorship, call number, or subject descriptor.

Natural language, especially the use of proper nouns, is likely to be effective for known item searches. In contrast, properties attributed to a set will apply indirectly to many objects which may share that property in varying degrees. Vocabulary control is generally needed for efficient subject searching. The difference between naming a wanted known item and specifying what is desired in a not-known item search corresponds to the distinction between referring and describing.

Directly assigned attributes can be expected to be specific and significant, but indirectly assigned properties likely less so. A property attributed to a kind will apply to—potentially—a very large set of individuals, who may share that attribute in varying degrees. Each individual can be expected to also have many other significant attributes. The indirect property may constitute a very small part of the attributes of some individuals. In consequence, generic (indirect) properties will be suitable for searches for any example or for all examples of some specified kind, as in subject searches. Generic properties can also be used to identify particular individuals, but that process tends to be relatively inefficient. In brief, the selective power of generic concepts is relatively weak when known items are sought compared with using specific properties directly attributed to the sought particular.

Combining the distinction between particular documents and kinds of documents with a division between material objects and non-material properties provides a two-by-two structure that is useful

for discussion. This matrix coincides with the four-category ontological framework advanced by philosopher E. J. Lowe.

The known item search and subject search distinction can be understood as a distinction between searching for a particular document and searching for a property.

Acknowledgments

I thank the University of California, Berkeley, France-Berkeley Fund for its support.

References and Notes

1. Andrei Broder, "A Taxonomy of Web Search," *ACM SIGIR Forum 3* 36, no. 2 (Fall 2002), 3–10.
2. For example, F. Wilfred Lancaster, *Information Retrieval Systems: Characteristics, Testing and Evaluation*, 2nd ed. (New York: Wiley, 1979), 6; Jennifer Rowley and Richard Hartley, *Organizing Knowledge: An Introduction to Managing Access to Information*, 4th ed. (Aldershot, England: Ashgate, 2008), 12.
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), 14, 20.
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.
5. For more detailed discussion of this and similar distinctions see Jonathan Furner, "Type–Token Theory and Bibliometrics," in *Theories of Informetrics and Scholarly Communication*, ed. Cassidy R. Sugimoto (Berlin: De Gruyter, 2016), 119–47. See also Wayne de Fremery and Michael Buckland, "Copy Theory," *Journal of the Association for Information Science and Technology* 73, no. 3 (2022): 407–18, https://escholarship.org/uc/item/6vf642mz.
6. Michael Buckland, "What is a 'Document'?" *Journal of the American Society for Information Science* 48, no. 9 (September 1997): 804–9, http://people.ischool.berkeley.edu/~buckland/whatdoc.html.
7. My translation. French original: "46 Echantillons et auto-documents. 460 Un "objet unique" ou membre d'une collection de semblables peut-être conservé et servir de "document" sur un domaine auquel il se rattache, ou bien sur ses propriétés objectives directes, 1) individuelles ou 2) spécifiques. Dans le second cas il s'agit d'un échantillon : une momie anonyme d'Egypte, un gorille dans une cage, un morceau de Spath. Dans le premier cas le document transmet des renseignements sur lui-même (chapeau de Napoléon, aérolithe singulier). C'est alors un "auto-document." Robert Pagès, "Transformations documentaires et milieu culturel (Essai de documentologie)," *Review of Documentation* 15, no. 3 (1948): 53–64, 60. Reprinted with an introduction and an English translation in Robert Pagès, *Documentary Transformations and Cultural Context* [special issue], *Proceedings from the Document Academy* 8, no 1. https://ideaexchange.uakron.edu/docam/vol8/iss1/.
8. E. J. Lowe, "Instantiation, Identity, and Constitution," *Philosophical Studies* 44, no 1 (1983): 52 (emphasis in the original).

9. Joseph Felician, *The Genesis of the Contractual Theory and the Installation of the Dukes of Carinthia* (Cleveland, OH: Felician, 1967).
10. Reproduced in Felician, *Genesis*, 15, English translation, 13.
11. Donald W. Krummel, "Bibliography," In *Encyclopedia of Library and Information Sciences,* 3rd ed., ed. Marcia J. Bates (Boca Raton, FL: CRC Press, 2010), 1: 522–33, 530.
12. Robert Fugmann, "The Complementarity of Natural and Indexing Languages," *International Classification* 9, no 3 (1982), 140–4. Reprinted in *Theory of Subject Analysis: A Sourcebook*, ed. Lois M. Chan, Phyllis A. Richmond and Elaine Svenonius (Littleton, CO: Libraries Unlimited, 1985), 392–402.
13. Fugmann, "Complementarity," 392.
14. Marcia Bates, "The Design of Databases and Other Information Resources for Humanities Scholars: The Getty Online Searching Project Report no. 4," *Online & CD-ROM Review* 18, no.6 (1994): 331–40. Reprinted in her *Information Users and Information Systems Design: Vol III of the Selected Works* (Berkeley, CA: Ketchikan Press, 2016), 262–82.
15. Peter F. Strawson, "On Referring," *Mind* 59, no. 235 (July 1950) 320–44,
16. Lowe, "Instantiation, Identity and Constitution."
17. E. J. Lowe. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science* (Oxford: Clarendon Press, 2006).