

UC Berkeley

UC Berkeley Previously Published Works

Title

INFORMATION-RETRIEVAL AND THE KNOWLEDGEABLE SOCIETY

Permalink

<https://escholarship.org/uc/item/54n5c000>

Author

BUCKLAND, MK

Publication Date

1990

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

Proceedings of the annual meeting

American Society for Information Science.
New York, Greenwood Publishing Corp. [1964?]-

<https://hdl.handle.net/2027/uva.x002070008>



Creative Commons Attribution-NonCommercial-NoDerivatives

http://www.hathitrust.org/access_use#cc-by-nc-nd-4.0

This work is protected by copyright law (which includes certain exceptions to the rights of the copyright holder that users may make, such as fair use where applicable under U.S. law), but made available under a Creative Commons Attribution-NonCommercial-NoDerivatives license. You must attribute this work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work). Only verbatim copies of this work may be made, distributed, displayed, and performed, not derivative works based upon it. Copies that are made may only be used for non-commercial purposes. Please check the terms of the specific Creative Commons license as indicated at the item level. For details, see the full license deed at <http://creativecommons.org/licenses/by-nc-nd/4.0>.

INFORMATION RETRIEVAL AND THE KNOWLEDGEABLE SOCIETY

Michael K. Buckland

School of Library and Information Studies, University of California
Berkeley, CA

ABSTRACT

Information retrieval, previously known as Documentation, is examined in relation to the ideal of a knowledgeable society in which people know what they need to know or can easily get to know what they need to know. This goal will require changes to current views of information retrieval. The intelligibility and credibility of what is retrieved needs to be considered. Information retrieval should be regarded as having to do with access to evidence. Objects and events as well as communications can be informative. Relevant writings by Francis Bacon, Suzanne Briet, Paul Otlet, and Gernot Wersig are noted.

INTRODUCTION

The theme for this conference is "Information in the Year 2000: From Research to Applications." The Call for Participation challenges us to "explore the 'Who's' and the 'What's' of current research so that we may better plan applications for tomorrow's needs". The purpose of this paper is to examine information retrieval, the central concern of the two societies responsible for this conference, in relation to the goal of a achieving well-informed society by the year 2000, the last year of the twentieth century. To the extent to which information (storage and) retrieval can be mapped on to the broader framework of society's need, such a mapping should contribute toward improvements in: (i) Our appreciation of what needs to be done; (ii) Our definition of the boundaries, extent, and limitations of our present concerns; and (iii) Our understanding of how information retrieval as a specialty fits into the broader goal of a more knowledgeable society.

In the context of the two societies sponsoring this conference, it would be usual to speak of "information science in relation to the knowledgeable society". However, in this context "information science" generally means information retrieval, broadly viewed. Since the term "information science" is also used with other meanings in other contexts, I shall use the narrower term "information retrieval".

There is a substantial discrepancy

between current views of information retrieval and the broad goal implied by the theme of the conference. However, a modest broadening of our view of information retrieval could do much to increase its relevance--and our relevance--to the goal of a more knowledgeable society.

For most of its existence the American Society for Information Science was named the American Documentation Institute. The term "documentation" replaced "bibliography" in the early decades of this century as the preferred term in some quarters for information storage and retrieval, especially for science literature and other specialized information. Since 1950 the term "documentation" has increasingly been replaced by the phrase "information retrieval"--more fully, "information storage and retrieval"--and by "information science". In 1968 the American Documentation Institute was renamed the American Society for Information Science. However, publications and conference proceedings make it clear that documentation, alias information retrieval, has remained the central concern. For the present purposes, since reference is made to the periods before and after 1950, the terms "documentation", "information retrieval", and "information storage and retrieval" are used interchangeably.

It should, however, be noted that the European Documentation movement has been seriously underestimated in the modern information science literature. The widespread assumption that "the discipline of documentation was transported to the United States in the 1930s" [1] is historically inaccurate and misleading, as even a cursory examination of the original literature shows. There was, in fact, a remarkable intellectual discontinuity: Some of the more interesting ideas of the Documentation movement in the first half of this century concerning information science and information technology remain essentially unknown in the United States.

This paper examines some of the boundaries defining information retrieval, indicates how these boundaries might be redrawn, and suggests what needs to be done, by ourselves or by others, in order to

achieve a more knowledgeable society by the year 2000.

INFORMATION IS POWER; OR, WHY RETRIEVAL?

Why would one wish to retrieve information? The question of what, in general terms, one might wish to document and to retrieve goes to the heart of the justification of the existence of information retrieval systems. The justification for information is that, on occasion, one wishes to be informed. This has been phrased in different ways by different writers. Dervin writes of "sense-making": We encounter a doubt or uncertainty in a problematic situation and we need to "make sense" of it [2]. Belkin popularized the phrase "anomalous state of knowledge" [3]. These and other phrases such as "distressing ignorance" [4] are but variations on the general theme of needing to clarify or to extend what we believe concerning some matter. In daily speech we refer to the need for information. Retrieval-based information systems constitute one important means of meeting this need. Meeting this need provides the basis for defining the purpose of information retrieval systems.

The need for information and for information retrieval is reflected in a popular slogan attributed to Francis Bacon, an English politician and lawyer active around 1600: "Information is Power!" What Bacon actually wrote was significantly different. He observed that ignorance was a source of weakness and concluded that, therefore, knowledge, as the opposite of ignorance, was source of power [5]. The distinction between information and knowledge is important. Bacon did not say that "Literature is power" or that "Data are power" or that "Hypermedia are power". If we follow Bacon, then, we should say that what will be important in the year 2000 is not what information is retrievable but whether people know what they need to know or can easily come to know whatever they need to know. Our concern, therefore, is to relate information retrieval (alias documentation) to the social goal of increasing the extent to which people know or could easily know what they need to know in the year 2000. How, and in what ways, do our present concerns with information retrieval relate to the broad and noble goal of a knowledgeable society?

INFORMATION RETRIEVAL AND ACCESS TO INFORMATION

The term "access" is widely and loosely used in relation to information. "Access" is useful as a unifying concept [6]. If something does not have to do with some aspect of access to

information we should probably deny that it is part of information science. Access has several aspects including the three functions of information retrieval:

1. The delivery of information (physical access).
2. The locating of information (e.g. using catalogs and directories); and
3. The identification of information (bibliographic access).

Two other aspects of access that are (and should be) discussed in relation to information retrieval are:

4. The price of access to the would-be user, which extends beyond money to effort, time, and those things that "user-friendliness" is supposed to mitigate; and
5. The cost to the provider, which also extends beyond monetary considerations to include broader social values of security and propriety.

A problem with any one of these five aspects of access can effectively prevent access to information. But, even if no problem existed with any of these five, one has a successful information system only in the limited sense of an ability to deliver physical chunks of information. For anyone to become informed, to come to know what they need to know, two additional aspects of access become important:

6. Cognitive access: Can the individual understand the information that has been retrieved? and
7. Credibility: Even if the inquirer understands the information that has been retrieved, does he or she believe it?--and should he or she believe it?

Relatively little attention has been paid to the intelligibility and credibility of the masses of information that we store and retrieve so diligently [7]. The current level of interest in the intelligibility of the interfaces of retrieval systems is encouraging, but it is a small matter compared with the intelligibility and credibility of all that information that lies beyond the interface. We shall need to extend our interests in these two directions if we are to have a credible concern with whether people know what they need to know.

WHAT IS RETRIEVBABLE? DOCUMENT AND OBJECT

What sorts of information are retrievable? Robert Fairthorne,

commenting critically on careless usage of the terms "information" and "information flow", cautioned: "Any discipline must define its scope. That is, it must define what matters it will study explicitly. These matters must then be studied and talked about in their own terms... To begin with, the scope must include all those, but only those, phenomena that are essential to the nature of the study" [8].

The potential scope of information retrieval seems very extensive. What and where are the limits to information retrieval? The enormous literature on information retrieval contains surprisingly little discussion of the nature and range of phenomena that might be stored and retrievable. When we look we find a serious limitation: Leading theorists of information science have rather consistently limited the definition of information to intentional communication. Machlup wrote: "The noun 'information' has essentially two traditional meanings.... Any meanings other than (1) the telling of something or (2) that which is being told are either analogies and metaphors or concoctions resulting from the condoned appropriation of a word that had not been meant by earlier users." ([9] Emphasis in original). This view of information is dominant and seems largely unquestioned in information retrieval circles. The concern of information retrieval is regarded as being with recorded knowledge, with discourse, with representations of data, with text. Text is specified as an essential ingredient in Belkin and Robertson's account of the phenomena of information science, for example [10]. Fox took an even narrower view limiting information to propositions [11].

The notion of information retrieval, or even information science, being limited to recorded discourse is interesting in several ways. For example, texts and discourse are cultural artefacts, so this definition would appear to place information science as a discipline among the humanities and social sciences, at least in terms of the objects (though not necessarily the methods) of study. Further, this definition suggests that if information is always and only a form of communication, then information science and information retrieval should be seen as a specialty within communications studies. What is more important for our present concerns is that the restriction of information and of information science to communication is quite simply inadequate [12, 13]. Bacon did not say that "Text is power" or that "Receiving propositions is power". If you smell smoke, see flames, and feel the heat, do you have to wait for someone to communicate to you that there is fire?

One is informed by perceptions of evidence. Acts of communication are only one form of evidence. We are also informed by our perceptions of things that are not intentional communications, by seeing soil samples, dinosaur bones, and footprints in the sand.

The recognition that objects and events can be informative is long established outside the literature of information science in philosophy, in semiotics, and in common sense [14, 15]. The technical term for an informative object is a "natural sign". Lightning is a natural sign of thunder and a fever is a natural sign of illness. We are informed by these natural signs even though there is no communicative intent behind them.

The ignoring or rejecting in the English-language literature of information science of informative objects that are not intended communications is not only unnecessary and unhelpful, it also ignores prior work in our own field. Paul Otlet, 1868-1944, a leading figure internationally in the Documentation movement [16, 17], knew better than to limit the concept of information to intentional communications. Like Manfred Kochen [18], Otlet used the term "document" in a technical sense to denote that which is the subject of documentation, that which one stores and retrieves. In his Traité de documentation (1934), one of the completest, longest, and least-read books in our field, Otlet starts by defining "document" in a broad but customary way as the generic book. Documentation, he wrote, deals with any element that serves to indicate or reproduce thought expressed in any form [19]. But later he observed that even this broad interpretation of the customary sense of "document" is simply inadequate to serve people's need for knowledge. Objects that are not communications or expressions of human thought can be directly informative (i.e. not indirectly through written representations of them) and so constitute a category of information or "document". The written or graphic record, wrote Otlet, is a representation of physical things or of intellectual or abstract images of things. But the physical things themselves (objects) can be considered "documents" when they are studied directly or treated as evidence [20]. Otlet enumerated five types of object: Natural objects; artifacts; objects bearing traces of human activity; objects such as models, designed to represent ideas; and works of art. All of these, Otlet argued, are "documents" in the technical sense of being phenomena with which information retrieval (alias documentation) should be concerned.

INFORMATION RETRIEVAL AS EVIDENCE RETRIEVAL

A logical development of Otlet's position would be to speak of documentation (or information retrieval) as being concerned with access to evidence. The term "evidence" has very suitable connotations if you are interested in information from the users' point of view. Evidence can include objects and events as well as communications. Evidence is also passive. People do things to or with evidence: People perceive it, interpret it, hide it, fake it, suppress it, and draw erroneous conclusions from it. One does the things with evidence that one does with information. If something is not evidence in some broad sense, it is not clear that we should regard it as information. The "evidential" view of information science, that information science can usefully be regarded as being concerned with access to evidence, seems to be worth exploring.

This "evidential" perspective on our field has been quite clearly stated and systematically overlooked. A discussion of information retrieval as being essentially concerned with access to evidence can be found in a book by the Suzanne Briet published in Paris in 1951 entitled Ou'est-ce que la documentation? [21]. Briet quoted approvingly the statement that a document is evidence in support of a fact. The Union Française des Organismes de Documentation had defined "document", in the technical sense of that with which information retrieval (alias documentation) deals, as any basis for knowledge that is physically fixed and capable of being consulted, studied, or used as evidence. Briet's position was that information retrieval was, or should be, concerned with anything that was treated (or, more exactly, handled) as evidence. Further, she pointed out that written descriptions of objects were secondary, derivative documents. The object being described was the "primary document". This position can be taken one step further by saying that something is information if it is regarded as information [22].

Unfortunately this whole line of thinking in our field has been ignored in the pursuit of information as being and only being recorded knowledge. The Science Citation Index and the Social Science Citation Index indicate only two examples of Briet's book being cited and in neither case is the substance of her work discussed. Being interdisciplinary is not supposed to mean examining what is in other disciplines instead of what is in one's own.

EVENTS AND INFORMATION RETRIEVAL

In addition to being informed by objects as well as by communications, we are also informed by our perceptions of events. Unfortunately events lend themselves to being stored and retrieved even less than objects do. How different the study of history would be if events could be stored and retrieved! Events are (or can be) informative phenomena and so need to be included in any complete approach to information science. In practice we find that the evidence of events is used in three different ways:

1. Objects, which can be collected or represented, may exist as evidence associated with events: bloodstains on the carpet, fingerprints on the gun;
2. There may well be representations of the event: photos, newspaper accounts, memoirs. Such documents can be stored and retrieved; and, also,
3. Events themselves can, to some extent, be created or re-enacted. In the experimental sciences it is regarded as being of great importance that an experiment, an event, be designed and described in such a way that it can be replicated subsequently by others. Remember, however, that a description of an experiment results in a retrieved article is no more than hearsay evidence.

Regarding events as informative phenomena, even though events themselves cannot be retrieved, constitutes an important element in the full range of information resource management. If a recreated event is a source of evidence (i.e. information), then it is not unreasonable to regard laboratory (or other) equipment as being somehow analogous to and/or substitutable for the objects and documents that are usually regarded as information sources. In what senses does it matter whether the answer to an inquiry is retrieved from a database or from re-enacting an experiment?

In practice we find that we have to deal increasingly with interactive information systems that involve combinations of input, processing, retrieval and experiment [23]. We need to view information retrieval in relation to the broader context of information service that includes, increasingly, the re-enactment as well as the representation of events.

THE USER'S PERSPECTIVE

If we consider information retrieval

from the perspective of the person who wants to know something, then we find, following Gernot Wersig [24], that retrieving information is not always necessary for beoming informed:

1. Thinking may suffice. An anomalous state of knowledge might cease to be anomalous if we think about it. It is often helpful to have somebody to act as a sounding board or to argue with, not in order to acquire additional information but to help one sort out one's thoughts and beliefs. A consequence for information system design is that it could be helpful to have question-asking machines as well questioning-answering machines;
2. Observation or inspection of objects or events may enable us to know what we want to know even though the "natural signs" that are perceived are not associated with any communicative intent;
3. Communications that we receive may inform us; and
4. A distinguishable subset of communications, that class of communications that are indirect, having been stored and retrieved, may be what informs us.

We should note that contemporary information science is almost exclusively concerned with only one of these four categories, the last. Note also that any concept of information science that is, in practice, substantially restricted to information retrieval in the traditional sense (#4), or even to communication (#3 and #4) deals with a lot less than the range of phenomena that are important for a knowledgeable society. If we are to be concerned with "sense-making", with making states of knowledge less anomalous, and with reducing the amount of distressing ignorance, then we need to recognize the importance of all four of these categories.

BACON ON INFORMATION RESOURCES

Bacon, as we have noted, is famous for his dictum that "Knowledge is power". Less well known is a vision of the ideal information support system, the authorship of which has been attributed to Bacon. This ideal system, concerned with information in the year 1594, has four parts:

"First, the collecting of a most perfect and general library, wherein whosofer the wit of man hath heretofore committed to books of worth ... may be made contributory to your wisdom. Next a spacious, wonderful garden, wherein whatsoever plant the sun of divers climate, or the earth out of

divers moulds, either wild or by the culture of man brought forth, may be ... set and cherished: this garden to be set about with rooms to stable in all rare beasts and to cage in all rare birds; with two lakes adjoining, the one of fresh water and the other of salt, for like variety of fishes. And so you may have in a small compass a model of the universal nature made private. Thirdly, a goodly huge cabinet [museum], wherein whatsoever the hand of man by exquisite art or engine hath made rare in stuff, form or motion; whatsoever singularity chance, and the shuffle of time hath produced; whatsoever Nature has wrought in things that want life and may be kept; shall be sorted and included. The fourth such a still-house [laboratory], so furnished with mills, instruments, furnaces, and vessels as may be a palace fit for a philosopher's stone." [25]

If we are to be serious in our concern for a knowledgeable society, then we should follow Bacon and the European documentalists in emphasizing knowledge and in taking a broad view of information. Limiting our interests to recorded knowledge, messages, and communications excludes much that needs to be considered. We should think of information retrieval and information science as having to do with access to evidence. We should be concerned with whether people can understand or should believe what we store and retrieve. Information retrieval should be viewed in terms of enabling people to know what they need to know. As individuals we can properly choose to specialize in one small part of the overall picture, but we also need to relate our small parts to the broader social context in our planning for the year 2000. Doing both provides a rich agenda of research and applications.

NOTES

- (1) Dorothy B. Lilley and Ronald W. Rice, A History of Information Science 1945-1985 (San Diego, CA: Academic Press, 1989), 15.
- (2) Brenda Dervin and Michael Niles, "Information Needs and Users," Annual Review of Information Science and Technology, 21 (1986) 3-33.
- (3) Nicholas J. Belkin et al., "ASK for Information Retrieval: Part I. Background and Theory," Journal of Documentation, 38 (1982) 261-71.
- (4) Michael K. Buckland, Library Services in Theory and Context, 2nd

- ed. (New York: Pergamon, 1988).
- (5) In "De Haeresibus" ("Concerning heresies") in his Meditationes sacrae, 1597, Bacon observed parenthetically "Nam et ipsa scientia potestas est" ("For knowledge itself is power") Works (London: Longman Green, 1857-74), v. 7, 241, 253). Later, in his Novum Organum, Book 1, Part 2, Aphorism 3, he explained "Scientia et potentia humana in idem coincidunt, quia ignoratio causae destituit effectum" (Works, v. 2, 157). A translation is "Human knowledge and human power meet in one; for where the cause is not known the effect cannot be produced." (Works. New ed. (London: Longman Green, 1875-79), v. 4, 47).
 - (6) Buckland, Library, chap. 15.
 - (7) Patrick G. Wilson, Second-hand Knowledge: An Inquiry into Cognitive Authority (Westport, CT: Greenwood, 1983).
 - (8) Robert A. Fairthorne, "The Morphology of 'Information Flow,'" Journal of the Association for Computing Machinery, 14 (1967) 710-19.
 - (9) Fritz Machlup, "Semantic Quirks in Studies of Information," in Fritz Machlup and Uná Mansfield, eds. The Study of Information: Interdisciplinary Messages (New York: Wiley, 1983), 641-71, 642.
 - (10) Nicholas Belkin and Stephen E. Robertson, "Information Science and the Phenomena of Information Science," Journal of the American Society for Information Science, 27 (1976) 197-204.
 - (11) Christopher J. Fox, Information and Misinformation: An Investigation of the Notions of Information and Misinformation, Informing and Misinforming (Westport, CT: Greenwood, 1983).
 - (12) Michael K. Buckland, Information and Information Systems (New York: Greenwood, forthcoming).
 - (13) Michael K. Buckland, "Information as Thing," Journal of the American Society for Information Science, (Forthcoming).
 - (14) D. S. Clarke, Principles of Semiotic. (London: Routledge & Kegan Paul, 1987).
 - (15) Umberto Eco, A Theory of Semiotics. (Bloomington, IN: Indiana University Press, 1976).
 - (16) W. Boyd Rayward, The Universe of Information: The Work of Paul Otlet for Documentation and International Organization. FID Publication 520. (Moscow: VINITI, 1975).
 - (17) W. Boyd Rayward, "Otlet, Paul-Marie-Ghislain," in ALA Encyclopedia of Library and Information Services, 2nd ed. (Chicago: American Library Association, 1986), 626-28.
 - (18) Manfred Kochen, Principles of Information Retrieval (Los Angeles: Melville, 1974).
 - (19) Paul Otlet, Traité de documentation (Brussels: Editions Mundaneum, 1934. Repr. Liège, Belgium: Centre de Lecture Publique de la Communauté Française, 1989), 9.
 - (20) *Ibid.*, 217.
 - (21) Suzanne Briet, Qu'est-ce que la documentation? (Paris: Editions Documentaires Industrielles et Techniques, 1951).
 - (22) Buckland, Information and Information Systems.
 - (23) U.S. Congress. Office of Technology Assessment. Intellectual Property Rights in an Age of Electronics and Information (Washington, D.C.: Government Printing Office, 1986).
 - (24) Gernot Wersig, "The Problematic Situation as a Basic Concept of Information Science in the Framework of the Social Sciences," in International Federation for Documentation. Theoretical Problems of Informatics: New Trends in Informatics and its Terminology. FID Publication 568. (Moscow: VINITI, 1979), 48-57.
 - (25) Gesta Grayorum, ed. by D. Bland. (Liverpool: Liverpool University Press, 1968), 47-48.