Written Knowledge: A Literary Perspective on Indexing Theory

Jack Andersen

Royal School of Library and Information Science
Copenhagen, Denmark

Jack Andersen is a Ph.D. student at the Royal School of Library and Information Science, Copenhagen, Denmark. He is working on a project about the role of knowledge organization in written scholarly communication. He adopts a literary approach to subject literatures when analyzing the correspondence between the knowledge claims put forward in the scholarly literature and how these knowledge claims are organized and represented in bibliographic literatures.

Andersen, J. (2000). **Written Knowledge: A Literary Perspective on Indexing Theory.** *Knowledge Organization, 27(4).* 201-212. 33 refs.

ABSTRACT: The general concern of Bazerman's book *Shaping Written Knowledge. The Genre and Activity of the Experimental Article in Science* is written knowledge as it is produced by the academy. Bazerman discusses in particular the rhetoric, communicative, and epistemological issues of written knowledge. The article discusses these themes in a library and information science (LIS) perspective in terms of their implications for LIS research.

For several reasons, it is argued that this way of scrutinizing into written knowledge ought to be of special interest to LIS research. As an example of a particular field of research in LIS, the article discusses the relationship between indexing theory and written knowledge. Bazerman analyzes written knowledge from a literary point of view. Among other things, it is argued that indexing theory can be seen as part of literary theory in that some of the questions raised by the latter are also raised in indexing theory. Furthermore, it is put forward that the indexer can be considered an author. The indexer produces a text, the document representation, which is the text the user actually meets in the first place. That way, the producer of a document representation is to some extent responsible for the quality of the documents indexed.

Having discussed this relationship between written knowledge and LIS research in general and indexing theory in particular, it is concluded that LIS research ought to head toward more humanistic oriented research traditions, if the line of research presented by Bazerman should be considered useful for LIS.

Introduction

Bazerman (1988) approaches written knowledge from a rhetorical, communicative, and epistemological point of view. He scrutinizes the relationship between the style and content-oriented means of scholarly literature and its means and modes of productions.

I believe that Bazerman can deliver some insights for library and information science (LIS) research in general, and indexing theory in particular.

Because I use Bazerman's work --which can be considered a literary analysis of scholarly literature-- I imply that this kind of literary analysis could enhance our understanding of indexing. Bazerman's analysis of written knowledge will, therefore, serve as the premises when discussing indexing theory in the following.

Literary theory and indexing theory are normally worlds apart, despite the fact that they are both concerned with texts and interpretations thereof. The main difference is that indexing theory is not solely concerned with fiction and that it has an ultimately instrumental purpose, namely retrieval.

However, the questions literary theorists raise concerning literature --no matter if these are about the topic or theme, the author's intention, language use, or the role of the reader-- have in fact also been raised

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

by theorists of indexing (cf. e.g. Andersen & Christensen, 1999; Blair, 1990; Fairthorne, 1969; Hjørland, 1992; Hutchins, 1978; Mai, 2000; Wilson, 1968). However, studies into the relationship between literary theory and indexing theory are rare within LIS. Campbell (2000) is in this connection an interesting and worthy example. Campbell discusses literary theory and subject analysis in terms of the aboutness-meaning relationship. While this relationship is recognized in subject analysis theory, Campbell claims that literary theory does not make such a distinction when analyzing works of literature.

This article will be divided into three sections. The first section briefly outlines and comments on the main points of Bazerman's book. The second section will put Bazerman's book into the perspective of LIS, and thereby attempt to show its significance to the LIS research community. The third section will concentrate on the relationship between written knowledge, as it is treated by Bazerman, and indexing theory.

Shaping Written Knowledge

The author, Charles Bazerman, is a professor of Education and English at the University of California, Santa Barbara. His writings cover, among other things, the implications of literacy, genre theory, scientific and technical writing.

The assumption of Bazerman's study is "...that the corpus of scientific writing is one of the more remarkable of human literary accomplishments" (Bazerman, 1988, p. 13). The objective of the book is to dispel the view that scientists never have been and never will be writers (Bazerman, 1988, p. 15). Bazerman claims that, even though scientists' rhetorical choices might be self-conscious or unselfconscious, "In whatever way these writing choices are realized and become institutionalized, they shape the kind of thing we consider contributions to knowledge" (Bazerman, 1988, p. 15). That is, an understanding of the way scientific writing is carried out will provide us with an understanding of what kind of thing scientific knowledge is. Bazerman considers, therefore, written knowledge to be an important point of departure for studying how scientific fields establish knowledge claims through language.

There are four reasons, according to Bazerman, an analysis of scientific writing is an important challenge. First, science, as a social institution in society, has social and cultural influences on us; and we rely heavily on the statements/insight science offers. Therefore, analyzing the statements made in scientific discourse can contribute to a holistic understanding of the influence of scientific writing. Second, even though scientific methods of formulating knowledge may seem abstract to us and not accessible to common sense, our belief in the scientific methods of formulating knowledge remains faithful and largely unquestioned. Third, the successful achievements of science have had and still have great influence on all other domains of intellectual inquiry. Fourth, considerations and treatments of scientific writing have been isolated from other forms of writings.

Communication of scientific knowledge takes place to a considerable extent through scientific literature. Bazerman investigates the human aspect of scientific writing because, he argues, scientific writing influences the very nature of science. Humans use language to gain understanding and exchange meaning. Therefore, one cannot view the production of knowledge in isolation from language. According to Bazerman, however, to ignore language has been the norm rather than the exception. Bazerman (1988, p. 14; emphasis added) notes that:

> "Yet to write science is commonly thought not to write at all, just simply to record natural facts. Even widely published scientists, responsible for the production of many texts over many years, often do not see themselves as accomplished writers, nor do they recognize any self-conscious control of their texts. *The popular belief of this past century that scientific language is simply a transparent transmitter of natural facts is, of course, wrong...*"

Two things need to be said about the emphasized statement. First of all, the belief that language is a transparent transmitter of facts is, after all, only wrong when one does not hold the standpoint that language in fact directly maps onto some state of affairs. One might say that the belief is true when one is convinced by Bazerman's rhetoric; that language is not a means of direct, unmediated representation. The rhetorical power of communication of knowledge is actually the central theme in Bazerman's book. That is, when Bazerman argues for his standpoint he cannot escape the power of rhetoric himself. Second, when taking a look at the history of philosophy of language one may argue that the belief "that scientific language is simply a transparent transmitter of natural facts" has in fact not been as popular in the past century as Bazerman asserts. The logical positivists believed that only meaningful sentences were sentences

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

that described some proposition or state of affair; this is also referred to as the 'principle of verification'. Logical positivists would probably claim that scientific language corresponds to some natural facts out there. However, as a philosophic group, the logical positivists were rather small and had a short lifetime.

A majority of practicing researchers, however, may not be aware of the literary dimension of their knowledge communication. They have at most only an implicit positivistic attitude about knowledge communication. If writing style and rhetoric is implicit and not discussed, it can be interpreted as positivism.

There have, however, been some strongly stated and widely accepted alternative views. Wittgenstein (1958) and Austin (1962) opposed the conception of language argued by the logical positivists. They have had a strong impact on philosophy of language from the middle of this century. They both stressed the pragmatic and social aspects of language. They didn't conceive language as a means of representation, but as something used to communicate or to do something. Wittgenstein (1958) actually argued against his earlier picture theory of meaning, which he presented in *Tractatus Logico-Philosophicus* (1961).

By paying close attention to the textual form of written knowledge it should become clear what kind of thing knowledge is: i.e. the written form matters (Bazerman, 1988, p. 18). Bazerman examines three articles by investigating how four types of contexts are referred to, invoked, or acted on in the articles. These four contexts are:

1) The object under study
2) The literature of the field
3) The anticipated audience, and
4) The author's own self (Bazerman, 1988, pp. 24-25).

These four contexts are also named:

1) Language and reality
2) Language and tradition
3) Language and society, and
4) Language and mind (Bazerman, 1988, p. 26).

In examining three scientific articles from molecular biology, sociology and literary criticism respectively --i.e. articles representing the natural sciences, social sciences and the humanities-- Bazerman shows that the rhetorical strategies used by the authors to gain acceptance of their knowledge claims are very different. Even though close attention is paid to the symbols or words under examination, Bazerman does not intend to restrict his analysis to the symbols alone, because the symbols themselves interact with the world. As Bazerman says in almost Wittgensteinian terms: "Without use and activity there is no language" (Bazerman, 1988, p. 18).

Although Bazerman cautions against making generalizations of the three knowledge domains on the basis of this examination, it nevertheless tells us something about the diverse epistemological nature of these different fields. This stresses that language is of great significance in understanding how the knowledge that scientists communicate is given legitimacy. Not all scientists are aware of or will acknowledge that they, in fact, are authors and narrators reporting their research findings through language. Nevertheless, this does not change the fact that language influences the formation of what we conceive as knowledge in society, which in turn, does not mean that sciences cannot be objective. It merely means that what is conceived as objective can be formulated in various ways and with various rhetorical methods. Writing science is basically, according to Bazerman, solving rhetorical problems.

Bazerman furthermore shows that the experimental article, which was developed in the 17th-century, is a genre, which supports the empiricist view of knowledge. The experimental article reports an experimental situation, and the argumentative function of the article lies in the description of the experiment. Isaac Newton (1642-1727) is presented as influential in setting the norms for scientific argumentation and publication. What Bazerman here attempts to emphasize is that form influences content. This epistemological point made about form and content is further put into perspective when Bazerman shows that 20th-century social science articles have adopted and adapted the form of the experimental article in order to communicate their knowledge.

The impact of empiricism has also been stressed in recent years in LIS by, among others, Budd (1995), Cornelius (1996), Hjørland (1997), and, lately, Dick (1999).

Bazerman analyzes how scientific communication has developed historically. This analysis of scientific communication stresses that communication is at the heart of science as a social system: From announcement of scientific discoveries in the first scientific journal in 1665, *Philosophical Transactions of the Royal Society of London*, to today's peer review system.

According to Bazerman, various disciplines shed light on the problem of writing knowledge and on what the activity of texts is. Sociology of science con-

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

tributes with an understanding of how communication is organized in various academic communities, "and how texts fit in with the larger systems of disciplinary activity" (Bazerman, 1998, p. 4). Philosophy of science contributes with an understanding of how disciplinary activity is conceived of by people, because what people think they are doing can provide insights into how they use language to accomplish those things. History provides an understanding of how any scientific writing activity is conditioned by history, because to write is to be part of a history. Psychology provides an important insight into the problem of writing knowledge, because "As a historically realized, social, epistemological activity, writing is carried on through people. People write. People read" (Bazerman, 1988, p. 5). What a text is must therefore take into consideration how people create it and use it. By putting writing into these perspectives, Bazerman touches upon language theory. He does not think that the single (written) word can be considered in isolation from its use, because "If the written word could only be understood within a historical, social moment, that would vex many of our habits of looking at language and texts as fixed structured systems of meanings" (Bazerman, 1988, p. 5).

In general, Bazerman's book demonstrates how the origins of modern science influence scientific activity. The goal for science is text production, which cannot be understood without taking the role of language into consideration. It seems reasonable, therefore, as Bazerman does, to consider scientists as writers of knowledge. By stressing this point about science and scientists, Bazerman shows that scientists are faced with the same linguistic and rhetorical problems as other professions whose job is some kind of text production. In short, analyzing written knowledge produced by the academy matters.

Shaping Written Knowledge and Library and Information Science

Bazerman's analysis of written knowledge seems to present fruitful research potentialities to the library and information science (LIS) research community. Research into subject literature (or non-fiction as Bazerman chooses to name it) and its types, concepts and theories is a field of study that lacks attention in LIS. Bibliometrics is, of course, a field of study that has subject literature as an object. But due to the quantitative nature of bibliometrics, it cannot tell us something about the epistemological implications of the knowledge claims put forward in texts, or how texts are composed. Therefore, studies into composition of texts can be a very fruitful supplement to bibliometrics (cf. Hjørland, 1998a, p. 23).

Thus, LIS should acknowledge studying subject literature, because it is, primarily, subject literature that is the object for theories of knowledge organization, bibliometrics, and information seeking and use. Studying how scientific documents are composed, the different types of documents, and how language is an instrument in the process of creating scientific knowledge has implications for research into knowledge organization, full text retrieval, information seeking and use, and scientific communication in LIS. The reason for this is that subject literature, as Bazerman (1988, p. 60) points out, can be said to be the representations of worlds in words, and because "...nonfiction creation incorporates procedures tying texts to various realities." (Bazerman, 1988, p. 60).

Buckland (1999) makes a distinction between two traditions within LIS: 'A document tradition' and 'A computational tradition'. As the name might indicate the 'document tradition' is concerned with documents, with signifying records (Buckland, 1999, p. 970). The 'computational tradition' is concerned with finding uses for formal techniques such as mechanical or mathematical (Buckland, 1999, p. 970). Due to this distinction, Buckland links the concept of document to the 'document tradition'. About the 'document tradition' Buckland (1999, p. 971; my italics) writes that it "...has to do with knowledge, meaning, learning, description, and language and ambiguity, therefore, any view of it remains incomplete *unless some roots in cultural studies, in the humanities and qualitative social sciences, is acknowledged*." Several authors have in a similar way stressed the importance of the concept of document in LIS research (cf. Hjørland, 2000; 1998a; 1998b; Hjerppe, 1994 & Houser, 1986a), and LIS needs to acknowledge these roots in cultural studies, the humanities and qualitative social sciences when studying subject literature. Studies of literature (fiction) take place in a humanistic-historical context. The same applies to studies of subject literature. Because it places the study of subject literature in a humanistic-historical context, Bazerman's book and the kind of research it presents can suggest the importance of studying the concept of the document in LIS.

Bazerman himself states that "Important stories remain to be told about theoretical articles, reviews of literature, speculative articles, handbooks and other reference works, proposals, and various pedagogic genres – their separate histories and interrelation-

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ships." (Bazerman, 1988, p. 7), and further that "Literary studies of non-fiction rarely ventured beyond belletristic biographies and autobiographies." (Bazerman, 1988, p. 11). Authors in LIS ought to contribute to these important stories. For instance, seen from the point of view of information seeking, what does LIS actually know about handbooks as information seeking tools? And more important, if LIS has some knowledge of this, are others outside the LIS community aware of this knowledge?

However, LIS researchers have made some contributions. Concerning reviews of literature Lindholm-Romantschuk (1998) has made an interesting contribution. She analyzed quantitatively the information flow in the social sciences and humanities in terms of book reviews. What Lindholm-Romantschuk found out was, among other things, that the humanities tend to review many books from the social sciences.

In terms of review articles Woodward (1977) has analyzed the roles of these in information transfer, and distinguished between historical and contemporary functions of review articles. But research into primary literature (e.g. theoretical articles) and its implications for and relation to, for instance, knowledge organization is hard to find in LIS. One example, though, is Houser (1986b).

Another example from LIS, advocating the line of research presented by Bazerman, is Hjørland (1998a; 1998b). Research into subject literature is related to the study of documents. In this connection, Hjørland (1998b, p. 616) claims that LIS has not produced a theory of documents. According to Hjørland a theory of documents must be connected to an analysis of functional differentiation of primary, secondary and tertiary documents (Hjørland, 1998b, p. 616-617). Thus, Hjørland's work connects to Bazerman's work, providing a way of studying documents in epistemological terms.

One could claim that information retrieval (IR) research, as described by Ellis (1996), has contributed to moving attention away from the concept of document and its significance to LIS, and thereby making LIS and, in particular, mainstream information retrieval (IR) extremely sterile and naïve. In this connection it is strange to notice that Buckland (1999, p. 971) connects IR with the above mentioned 'document tradition'. But when developing IR-techniques such as search algorithms is it the same algorithm to be applied whatever type of document (e.g. primary or tertiary documents) it is applied to? How can algorithms for IR be constructed if no consideration is given to the style and conventions of, for instance, scientific writing, and the conditions under which it is produced? Algorithmic IR techniques do not seem to take into account that

> "An article in a journal can be written from one epistemology in a journal following principles inspired by a second epistemology. The same article can be indexed in a database influenced by a third epistemology, and used by a user interested in e.g. anorexia from a fourth epistemological point of view" (Hjørland, 1998a, p. 28)

Algorithmic IR techniques do not seem to consider documents as instruments or tools in communication; i.e. they do not conceive of documents as developed and produced in a social, historical, and cultural context, determining the materiality they cannot do without (cf. Duguid, 1996). Rather, algorithmic IR techniques seem to consider documents as ahistorical entities that can be exposed to algorithmic treatment distilling 'pure' information.

However, from the point of view of IR research, Ingwersen (1996, p. 30-34) seems to be aware of writing style and conventions, when talking about paragraphs or sections as possible access points to full text documents. But Ingwersen does not analyze this further in terms of epistemological and interpretative issues of texts and their role in communication, as is done in the work of Bazerman and others (cf. Bazerman & Paradis, 1991). In this connection Hjørland (1998a) is arguing for the relevance and importance of text composition, epistemology, and semantics when discussing subject access points in IR, and connects Bazerman's work into this framework. But in general, mainstream IR research does not seem to be aware of or acknowledge such kinds of problems, which are problems of how documents are composed, and of how documents structure their knowledge claims. Mainstream IR research is simply not that much concerned with epistemological and interpretative issues of text structures. Bazerman can be useful to LIS, because he delivers a foundation as to how to understand the composition of documents.

Written Knowledge and Indexing Theory

Indexing theory is here considered as being concerned with providing a coherent foundation as to speak of and understand the assignment of words and concepts, when organizing and representing texts (i.e. written knowledge) in information systems. That is, a

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

theoretical perspective from where LIS can speak of organization and representation of texts.

When discussing this, the epistemological issues of written knowledge, as recognized by Bazerman, cannot be excluded here. Hence, the epistemological issues must logically serve as the presuppositions for a discussion of the organization and representation of texts.

The way Bazerman analyzes subject literature is done by the use of the aforementioned four contexts: 1) the object under study (language and reality), 2) the literature of the field (language and tradition), 3) the audience/readers (language and society), and 4) the author's self (language and mind). This analysis is carried out in order to show how these contexts are referred to, invoked, or acted on. How a text refers to, invokes, or acts on, each of these contexts is explored through the specific features of language (Bazerman, 1988, p. 25).

The analysis Bazerman is applying to subject literature is basically some sort of literary analysis. It is therefore open to discussion whether such kind of analysis applies to subject literature at all. But what does an argument against applying such kind of analysis on subject literature produced by the sciences look like? An argument could be that there is a fundamental theoretical distinction between fiction and subject literature. However, while this might be true, it is still not an argument in favor of *not* applying literary methods, when analyzing literature produced by the sciences. Who can write a text (fiction or non-fiction) without using rhetorical and stylistic strategies? And as Bazerman (1988, p. 26; emphasis added) states, when being a narrator of scientific texts:

> "An author, in deciding which words to commit to paper, must weigh these four contexts and establish a workable balance among them. *A text is, in a sense, a solution to the problem of how to make a statement that attends through the symbols of language to all essential contexts appropriately*."

Accepting this statement by Bazerman, ought to justify the application of literary methods when analyzing subject literature.

By analyzing texts in relation to these four contexts, Bazerman believes that it will tell us something about what kind of thing knowledge is. In doing this Bazerman recognizes the epistemological implications of such an analysis (Bazerman, 1988, p. 26), because to say something about what kind of thing knowledge is, is in itself an epistemological issue.

If one substitutes 'knowledge' with 'a subject' in the sentence, then the sentence goes like this: "By analyzing texts in relation to these four contexts, it will tell us something about what a subject is". Thus, the point to be established is that the concept of subject is connected to the concept of knowledge. That is, indexing theory implies epistemological issues. This has also been recognized and stressed by Hjørland (1992), when he argues that the concept of subject should have a central role in indexing theory, and, in continuation of this, defines a subject as the epistemological potentialities of documents. Hjørland's definition of the concept of subject is not supposed to be confused with Wilson's (1968, 1968, p. 67) utility of a writing: "...for the utility of a writing, if any, is by no means bound to lie in its contribution to the understanding of its subject."

The implication of Hjørland's concept of subject to indexing theory is that to determine the subject of a document is tantamount to assertion of an epistemological statement about that particular document. To assert an epistemological statement about a particular document is, among other things, to say something about what kind of knowledge that document expresses through the subject(s) treated.

In the following, the four contexts presented by Bazerman will be discussed in relation to indexing theory. If one accepts Bazerman's way of analyzing subject literature, then it seems reasonable to assert that indexing theory is a kind of literary theory. If Bazerman's way of analyzing subject literature should be of any interest at all to LIS, it must be in relation to indexing theory. As stressed by Andersen & Christensen (1999), an indexing theory is also a theory of documents (texts), and it must account for the circumstances that constitute documents. The four contexts presented by Bazerman can, to some extent, be said to account for these circumstances, because Bazerman's examination "...will not be of dormant symbols lying quietly on flat pages. *The symbols will constantly lead us outward to the many worlds they interact with*." (Bazerman, 1988, p. 18; emphasis added).

Before proceeding with the discussion of Bazerman's four contexts in relation to indexing theory, it should be noted that LIS is in fact not unaware of such discussion. From a different angle, Wilson (1968, pp. 69-92) has discussed four methods of analysis to come up with the subject(s) of a document[1]. Wilson names his four methods as the purposive way, the fig-

[1] Wilson does not use 'document' but 'writing'

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ure-ground way, the objective way, and the appeal to unity (or rules of selection and rejection). The purposive way is concerned with the author's aim or purpose with the document. Hence, the purposive way has similarities with classical hermeneutics. The figure-ground way is concerned with what stands out or what is most emphasized in the document. The objective way of determining the subject of a document is concerned with counting references to items in the document. The idea of the objective way could be interpreted as the one behind automatic indexing. However, the fallacy of automatic indexing is that it confuses syntax with semantics. The appeal to unity (or rules of selection and rejection) is concerned with what makes the document coherent, unified and complete, and what is selected and rejected in order to make the document coherent, unified and complete.

As with Bazerman's four contexts, Wilson demonstrates that these four methods are each looking for something different, and, hence, will tell something different as to what the subject is. But in contrast to Bazerman's four contexts, which are applied at once on the same text in order to say something about what kind of thing knowledge is, Wilson doesn't seem to advocate his four methods to be applied simultaneously on the same document. Nonetheless, Wilson's appeal to unity can be interpreted to be the one method, which to a certain extent is equivalent to all Bazerman's four contexts.

Another example of approaching indexing by stressing the same things similar to Bazerman's approach is Swift, Winn & Bramer (1977). In their critique of the concept of aboutness, they put forward their 'multi-modal approach'. Swift, Winn & Bramer do not believe that established methods of indexing are appropriate for social scientists. Instead of 'carving things up', these authors suggest that documents ought to be analyzed in terms of theoretical orientation, method of research, empirical situation under study, data collected and form (Swift, Winn & Bramer, p. 91, 1977). The elements mentioned are elements that are characteristic of and obligatory in primary literature. Thus, by stressing these aspects and their importance in document representation, Swift, Winn & Bramer's (1977) 'multi-modal approach' actually recognizes the importance of research into primary literature. To a certain extent these elements are what Bazerman analyzes in terms of how they are referred to, invoked, or acted on.

Further, when defining a subject as being the epistemological potentials of a document, Hjørland's (1992) concept of subject comes very close to the 'multi-modal approach'. A document's epistemological potentials are, among other things, what theoretical orientations the document relies on, methods used etc.

The Four Contexts

Intuitively, the first context, the *object under study*, could be considered essential in determining the subject of a document. It is rather tempting to claim that the object under study must be the subject (cf. ISO, 1986). But what leads to this seemingly logical conclusion? The object under study does not in itself suggest such a conclusion. The object under study might just be there in some way or another, but that doesn't automatically turn it into the subject of a document. A given document (e.g. primary literature) probably has an object under study, but various subjects and concepts can contribute to illuminate or mention the object under study, and the way the object under study is present in a document might vary according to what knowledge domain the document is representing. But it still doesn't establish the object under study as the subject of a document per se. The object under study may as well arise as the document moves on. This implies that the object under study can be conceived of as rather insignificant, when it comes to determine the subject of the document.

According to Bazerman, the way the object under study is referred to, invoked, or acted on, is not just a matter of how it is plainly described in words on the single page(s), but also of how the object under study is linguistically constructed in the particular document. An examination of the lexicon of the document reveals this. That is, what kind of information is conveyed in order to *talk* and *write* about the object under study, and thereby to "indicate the quality of tie between text and the world" (Bazerman, 1988, p. 25). This information might contribute to determine the subject of a document, because it depends on how the object under study is linguistically treated in the particular document. This may lead to determine whether the object under study is the subject of the document.

When analyzing a text in relation to the *literature of the field*, Bazerman thinks of how the literature is being used in the document. That is, are explicit citations or implicit citations being used, or is the literature used in a defensive or offensive way. If the way the literature of a subject field is referred to, invoked, acted on in a document can contribute to determine

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

what the subject of the document is, then one is to some extent moving outside the actual document. In this way, Andersen & Christensen (1999) have argued that indexing is a representation of meaning. This meaning is the result of what constitutes the document: *Its production, distribution, and consumption*. In other words, processes that lie outside the particular document. To some extent, then, the way the literature of a field is referred to, invoked, or acted on in the document can give indications of what the subject of the document is, because the literature of a field indicates the document's relationship to the previous literature on the subject, or how the document is in dialogue with the previous literature in the field. The previous literature on the subject might be said to represent a relatively stable body of knowledge, and therefore in some way constitutes the actual document.

The way the *anticipated audience* is referred to, invoked, or acted on in a document is thought of in the sense of what knowledge and attitudes the document assumes the readers will have, how the anticipated audience is addressed, how a given argument or knowledge claim is structured in the document (cf. Bazerman, 1988, p. 25). However, this context is not supposed to be confused with a subjectivist view of knowledge and subjects. It is not a question of how the audience conceives of and interprets the document, but rather of how the author of the document constructs and places his various knowledge claims in the document in order to attempt to persuade the audience of the validity of his knowledge claim; i.e. a kind of social negotiation between author and audience. In doing this, the author assumes something of his readers and their response to the document and the knowledge it delivers or argues for. However, if the author of a document assumes or expects something about his readers, then it follows that the author himself is not free to put on a page what he likes. This can be explained by the fact that documents are instruments in scientific communication, and as such they are linguistic devices in the communication of knowledge. The author must use a language common to the audience. The author of a document must recognize this, when having a desire to communicate knowledge to his audience. This common language can consist of the technical terminology applied in a particular field of study.

Seen from the point of view of indexing theory, how can this context contribute to determining the subject of a document? If the author assumes something about the knowledge the readers will have, then it is not unthinkable that the subject of a document can be rather implicit. For instance, if a literary critic comes up with a new interpretation of a given poem, the critic has to convince the audience of this new interpretation. The audience is assumed to know and be acquainted with the previous readings of the poem in question. The task is to show what's new in the interpretation delivered. The interpretation is only new if the audience can recognize it as such, and if the previous readings of the poem do not suggest such an interpretation. The new interpretation must be the subject of the document, not the previous readings of the poem. However, it is the previous readings that are under discussion in the document. The previous readings can be said to be the object under study, because it is the previous readings that the new interpretation reacts on. The subject of the document is the new interpretation given. Thus, the subject of the document is negotiated between author and audience and, therefore, is not something that the author created himself in advance.

The *author's self* is present in various ways in the document. First of all, the author is the mediating link between the knowledge to be argued for and the audience to perceive this knowledge. This mediation of knowledge is of course done by the use of language. It is at this point the author of a document becomes a narrator. The author has to tell a story that sounds reasonable to the anticipated audience. In fields of study where the degree of consensus as to what counts as knowledge is relatively stable, the author will probably be rather 'invisible' or objectified in the document. That is, the story exposes itself without the author sitting behind the wheel. In less paradigm-laden fields of study, the author of a document will probably be more visible, more subjective, and active in the document as to tell the story. Seen from the point of view of this context, the subject of a document is dependent on where the author is in the document, because the author is a public face "...which makes the reader aware of the author as an individual statement-maker coming to terms with reality from a distinctive perspective." (Bazerman, 1988, p. 26). From the point of view of this context, it is these individual statements made by the author that are of interest in determining the subject of the document, because it can be argued that indexing, at some point, is an interpretation and a representation of these statements. This is not to say, however, that it is the author who determines the subject of the

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

document. On the other hand, it is the author who has the power to decide which words to put on a page. It is also these individual statements that reveal the value assumptions made by the author. These value assumptions can themselves contribute to determining the subject of the document, because they might give clues as to what the document is or is not about. In the determination of the subject of the document then, one cannot escape paying attention to the value assumptions.

The above discussion of Bazerman's four contexts in relation to indexing theory has shown that, as also argued by Mai (2000), to try to say something about the subject of a document is not a matter of checking out titles, abstract, headlines, or counting words and the like. The way a document refers to, invokes, or thereby reproduces a document, namely a document representation (see figure 1). This way of conceiving the indexer and her connection with the document has strong similarities with a documentation science conception of the production and reproduction of documents. The reproduction process is here indexing, classification, and use of documents. According to Mangen (1999, p. 57), a documentation science analysis of documents looks at every step in a production-reproduction process and sees the document as the result of a process involving many agents and producers (Mangen, 1999, p. 58). Such an analysis recognizes the importance of paying attention to the processes constituting the actual document in order to understand that document. This is also illustrated in figure.

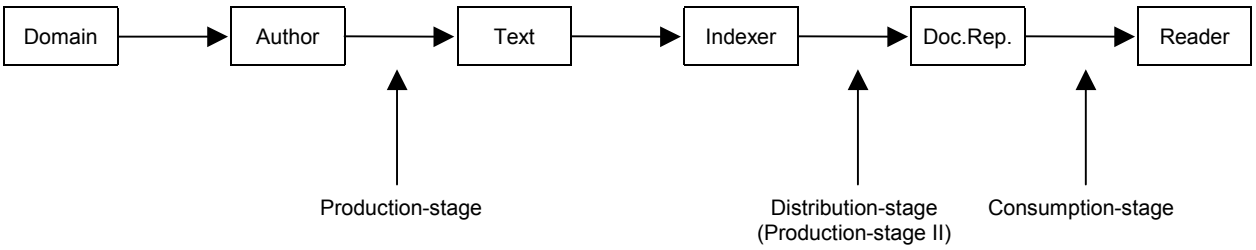

Figure 1: The indexer as author

acts on the object under study, the literature of the field, the audience or readers, and the author's self reveals that these four contexts are in interaction with each other, and shows that they are components that make the document coherent as a linguistic device in scholarly communication. The text in itself contains a lot of 'life' and is a product of a variety of circumstances (i.e. the four contexts) that contribute to a complex understanding of what the subject is.

To the extent Bazerman is right in claiming that these contexts and their presence in a document tell something about what kind of thing knowledge is, the concept of subject is a theoretical issue implying epistemological issues as well. An author of a piece of subject literature must take into account these contexts when producing his statements of knowledge. Here is the connection between the concept of subject and the concept of knowledge with regard to indexing. An indexer must also take into account these contexts when producing a representation of a text with descriptors. That way an indexer can be considered an author, because the indexer produces a statement of knowledge when representing a document and

Seen from an indexing and information seeking point of view, the reader (or user), as conceptualized in the LIS-vocabulary, does not meet the actual document, but a representation and reproduction of it that is an interpretation of the document. Thus, when the indexer produces a new text, Wilson explains in his discussion of how to determine the subject(s) of a document, why "...nothing definite *can* be expected of things found at any given position." (Wilson's emphasis, 1968, p. 92).

Also, by introducing the production, distribution, and consumption stages in the model, a link between indexing theory and Egan & Shera's (1952) social epistemology is established. They defined social epistemology as "...the analysis of the production, distribution and utilization of intellectual products." (Egan & Shera, 1952, pp. 133-134). Egan & Shera launched social epistemology, when arguing for a theory of bibliography, and through this emphasis was put on the epistemological issues of communicating knowledge through documents. There's need for a social epistemology, according to Egan & Shera, because 'traditional' epistemology is mainly concerned with "...the

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

intellectual processes of the individual" (Egan & Shera, 1952, p. 132); i.e. epistemologies like empiricism and rationalism. That way, epistemology does not come up with an answer to what the social function of knowledge in society is. That social function is illustrated by the fact that humans are dependent on recorded (written) knowledge:

> *"...the personal carrier of today, especially the "expert" in any field, himself relies upon graphic records far more than his memory or upon direct observation of his own limited environment."*
> *(Egan & Shera, 1952, p. 127)*

By connecting social epistemology with indexing theory, the epistemological aspect of indexing is emphasized, and enforces the social role of indexing in the communication of knowledge is enforced.

Being an author means to be part of a social context and social activity. Writing and indexing are social activities. The indexer as author, and the social activities embedded in it, can be illustrated by Austin's (1962) conception of language. Austin (1962) puts forward the conception of language that one can make something happen through language; that is, we are able to *perform* something with language rather than just describing something with language. By communicating we can do something. This is exactly what authors (and indexers) do. Austin distinguished between three types of performatives: locutionary acts, illocutionary acts and perlocutionary acts. Locutionary acts are thought of in the sense that to say something *is* to do something; the illocutionary acts are thought of as *in* saying something we do something and, finally, by perlocutionary acts are meant that *by* saying something we do something. Through indexing one can actually make something happen. One can advise some kind of usability or visibility of the documents indexed, or express its intensional or extensional aboutness (Fairthorne, 1969).

Considering language as something with which we can do or obtain something implies for a theory of indexing that it is also a theory of "how to do things with words". This aspect of language, recognized by Austin (1962), pays particular attention to the social activities of language that are vital for a theory of indexing to recognize. Indexing practice is itself a social activity, and an indexing theory should be the framework from which the actual indexing practice is supposed to be understood. In this way indexing can be interpreted as at one and the same time being a locutionary, illocutionary and perlocutionary act.

There is no reason, then, not to consider indexing theory as some kind of literary theory, to the extent that the latter is concerned with texts and interpretations thereof. However, while literary theory is concerned with the actual literary text as an object, indexing theory must be concerned with the universe of texts. This universe of texts is, following Wilson (1968), that with which bibliographical control is concerned. Indexing theory, then, is also connected to the enterprise of bibliographical control.

Conclusion

Many discussions within LIS have centered around what kind of thing LIS is and what it should be. However, one thing should be clear. As has been stressed in section two of this article, the concept of document ought to have a central position in LIS research. This conclusion may sound rather trivial, but the point is that documents have been developed in a historical context that has shaped the form and content of documents. This cannot solely be studied by quantitative research methods, as IR research usually makes use of, and this is where a humanistic research tradition enters the stage. The humanistic research tradition is basically founded in hermeneutics. If the LIS research community is to recognize the overall research potentialities presented by Bazerman's book as demonstrated in section three of the article, it is clear that LIS needs to reorient itself towards a more humanistic research tradition.

Acknowledgements

The author wishes to thank a couple of people for having contributed to the making of this article. I want to thank my Ph.D.-supervisor, Birger Hjørland, Professor at the Swedish School for Library and Information Studies, Borås, Sweden, for having introduced me to Bazerman's writings and encouraged me to write this article. I also want to thank Professor Robert P. Smiraglia, Palmer School of Library and Information Science, Long Island University, and Assistant Professor at Information School, University of Washington, Jens-Erik Mai for their valuable comments and suggestions on grammar and content. My gratitude also goes to Professor at Graduate School of Education, University of California, Santa Barbara, Charles Bazerman for having discussed some of the themes in his book with me and for having commented on my grammar. Finally, I want to thank the three anonymous referees for valuable comments.

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

References

Andersen, J. & Christensen, F. S. (1999). Wittgenstein and indexing theory. *Proceedings of the 10th ASIS SIG/CR. Classification Research Workshop*. Washington D.C. October 31. Ed. Hanne Albrechtsen & Jens-Erik Mai. (IN PRESS)

Austin, J. L. (1962). *How to do things with words*. Cambridge: Harvard University Press.

Bazerman, C. (1988). *Shaping Written Knowledge. The Genre and Activity of the Experimental Article in Science*. The University of Wisconsin Press.

Bazerman, C. & Paradis, J. (ed.).(1991). *Textual Dynamics of the Professions. Historical and Contemporary Studies of Writing in Professional Communities.* The University of Wisconsin Press.

Blair, D. C. (1990). *Language and representation in information retrieval*. Amsterdam: Elsevier.

Buckland, M. K. (1999). The Landscape of Information Science: The American Society for Information Science at 62. *Journal of the American Society for Information Science*, 50(11). 970-974.

Budd, J. (1995). An epistemological foundation for library and information science. *Library Quarterly*, 65(3). 295-358.

Campbell, G. (2000). Aboutness and Meaning: How a Paradigm of Subject Analysis Can Illuminate Queer Theory in Literary Studies. In: *Dimensions of a Global Information Science.* Edited by Angela Kublik. Proceedings of the 28th Annual Conference of the Canadian Association for Information Science. (http://www.slis.ualberta.ca/cais2000/campbell.htm)

Cornelius, I. (1996). *Meaning and method in information studies*. Norwood: Ablex Publishing Corporation, Norwood, New Jersey.

Dick, A. (1999). Epistemological positions and library and information science. *Library Quarterly*, 69(3). 305-323.

Duguid, P. (1996). Material matters: The past and futurology of the book. In: Geoffrey Nunberg (Ed.). *The Future of the Book.* Berkeley: University of California Press. 63-101.

Egan, M. & Shera, J. H. (1952). Foundations of a theory of bibliography. *Library Quarterly*. 22(2). 125-137.

Ellis D. (1996). *Progress and problems in information retrieval*. London: Library Association Publishing.

Fairthorne, R. A. (1969). Content analysis, specification and control. *Annual Review of Information Science and Technology*, 4. 73-109.

Hjerppe, R. (1994). A framework for the description of generalized documents. pp. 173-180. In: *Knowledge Organization and Quality Management: Proceedings of the third ISKO conference*, 20-24 June 1994, Copenhagen, Denmark. Ed. By Hanne Albrechtsen and Susanne Ørnager. INDEKS: Frankfurt/Main.

Hjørland, B. (1992) The Concept of "Subject" in Information Science. *Journal of Documentation*. 48(2). 172-200.

Hjørland, B. (1997). *Subject representation and information seeking. An activity-theoretical approach to information science*. Westport, Conn: Greenwood Press.

Hjørland, B. (1998a). Information retrieval, text composition and semantics. *Knowledge Organization*, 25(1/2). 16-31.

Hjørland, B. (1998b). Theory and metatheory of information science: A new interpretation. *Journal of Documentation*, 54(5). 606-621.

Hjørland, B. (2000). Documents, memory institutions and information science. *Journal of Documentation*, 56(1). 27-41.

Houser, L. (1986a). Documents: The domain of library and information science. *Library and Information Science Research*, 8. 163-188.

Houser, L. (1986b). The classification of science literatures by their "hardness". *Library and Information Science Research*, 8. 357-372.

Hutchins, J. W. (1978): The concept of "aboutness" in subject indexing. *Aslib Proceedings*, 30. 172-181.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation*, 52(1). 3-50.

ISO. (1986). *Documentation -- Guidelines for the establishment and development of monolingual thesauri*. International Organization for Standardization.

Lindholm-Romantschuk, Y. (1998). *Scholarly book reviewing in the social sciences and humanities. The Flow of Ideas Within and Among Disciplines*. Greenwood Press: Westport, Connecticut & London.

Mai, J. E. (2000). *The subject indexing process: an investigation of problems in knowledge representation.* Graduate School of Library and Information Science, The University of Texas at Austin. (Unpubl. PhD.-dissertation)

Mangen, A. (1999). Fra dokumentkompleks til bok til verk til tekst. Noen refleksioner omkring forholdet mellom dokumentasjonsvitenskap og litteraturvi-

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

tenskab. *Norsk Tidsskrift for Biblioteksforskning*, 12. 48-65.

Swift, D. F., Winn, V. A. & Bramer, D. A. (1977). A Multi-Modal Approach to Indexing and Classification. *International Classification*, 4(2). 90-94.

Wilson, P. (1968). *Two kinds of power. An essay on bibliographical control*. Berkeley: University of California Press.

Wittgenstein, L. (1958). *Philosophical Investigations*. New York: Macmillan Publishing.

Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. Translated by D. F. Pears & B. F. McGuiness. London: Routledge & Paul.

Woodward, A. M. (1977). The roles of reviews in information transfer. *Journal of the American Society of Information Science*, 28(3). 175-180.

NEW PUBLICATION

Content Organization in the New Millennium

(Papers contributed to the Seminar on Content Organization in the New Millennium, Bangalore, 2-4 June 2000)
Edited by A. Neelameghan and K.N. Prasad
Published by the Sarada Ranganathan Endowment for Library Science
ISBN: 81-900957-1-4

Content Organization in the New Millennium is a compilation of papers contributed to the Seminar on 'Content Organization in the New Millennium' (2-4 June 2000). There were nine invited presentations on various aspects of content organization.

The rapid developments in and widening range of use of the Internet worldwide is enabling easy access to information and data globally and almost seamlessly. The quantity, range and variety of information – as text, image, graphics, sound, and multimedia – that is accessible is indeed vast. On the other hand, the ease with which almost any data or information can be placed on and disseminated via the Internet is causing problems for information seekers. One of the causes of these problems is the amorphous nature of the information accessed which has only minimal organization. This results in, among other things, retrieving too much information that is irrelevant to the subject of interest to the user, and, many a time, it is like searching for a needle in a haystack. Recently, information professionals and subject specialists have become concerned with the situation and have experimented with tools, techniques and strategies, and with the use of time-tested classificatory ideas and other knowledge organization tools, such as thesauri, to mitigate the problems.

The paper on "Knowledge Management and Content Organization" by L.J. Haravu places the subject of content organization in the broader canvas of knowledge management (KM). Content organization and the tools necessary to aid knowledge discovery, a basic objective of most information seeking activity, is discussed in the paper "Content Organization as an Aid to Knowledge Discovery" by A. Neelameghan. In that paper, the role of statistical, informetric and scientometric techniques are mentioned, but elaborated on by I.K. Ravichandra Rao in his paper "Quantitative Techniques for Content Analysis."

HTML forms for web publishing and embedding metadata have been in wide use; but they are being extended or replaced by XML, XSL, etc. for customizing "Data Type Definitions" to enhance retrieval effectiveness. Shalini R. Urs and K.S. Raghavan discuss this aspect of content organization based on the experience of building a database of theses.

The variety of factors to be taken into consideration in content organization for Internet-based information services is elaborated by T.B. Rajashekar on the basis of practical experiences at the Indian Institute of Science. S.B. Viswakumar identifies factors that may affect content organization in multimedia databases.

Handling the scripts and vocabulary of Indian languages in organizing the contents of databases raises additional problems and issues, and these are being examined in an increasing measure as more and more such databases are being constructed in this country. B.A. Sharada considers some aspects of the problems of preparing databases in Kannada language.

The papers by M.A. Gopinath and G. Bhattacharyya deal, respectively, with the training required for and professional aspects of, content organization.

Content Organization in the New Millennium is the first publication in the Seminar Series of the Sarada Ranganathan Endowment for Library Science.

Price: Rs. 300/-; US $25.00
Mechanical details: iv + 148; Crown ¼; Paperback

For further details, please contact:
The Director, Institute of Information Studies (Sarada Ranganathan Endowment for Library Science),
702, 'Upstairs', 42nd Cross, III Block, Rajajinagar, Bangalore 506 010, India. Ph: (080) 3305109. Email: srles@vsnl.com

https://doi.org/10.5771/0943-7444-2000-4-201 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY