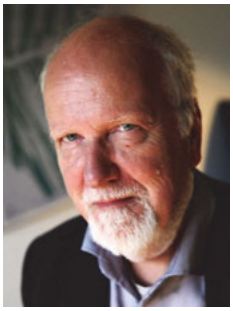
Indexing: Concepts and Theory†

Birger Hjørland

University of Copenhagen, Department of Information Studies,
Njalsgade 76, DK-2300, Copenhagen S, Denmark,
<birger.hjorland@hum.ku.dk>

Birger Hjørland holds an MA in psychology and PhD in library and information science. He is Professor in knowledge organization at the Department of Information Studies, University of Copenhagen (formerly Royal School of Library and Information Science) since 2001 and at the University College in Borås 2000-2001. He is chair of ISKO Scientific Advisory Council and a member of the editorial boards of *Knowledge Organization*, *Journal of the Association for Information Science and Technology* and *Journal of Documentation*. His h-index on 2018-03-30 is 45 in Google Scholar and 26 in Web of Science.



Hjørland, Birger. 2018. "Indexing: Concepts and Theory." *Knowledge Organization* 45(7): 609-639. 178 references. DOI:10.5771/0943-7444-2018-7-609.

Abstract: This article discusses definitions of index and indexing and provides a systematic overview of kinds of indexes. Theories of indexing are reviewed, and the theoretical basis of both manual indexing and automatic indexing is discussed, and a classification of theories is suggested (rationalist, cognitivist, empiricist, and historicist and pragmatist theories). It is claimed that although many researchers do not consider indexing to be a theoretical issue (or consider it to be a field without theories) indexing is indeed highly theory-laden (and the idea of atheoretical indexing is an oxymoron). An important issue is also the subjectivity of the indexer, in particular, her socio-cultural and paradigmatic background, as for example, when authors of documents are the best indexers of their own documents. The article contains a section about the tools available for indexing in the form of the indexing languages and their nature. It is concluded that the social epistemology first proposed by Jesse Shera in 1951 provides the most fruitful theoretical framework for indexing.

Received: 20 April 2018; Accepted: 10 May 2018

Keywords: indexing, index, indexes, subject content, documents, knowledge

† Thanks to Maja Žumer who served as editor and two anonymous reviewers who provided valuable feedback. Also, thanks to my colleagues Niels Ole Finnemann and Volkmar Engerer for reading, commenting and improving the article.

1.0 Definition of the terms[1] "index" and "indexing"

The word "index" comes from Latin and meant, according to Harper (2017), "one who points out, discloser, discoverer, informer, forefinger (because used in pointing), pointer, sign, title, inscription, list." Knight (1979, 17) wrote that the Latin word had the meaning "he who, or that which, points the way." In *Oxford English Dictionary* (2018) the following senses, among others, are given:

> Sense 4b: A sign, token, or indication of something
> Sense 5b: An alphabetical list, placed (usually) at the end of a book, of the names, subjects, etc. occurring in it, with indication of the places in which they occur
> Sense 5d: Computing. A set of items each of which specifies one of the records of a file and contains information about its address.

Today, the terms are used in different senses, for example, in economics about "cost-living indexes." In semiotics, Charles Sanders Peirce used index or indexical sign as one of three sign modes, the two others being icon and symbol.[2]

In library and information science (LIS), there have been different suggestions on how to define an index[3] and the process of indexing. Borko and Bernier (1978, 8) defined indexing as "the process of analyzing the informational content of records of knowledge and expressing the informational content in the language of the indexing system;" the ISO standard 5963:1985 defines indexing as "(t)he act of describing or identifying a document[4] in terms of its subject content," while Chan (1994, 166) pointed out that indexing involves basically three steps: 1) determining subject content of the item; 2) identifying multiple subjects and/or subject aspects and interrelationships; and, 3) representing them in the language of the subject headings list. While these definitions best suits manual indexing, other definitions may cover both manual and automatic indexing. Mulvany (2010, 486) wrote:

> In the United States, the National Information Standards Organization (NISO) [Anderson 1997]

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> defines an index as "a systematic guide designed to indicate topics or features of documents in order to facilitate retrieval of documents or parts of documents" (p. 39). The International Organization of Standardization's (ISO) ISO 999 [ISO 999:1996] defines an index as
>
> > an alphabetically or otherwise ordered arrangement of entries, different from the order of the document or collection indexed, designed to enable users to locate information in a document or specific documents in a collection (Section 3.5).

Many find these definitions too broad and imprecise. More thorough and lengthy descriptions of the purpose of an index can be found in the British Standard's "Function of an Index" [BSI 1988] and the American Society for Indexing's (ASI) "Criteria for the H.W. Wilson Award."[5] For general purposes, I find this definition useful (Mulvany 2005, 8): "An index is a structured sequence—resulting from a thorough and complete analysis of text—of synthesized access points to all the information contained in the text." A computer-generated list of words in the text, even arranged alphabetically, is not an index. For example, a concordance does not require analysis and synthesis of a text and its meaning. The concordance can only list words that appear in the text; it cannot include concepts or indicate relationships between topics. An alphabetical list of words does not truly qualify as the structured sequence that we associate with a proper book index.

While indexes are often alphabetically arranged, this is not always the case, as also reflected in Mulvany's definition above. (Alphabetization will be treated in an independent article in this encyclopedia). Another definition is (Taube 1953, 40):

> An index is an array of symbols, systematically arranged, together with a reference from each symbol to the physical location of the item symbolized. The items themselves may be stored in any arbitrary arrangement and yet located by virtue of the correspondence between them and their symbols. When names or verbal descriptions constitute the symbols, the established order of letters in the alphabet provides a convenient, searchable order of arrangement.

Weinberg (2017, 1978) suggested the following definition:

> An index leads from a known order of symbols to an unknown order of information. An index is in a different order from the document or collection to which it provides access.[6]

Building on Weinberg's definition, an index can be considered a kind of document, whether an independent document (e.g., printed or electronic as a database), a part of a document (e.g., a back-of-the-book index) or a structure embedded[7] in a document (e.g., in an XML document). The function of indexes is to provide access to information in or about other documents. Borrowing the terminology from translation studies, from a construction perspective[8] an index may be considered "a target document" and the documents indexed (whether independent documents or collections) may be considered "source documents." The task of providing access to information in source documents is done:

1) By deriving symbols from source documents or by assigning symbols about source documents (or by deriving/assigning symbols to specific places in source documents as in back- of-the-book indexes);
2) by providing a known order of symbols (e.g., alphabetical order);
3) by providing semantic relations between the symbols in the index (helping the users finding the right symbols); (this third step is possible, but not mandatory).

Contrary to Mulvany's quote above, a computer-generated list of words in the text, arranged alphabetically, fulfills a definition of an index, but is not a quality index (many human-made indexes may, however, be even lower quality).

Therefore, the following definition is here suggested: An index is a kind of target document, which has the function of providing access to information in or about some source documents by deriving symbols from the source documents or by assigning symbols about the source documents, thereby providing users access from a known order of symbols (e.g., A-Z) to an unknown place of information. An index often provides an order that is different from the document or collection to which it provides access; if not, it provides more or alternative entry points. In addition, an index may assist users finding the needed terms (or symbols) by providing semantic relations between indexing terms.[9]

Index and indexing can be defined reciprocally: indexing is the process of producing an index and an index is the product of an indexing process.[10] The process can be done by humans, by computer programs (which are made by humans and, therefore, also reflects human subjectivity) or by combinations. To provide an impression of the great variation of indexing processes, Section 2 provides a systematic overview of kinds of indexes.

2.0 Kinds of indexes

It follows from the definition on Section 1, that indexes may be classified by three overall sets of criteria: 1) criteria related to the kinds and attributes of source documents; 2)

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

criteria related to the attributes of the indexes themselves (target documents); and, 3) criteria related to the indexer, the indexing process, the context in which the indexing is taking place and the tools used.

2.1 Indexes classified according to kinds and attributes of their source documents

Indexes may be classified by the "kinds of documents being indexed." The most important categories being book indexes,[11] journal indexes, database indexes,[12] other text indexes (including mixed indexes), image indexes (still[13] and moving[14] images), sound and music[15] indexes, multimedia and non-text indexes,[16] and computer- and web-indexes.[17] As museum objects may also be considered documents, indexes of objects qualify as indexes in the present sense.[18] Kinds of documents may also refer to different domains and genres, such as historical,[19] medical[20] and legal[21] indexes or indexes of court decisions.

Indexes may be classified according to "indexable matters"[22] or "access points"[23] in source documents. Examples are title indexes, author indexes, descriptor- and keyword indexes, analytical indexes, citation/reference indexes and full text indexes.[24] Traditionally, words and phrases have been considered units (words defined as a sequence of alphanumerical characters surrounded by spaces) but ngram[25] indexes may be based on sequences of signs whether they include spaces. Theoretically, the most important attributes of documents are their subjects (www.isko.org/cyclo/subject). Titles, descriptors, references and full text may be considered different subject access points or means to determine the subjects of documents (cf. Hjørland and Kyllesbech Nielsen 2001). Descriptive cataloging (or descriptive indexing) refer to features other than a document's subject.[26]

Indexes may cover analytical entries or just comprehensive entries. An analytical entry indexes a part of a work (chapter in a book) or an entire work (story, play, essay or poem) contained in an item, such as an anthology or collection, for which a comprehensive entry is also made (See also note 32 concerning micro-documents and passage retrieval).

Indexes may be classified according "the coverage of the source documents;" for example, in cumulative indexes (=retrospective indexes) versus current indexes or comprehensive indexes versus selective indexes.[27]

2.2 Indexes classified according to the attributes of the indexes themselves

1) Indexes may be classified "according to their organization;" for example, alphabetical indexes and systematic indexes (as mentioned, a special article on alphabetization is under construction for ISKO Encyclopedia of Knowledge Organization). A systematic index may be arranged, for example, according to a subject classification, a chronological classification or a place classification).[28]
2) Indexes may be classified according to the kinds of signs used. This is, in particular, visible in the case of picture indexing. A back-of-the-book index is normally based on the same signs (mostly words) as the book itself. However, in picture indexing, for example, the picture may consist of signs other signs than those used in the index, for example, color, shape, and texture, or abstract attributes such as the significance of the scenes depicted, the latter using words (Chu 2001, 1011).[29]
3) Indexes may be classified according to their use of syntactical devices such as pre-coordinate indexes versus post-coordinate indexes and by their use of devices such as roles and links.[30] String indexes is one family of indexes.[31] Post-coordinated indexes were developed in parallel with information retrieval and have sometimes replaced pre-coordinate indexes also in the print environment.[32] Milstead (1984, 187) noted that it is not a question of better or worse; pre-coordination is the only appropriate method in the print environment (see also 2.3 §3).
4) Indexes may be classified according to their "locator[33] information," e.g., specific indexes (e.g., locators referring to page numbers or record numbers) versus relative indexes (locators referring to a concepts place in a classification system).[34]
5) Indexes may be classified in non-probabilistic versus probabilistic indexing systems. In non-probabilistic indexes, a given term is either assigned or not assigned to a given document. In probabilistic indexes, terms are assigned with an indication of their probability of being relevant (instead of a yes/no decision). Probabilistic indexing systems goes back to 1960 (cf. Maron 2008), when theory on indexing in information retrieval was still dominated by human indexing. Today, the dominant trend in probabilistic indexing is computer assigned probabilities.
6) Indexes may be classified according to the information they provide in annotated indexes and naked indexes.
7) Indexes may be classified according to their medium/physical form, e.g., printed indexes, card indexes and electronic indexes.
8) Indexes may be classified according to whether they are static (as a printed index) or dynamic (in which new links between source documents and indexes are added, deleted or changed).

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

2.3 Indexes classified by criteria related to the indexing process, context and tools

1) Indexes may be classified as human-based indexes versus computer-based indexes (each with many subtypes and possible combinations). Smiraglia and Cai (2017, 230) wrote about computer-based approaches:

> We have discovered a lively group of scholars centered around the use of what is called "clustering" and also around what is called "automatic classification." We have discovered that "automatic indexing" is often thought to be the same as "automatic classification" (although we acknowledge that it is quite different), and that "machine learning" has become a computer science paradigm that is larger than the problems of KO. In other words, we have demonstrated the fact that there are scholars involved in "clustering" and "automatic classification," and that they have a rich series of precedents over two decades, and that they share common thematic emphases.

In addition, software is often used in indexing, either for automatic indexing or as a tool in manual indexing (see e.g., Schroeder 2003, Browne and Jermey 2007, Chapter 10, 175-94 and American Society for Indexing 2017).[35]

2) Indexes may be classified as "derived" or "extracted" (all symbols used as headings in the indexed are taken directly from the source documents[36]) versus "assigned" (in which case the indexer may assign terms or symbols in the index that does not occur in the source documents).

3) Indexes may be classified based on the source of the assigned terms: for example, free assigned, indexes based on subject headings (pre-coordinated systems), indexes based on thesauri (post-coordinated systems), indexes based on systematic subject classifications (see also 2.2 §2).

4) Indexes may be classified based on "theoretical assumptions underlying the indexing process," for example, "document-oriented" indexes (based on words or concepts in the source documents, e.g., applying the 20% rule[37]) or "request oriented" indexes (based on the information provided and subjects). This principle is probably the most important in indexing theory and was suggested by Soergel (1985) and Lancaster (1991, 8).[38] It is discussed further in Section 3. Indexes may also be classified according to who the indexers are (e.g., author assigned keywords, indexing by subject specialists or by general information specialists).

This encyclopedia plans to provide specific entries for many of the above listed kinds of indexes, and no kind is, therefore, considered in depth in the present article.

3.0 Indexing theory

3.1 Atheoretical views

Weinberg (2017, 1984) in the section "Theory of Indexing" states "Indexing is not really a theory-based profession" and she concluded the section in the following way (1985):

> Indexing is an art, not a science. Many intelligent people lack the ability to distill the essence of a document and to represent its main topics in a few words. Some people who have gone through formal training will never make good indexers, while others who are self-taught are excellent indexers and have even won awards in the field.
> An indexer must be something of a prophet—envisioning the concepts likely to be sought by users of a document, expressing those concepts in terms likely to be sought by users, and providing cross-references from synonyms and alternative spellings as well as links to related terms to assist users in finding all the information that is relevant to their topics of interest.

Lancaster (2003, 35-37) shortly discusses theories of indexing and wrote (35): "A number of 'theories' of indexing have been put forward, and several have been reviewed by Borko (1977), but these tend not to be true theories and they offer little practical help for the indexer." Moreover (36):

> In fact, I have not been able to find any real theories applicable to the process of indexing although there are some theories (see, for example, Jonker 1964) that relate to the characteristics of index terms. Furthermore, I believe that it is possible to identify only two fundamental rules of indexing, one related to the conceptual analysis stage and one to the translation stage, as follows:
>
> 1. Include all the topics known to be of interest to the users of the information service that are treated substantively in the document.
> 2. Index each of these as specifically as the vocabulary of the system allows and the needs or interests of the users warrants.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

However, both Weinberg's expression "to distill the essence of a document" and Lancaster's "include all the topics known to be of interest to the users of the information service" may be considered theoretical views that are confronted with alternative theoretical views (as discussed below). Below it will be argued that "to distill the essence of a document" is a sentence that represents a rationalist view that is different from the view expressed as "include all the topics known to be of interest to the users of the information service," which represents a pragmatic view of indexing. Therefore, Weinberg's and Lancaster's views represent conflicting theories (but as demonstrated in the next section, Lancaster is not fully consistent in his theory and, therefore, fails to see his own view as a theory conflicting with other theories of indexing).

Theories of indexing are basically related to issues of subjectivity/objectivity (i.e., theories of knowledge), such as the indexer's interpretation and whether subjects are considered as something inherent in documents or as something related to the needs of users and purpose of the information service.

In the literature, discussions of the theoretical basis of indexing are often separated into human based indexing[39] versus computer-based indexing (see, for example, Anderson and Pérez-Carballo 2001a and 2001b; Stock and Stock 2013). This is, however, not a proper theoretical distinction, because human indexing and computer indexing are based on different theories of a deeper nature (human indexing may, for example, be very computer-like if the indexer follows a simple set of rules, see further in Hjørland 2011).

Jonathan Furner (2012) wrote in relation to the work about IFLA's principles known as "Functional Requirements for Subject Authority Records" (FRSAR):[40]

> Ultimately, the FRSAR Working Group does not take a philosophical position on the nature of aboutness; rather, it looks at the problem from the user's point of view (Zeng, Žumer and Salaba 2010, 8). The implication here is that, not only is it desirable to refrain from taking a philosophical position on the nature of aboutness when modeling bibliographic and authority data, but also that it is indeed possible to so refrain. On reflection, I have to admit that I am not comfortable with the Working Group's implicit endorsement of the latter claim. I am not sure that it is possible to avoid taking a philosophical position on this matter.

There are no such things as atheoretical views of indexing, but indexing theories may be implicit, unrecognized or unarticulated. To make progress in theory and practice, fruitful and explicit theories are needed. Regardless of whether the underlying theory is made explicit or not, it will nevertheless have a profound impact on the ways in which indexing is practiced, taught and researched. This will hopefully be made clear in the present article.

The field of epistemology covers a bewildering range of different theories and positions, but four of these are more fundamental and provides a very powerful way of analyzing theories of knowledge in general as well as single fields like indexing. All other theories of knowledge may be considered variants or combinations of one of these four positions: rationalism, empiricism, historicism and pragmatism.[41] This classification provides a systematic classification of theories of indexing. It has been pointed out that specific contributions may combine these epistemologies (Hjørland 2013a, 173). Dousa and Ibekwe-Sanjuan (2014) represents probably the most well-supported argument for eclecticism (or "epistemologico-methodological hybridity") in the construction of KOS. Although their paper made an important argument, it will be shown in the following that these epistemologies to some degree represent conflicting ideals and, therefore, resists eclecticism. It should also be noted that the idea of a pure rationalism or a pure empiricism is unattainable but nonetheless visible (if not dominant) as methodological ideals in the literature. Thus, the suggested classification of theories of indexing in epistemological schools provides the deepest and most fruitful understanding available.

3.2 Rationalist views

Rationalism (as opposed to empiricism) is a view found in, in particular, Descartes, Spinoza, and Leibniz, who tried to provide a foundation of knowledge on a few bedrock truths inhering in the rational soul prior to experience (cf., Fraenkel, Perinetti and Smith 2011, 6). The main methods associated with rationalism are logical intuition, logical deduction and *a priori* thinking (what is taken to be universal true independent of experience). Rationalism played an important role in "the cognitive revolution" in the twenthieth century, and one of its founders, Noam Chomsky, explicitly acknowledged his rationalist influence from Descartes.

Rationalist theories of indexing (such as Ranganathan's theory of classification and indexing[42]) suggest that subjects are logically constructed from a fundamental set of categories and that logical rules exist to determine the subjects of documents, index documents and afterwards search them. Rationalism is associated with the idea of an ideal language (Laporte, current issue), with logical atomism (the construction of all knowledge from a set of basic concepts) and with the idea of mechanical analysis. The basic method of subject analysis is "analytic-synthetic," to isolate a set of basic categories (=analysis) and then to

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

construct the subject of any given document by combining those categories according to some rules (=synthesis). The application of rules such as logical division is by principle part of the rationalist view (Hjørland 2011, 74). The rules and principles are understood as universal and neutral and reflecting an underlying order. We encountered, above, Weinberg's expression "to distill the essence of a document," which may be interpreted as a rationalist view. The idea that things (including documents) have essences is widespread but disputed.[43] In indexing, Fugmann (1979, 1985, 1992, 1993) proposed an advanced theory, "five axioms of indexing and information supply,"[44] which claims that indexing should be based on

> 1) the selection of the "essence" of the document to be indexed and
> 2) the description of this essence with a sufficient degree of predictability and fidelity. (Representational predictability is a core concept in this theory)

The philosophical assumption behind this view may be considered rationalist. Fugmann's approach is rationalist by construing a set of axioms for indexing based on logic rather than empirical research, but it is primarily rationalist by assuming that a given document has a determinate number of essences that can be identified and described in a way that is neutral to different views and needs[45] (it is characteristic for rationalist philosophy that it assumes that an order exists behind the unordered empirical knowledge). Two well-known indexing researchers reviewed Fugmann's book: Anderson (1994) and Lancaster (1994). Anderson (477) concluded:

> Fugmann's axioms, theories, and advice are important, and they merit careful consideration by all students, practitioners, and researchers of indexing and information retrieval, but his theory of indexing must be made subordinate to a theory of users and their use of retrieval systems.

Lancaster (1994, 150) wrote:

> Robert Fugmann is perhaps the only individual since Ranganathan who has made a serious attempt to produce some theoretical foundation for subject indexing" (149) and …
> It is difficult to quarrel with the axioms themselves; they are eminently sensible … Fugmann chooses to ignore economic realities and is prone to sweeping generalizations for which he offers no empirical (or any other form of support).

Both researchers recognized the importance of Fugmann's theory but had some reservations. Anderson found that the theory should be subordinate to a theory of users, but he did not put forward any developed criticism of Fugmann, just a loose hint.[46] Lancaster (2003, 12, 86), expresses viewpoints clearly opposed to Fugmann's axioms (see also endnote 38). Lancaster's view is that documents do not have essences, and that there is, therefore, no single correct set of index terms for a document: indexing should vary with the anticipated uses of the index.[47] How then, can Lancaster (1994, 150) state "It is difficult to quarrel with the axioms themselves; they are eminently sensible"?[48] Fugmann and Lancaster (as here cited) seem to subscribe to deeply conflicting theories, but Lancaster did not recognize this. One reason for this is that Lancaster is not consistent in his view, not even in the same book (Lancaster 2003). In the indexing exercises (chapter eighteen), for example, no hint is given to consider the possible different requests for which the indexing should be done. And in chapter five titled *Consistency of Indexing*, the possibility that indexer inconsistency is related to different views about what requests the indexing is supposed to serve is not set up as a possible explanation (although this is partly done in chapter six titled *Quality of Indexing*). Despite Lancaster's pragmatic formulations cited above, it seems that, in practice, this has not been followed systematically in his research and writings, and Lancaster may also be influenced by a conflicting rationalist philosophy. Probably, this is the reason he fails to disagree with Fugmann's axiom about the essence of documents and that he fails to identify any theories in indexing.

The idea that there is one correct way to index a document (or a collection) is associated with the view that such a correct indexing will be performed by all properly trained indexers and reflected by inter-indexer consistency (a measure of degree to which indexers agree in assignment of terms representing subject contents of document). Much research has been made from this assumption (see, for example, Lancaster 2003; Leonard 1977; Markey 1984; Soler Monreal and Gil-Leiva 2011). Findings from inter-indexer consistency studies has been met with disappointment, because human indexing reflects a large degree of inconsistency. The main conclusion that can be drawn from the tests is that inconsistency is an inherent feature of indexing, rather than a sporadic anomaly. Leinnger (2000, 4) gives some indication of the degree of this inconsistency:

> Consistency ratings for indexing methods that employ uncontrolled vocabulary have ranged between 4% and 67%, with an average of 27.05%. These ratings improve considerably, to a range of 13% to 70% and an average of 44.32%, when using a controlled vocabulary.[49]

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Frohmann (1990, 94) argued that lack of inter-indexer consistency should be met with more explicit guidelines for indexing and better trained indexers:

> The problem of indexer inconsistency … is not solved by first discovering and then bringing order to the motley of tacitly known rules unconsciously followed by indexers, but by replacing prevailing vague rules, for example, those providing no more guidance than 'express' the subject of this text in a concise statement', which indexers perforce interpret variously, with rules sufficiently precise to serve as justifications, as standards of correctness, and as instruments of indexer training.

However, the assumption that consistent indexing is good indexing is problematic as argued by Cooper (1969), because indexing may be consistently bad instead of consistently good. This is easy to understand if it is assumed that indexing theories are developed in research and taught to indexers. Theories may be more or less fruitful, and bad theories may cause indexers to index consistently bad.

That human indexing is sometimes taken as the golden standard to which computer indexing is adjusted is of course problematic in the light of the large degree of inconsistency found in empirical investigations and the uncertainty about how indexing should be evaluated.

3.2.1 The cognitive view

The cognitive view is a position mainly based on rationalism. It implies a view of indexing (as well as information seeking, and other processes related to information science) that the most fundamental intellectual operations are, in principle, explicable by internally realized and tacitly known rules that generate an indexing phrase from a given text. The human mind is understood in analog with a computer with certain universal attributes. It is assumed that there must be some rules guiding the mental activities of indexers and that the research problem in information science is to discover the precise form of these rules. Farradan's (1977) relational indexing is a case in point, because it represents "an attempt to simulate the structure of thought" (Farradane 1980, 76). Frohmann (1990, 94), on the bases of Wittgenstein's philosophy, criticized this view:

> First and foremost, it [mentalism, the cognitive view] conceals problems pertaining to the construction of rules. This paper assumes that information retrieval systems depend at a preliminary stage of their development upon rules governing the derivation of indexing phrases from texts. Wittgenstein's remarks on rules shift indexing theory away from rule *discovery* and toward rule *construction*. By Wittgenstein's lights, indexing rules governing the derivation of indexing phrases from texts are properly seen as instruments of particular social practices. Theory in indexing is therefore confronted with the challenge, not of discovering rules followed unconsciously, but of constructing, consistent with stated purposes, explicit, well-formulated, and strict rules which may be used to yield indexing phrases from texts.[50]

Based on this insight, the cognitive view seems paradoxical. Schools of LIS have for a long time before the cognitive view was developed studied and taught indexing. If it is assumed that indexers are influenced by what they have learned about indexing; how then, can LIS learn what should be taught by studying mental processes of people (whether or not they have received training in indexing)? The mental processes are not universal principles, hardwired in the human brain, but are learned and thus socially formed principles influenced by prevailing theories and technologies. With the words of Cooper (1978, 107; italics in original):

> Some of the studies have had the character of an investigation of how professional indexers currently *do* index, rather than how they *should* index. The upshot is that there is as yet no consensus among experts about the answers to even some of the most basic questions of what indexers ought to be told to do or of how an indexer's performance should be evaluated

Andersen (2004, 139-144) discussed "request, user and cognitive-oriented indexing" and wrote:

> A cognitive approach to indexing has been put forward in several writings by John Farrow (Farrow 1991; 1994 and 1995). Farrow's objective is to provide an understanding of the indexing process based on cognitive psychology and cognitive reading research. Reading research distinguishes between perceptual and conceptual reading. The former is relying on scanning the text for cues, whereas the latter is dependent on the background knowledge (e.g. knowledge of subject matter) a reader approaches the text with. Basically, Farrow argues that the indexing process may be viewed in light of these two modes of reading. It is, however, difficult to see what a cognitive approach to indexing offers and, if it offers something, what is cognitive about it. Turning indexing (and reading) into a cognitive matter is to remove attention away from the typified socio-cultural practices of document production and use, that

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> authors, indexers and readers are engaged in. Mai (2000, 123-124) also criticizes Farrow's cognitive model of indexing as it ... adds no further knowledge or instructions to the process. He simply says that indexing is a mental process, which can be explained by using models of human information processing from cognitive psychology. But these arbitrary models of minds, memory and cognition explain little about the indexing process.

Cognitive views on indexing are further presented or discussed by Beghtol (1986), David et al. (1995) and Hjørland (2013b).

3.3 Empiricist views

Empiricism is the view that all knowledge is based on experience, and the primary methods for obtaining knowledge must, therefore, be based on: 1) observations (or other sensations) made by individual observers; and, 2) inductions from pools of such observations (see further in Nickles 2005[51] and Suchting 2012).

Empiricist theories of indexing are based on the idea that similar (informational) objects share a large number of properties. Objects may be classified according to those properties. This should be based on neutral criteria, not on the selection of properties from theoretical points of view, because this introduces a kind of subjective criteria, which is not approved by empiricism (Hjørland 2011, 74).

The best examples of indexing based on empiricist assumptions are numerical, statistical procedures and retrieval techniques based on statistical measurement of similarity. However, empiricist ideals may also be found in manual indexing. The 20% rule (see endnote 37), for example, demands that the indexer chooses subjects that are contained in at least 20% of the document indexed (and thereby based on an empirical investigation of the document). The ISO 5963:1985 standard *Methods for Examining Documents, Determining Their Subjects, and Selecting Indexing Terms* seems also mainly empiricist although no real empirical procedure is put forward. Wilson (1968) examined—by thought experiments—the suitability of different methods of determining the subject of a document. One of his four methods may be classified as empiricist: "to group or count the document's use of concepts and references," which seems related to the 20% rule. The concept "literary warrant" (Barité 2018: http://www.isko.org/cyclo/literary_warrant)—at least in its extreme interpretation—also represents an empiricist philosophy.

The problem with empiricism is that it presumes that investigations should be made without subjective interpretations and theoretical assumptions. Perception is regarded a passive process, and data are supposed "to speak for themselves." Specifically, empiricism fails to consider that similarity cannot be an objective relation. Any two documents or objects can be similar in many ways. Therefore, it is necessary to choose from which perspective similarity is considered and which elements should be considered important, which, however, falls outside the ideals of empiricism, because it introduces an element of subjectivity.

3.4 Historicist views

Historicism is an insistence on the historicity of all knowledge and cognition. It is opposed to mainstream cognitive science and cognitivism in not regarding human beings as having universal cognitive characteristics, but to regard cognitive functions as historically and culturally specific and situated. It is intended as a critique of the normative, allegedly anti-historical, epistemologies of enlightenment thought.[52] Historicist approaches to indexing are based on the historical development of both object[53] as well as subject. Whereas rationalism and empiricism are individualist epistemologies, historicism, pragmatism and Kuhnian paradigm theory are examples of social epistemologies. Social epistemologies deny the rationalist assumption that human thinking is based on universal cognitive processes and that human perception and cognition are independent of the social and cultural context in which they take place. The mentioned theories see knowledge shaped in scientific and other traditions and paradigms. For indexing theory, this means that the way a document is perceived, interpreted and indexed varies from one social or cultural context to another (or from one paradigm or theoretical perspective to another). In addition, the users of the index will interpret the terms in the index from their knowledge and cultural or paradigmatic background. Such perspectives tend not to speak of "the essence" of documents but consider that different views tend to emphasize different aspects of documents. By consequence, documents must be indexed from explicit theoretical points of view to support the work of particular traditions and views (a fine example by Swift, Winn and Bramer (1973) is presented below in Section 3.8.4).

Hjørland wrote (Hjørland 2017, 4.2c):

> Hermeneutical theories of indexing suggest that the subject of a given document is relative to a given discourse or domain and is why the indexing should reflect the need of a particular discourse or domain. According to hermeneutics, a document is always written and interpreted from a particular horizon [note omitted]. The same is the case with systems of knowledge organization and with all users searching such systems. Any question put to such a system is

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

put from a particular horizon. All those horizons may be more or less in consensus or in conflict. To index a document is to try to contribute to the retrieval of "relevant" documents by knowing about those different horizons."

3.5 Pragmatist views

Pragmatism as an epistemological approach emphasizes the justification of theories and concepts by the examination of their consequences and of the goals, values and interests they support.[54] Hjørland wrote (2017, 4.2c):

> Pragmatic and critical theories of indexing are in agreement with the historicist point of view that subjects are relative to specific discourses but emphasize that subject analysis should support given goals and values and should consider the consequences of indexing. These theories emphasize that indexing cannot be neutral and that it is a wrong goal to try to index in a neutral way. Indexing is an act (and computer-based indexing is acting according to the programmer's intentions). Acts serve human goals. Libraries and information services [and classifications] also serve human goals, and this is why their indexing should be done in a way that supports these."

Jesse Shera, the originator to the term "social epistemology," expressed the pragmatic approach rather clearly (Shera 1951, 83-84 emphasis original):

> The pragmatic approach to classification through meaningful units of knowledge must be based on recognition of the obvious truth that any single unit may be meaningful in any number of different relationships depending on the immediate purpose. *Thus it is the external relations, the environment, of the concept that are all-important in the act of classifying.* A tree is an organism to the botanist, an esthetic entity to the landscape architect, a manifestation of Divine benevolence to the theologian, a source of potential income to the lumberman. Pragmatic classification, then, denies the existence of the "essence" of tree, for each of these relationships owes its existence to different properties of the tree. Relationship is not a universal, but a specific fact unique to the things related, and just as these relations reveal the nature of the relata, so the relata determine the character of the relationship."

The problems of inter-indexer reliability, for example, may partly be understood as related to different worldviews and interests by the indexers and not just considered errors determined by their individual cognitive capacities.

Feminist epistemology and other critical theories fall under pragmatism. Olson (2002) brings a critical feminist perspective to key issues in knowledge organization. The title of her book, *The Power to Name*, is in itself a powerful expression of an extremely important theoretical principle: the assignment of a subject to a document is not a neutral act but is a policy act contributing to facilitate certain uses of that document at the expense of other uses. Olson's book and other publications also emphasized that indexing is influenced by (mainstream) views, which does not consider the perspective of, for example, feminist epistemology. There are thus different ideological contrasts at play in indexing. The idea of neutral indexing must be given up and replaced by, for example, slanted indexing (cf. Guimaräes 2017).

3.6 Units of indexing

When the source documents are texts, the units to be indexed may be symbols, words, phrases, concepts or subjects. In other media, other kinds of signs may be used (see also 2.2 §2).

3.6.1 Word based indexes

Early mechanically produced word indexes include KWIC indexes (keyword in context indexes) coined at IBM in 1958 by Hans Peter Luhn, although it has an older history as keywords in titles (variations are KWAC (Keyword and context) and KWOC (Keyword out of context)) (See further in Landry and Rush 1975 about early developments in automatic indexing).

> An example of a form of mechanically derived full-text indexing is to let an algorithm make an alphabetical list of all the words in a document with locators indicating its place in that document. (A modified version is to have a list of "stop words" that should not be indexed. This is a bit less mechanical in that human interpretation is involved in the decision of which words should be considered stop words).

Such a word index has important limitations but is nonetheless often far more useful compared to no index or to poor manual indexes. It is limited, because, for example, some words are synonyms and, therefore, users should be guided to alternative words. Other semantic relations are also missing in derived word indexes, making the index less than optimal. This kind of index also suffers by not distinguishing important and unimportant words, and common

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

words like "indexing" may get a very large number of locators, thereby overwhelming the user who must examine them all to decide which places provide information.

Another kind of derived word indexing is made by highlighting important words in text and then making an alphabetical indexed based on the highlighted words (in text processing software there may be a way for authors to mark terms with a hidden code whereby the software can generate the index based on the marked terms, or by using a markup language such as XML). This kind involves a large amount of interpretation about which terms should be highlighted or marked. People may differ much in their choice of terms. Indexing research aims at (or should aim at) providing guidelines on how to select terms. This requires investigations into difficult problems such as "what information in the document is relevant to users?" This eliminates one problem compared to the first alternative, but the missing semantic relations still makes such an index less than optimal. Also, it is sometimes necessary to assign words, because the text does not contain the terms the users may need.

3.6.2 Concept based indexes

Indexing by using, for example, a thesaurus, is to take the step from word-based indexing to concept-based indexing. In case of synonyms, the user is guided to the preferred term, and all information is collocated under this term. In the case of homonyms, the ambiguity is removed by parenthetical qualifiers. The user is also guided to broader and narrower terms and related terms. Instead of words, we are, therefore, dealing with descriptors representing word meanings or concepts as units in indexing. The terms in thesauri include phrases.[55]

3.6.3 Subject indexes

Many texts about indexing stops at the conceptual level although a proposal for the differentiation between concept indexing and subject indexing was given by Bernier (1980). In his opinion, subject indexes are different from, and can be contrasted with, indexes to concepts and words. Subjects are what authors are working and reporting on.[56] A document can have the subject of "chromatography" if this is what the author wishes to inform about. Papers using chromatography as a research method or discussing it in a subsection do not have chromatography as subjects. Indexers can easily drift into indexing concepts and words rather than subjects, but this is not good indexing. Hjørland (2017b, Section 2.6) (http://www.isko.org/cyclo/subject#2.6) argued that citation retrieval may be applied as a kind of (semi)-automatic way of indexing and retrieving subjects.

3.7 Thought and language

It is common in indexing theory to understand the process of indexing as containing a translation process from thought to language to indexing language. Ranganathan, for example, claimed a three-stage process: From "idea plane" over "verbal plane" to "notational[57] plane" (cf., Ranganathan and Gopinath 1967, 327-8). This is a theoretical view, and a view that is not without criticism. Spang-Hanssen (1974, 29) argued that the distinction between idea plane and verbal plane is problematic, because the description of the two planes will lead to one and the same structure. Of this reason, he claims, it cannot be considered fruitful to speak about two distinct planes.

Lancaster (2003, 9-18) distinguishes two steps in subject indexing: 1) conceptual analysis; and, 2) translation. Again, this is an important theoretical issue that Lancaster and others fail to recognize as part of indexing theory. These two steps may certainly occur, but language (including indexing languages) also influences the conceptual analysis. Conceptual analysis is important in indexing, but it is also a philosophical method (and has been considered what defines philosophy as opposed to the empirical sciences). Conceptual analysis was a movement that ran into serious difficulties and "by the end of the 1970s the movement was widely regarded as defunct" (Hanna 1998, 518). The idea that philosophers (or indexers) have "*a priori*" access to true conceptual relations is a rationalist view that has been widespread (cf., Hjørland 2015a for a criticism of this view in LIS). There are different views of conceptual analysis, and Hanna (2007) outlines the most important. Here, we just conclude that concept analysis is not simply "given" but requires theoretical clarification and that an unfruitful understanding of it may lead to suboptimal indexing.

Mai (2000, 211 ff.) analyses the indexing process by applying Peirce's notion of "unlimited semiosis." Because Peirce claims that all thought is in signs, the view of clearly distinguished stages between thought and languages is not maintained in this theory.

The literature about relations between thought and language is very big and covers fields like linguistics, philosophy, psychology and sociology, among many others. We shall not go deeper into this problem here, but just realize that it is a theoretical problem in indexing theory that needs further study. It is also related to the distinction between rationalism, empiricism, historicism and pragmatism. The two last positions are related to the so-called "linguistic turn" and tend to consider human cognition as socially and culturally shaped and tend to consider conceptual analysis influenced by language.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

3.8 The subjectivity of the indexer

Subjectivity means the knowledge, understanding, views, ambitions etc. characterizing the person doing the indexing (including tagging). Often, objectivity is claimed as the ideal (e.g., Bell 1991, 173) but objectivity demands that there is one right position that the indexer (or somebody evaluating the indexing) knows. This corresponds to the rationalist view described in Section 3.2 and is opposed to the pragmatic or critical understanding (see also Swift, Winn and Bramer 1979 for an important criticism of objectivism[58]). Therefore, subjectivity should not just be considered a bad thing but also a necessary precondition of any kind of professionalism. It should be acknowledged that indexing must be biased (or slanted[59] as introduced by Guimaräes 2017) but not in any way, of course; it should support the activities that it is design for. It is possible to investigate different aspects of indexer subjectivity contributing to qualify the indexing process. Different kinds of indexers may be distinguished: 1) authors as indexers; 2) indexers trained in LIS; 3) contributors to social tagging systems; 4) people with high formal subject knowledge; and, 5) people with high formal subject knowledge trained as professional indexers.

3.8.1 Authors as indexers

Some journals ask their authors to provide keywords (for example, *Journal of Documentation*) just as some publishers demand that authors index their own books. Mulvany (1994) discussed authors as indexers and the cooperation between authors and indexers. Morville and Rosenfeld (2007, 105) wrote about "content authors:"

> However, even when authors select terms from a controlled vocabulary to label their content, they don't necessarily do it with the realization that their document is only one of many in a broader collection. So they might not use a sufficiently specific label. And few authors happen to be professional indexers.
>
> So take their labels with a grain of salt, and don't rely them for accuracy. As with other sources, labels from authors should be considered useful candidates for labels, not final versions.

Diodato and Gandt (1991) examined back-of-the-book indexes produced by thirty-seven authors and twenty-seven nonauthors and found that the nonauthors, many or all of whom were probably professional indexers, provided significantly more index pages, modified headings and modifiers than did the author indexers. The two groups were almost identical in their frequency of cross reference use.

3.8.2 Indexers trained in LIS

A purpose of training people in LIS and knowledge organization has often been—explicitly and implicitly—to qualify them as indexers and as managers and evaluators of indexers. We have already seen that Weinberg (2017, 1984) stated:

> Many intelligent people lack the ability to distill the essence of a document and to represent its main topics in a few words. Some people who have gone through formal training will never make good indexers

Lancaster (2003, 365), on the other hand, in the introduction to chapter eighteen "Indexing Exercises," wrote: "Practice makes perfect, in indexing and abstracting as in other activities." One important issue is, however, if indexers trained in LIS have the proper subject knowledge. Many approaches to indexing tend to downplay this issue. If there are different theories of indexing, the qualification of indexers may depend on the theory their education is based on. Therefore, no general conclusion can be drawn about people educated in KO (often considered "professional indexers"); it depends on the specifics of their knowledge. If courses in indexing do not improve the quality of the work done by indexers, something must be wrong with the course; for example, its theoretical basis or low standards. We shall comment more on this in sections 3.8.4 and 3.8.5.

3.8.3 Contributors to social tagging systems

Social tagging has been called "democratic indexing" (Rafferty 2018, 501). However, it may be a mistake to consider it without ideological bias. Gartner (2016, 103) wrote:

> The great strength of folksonomy is often claimed to be that it has a degree of authority because it comes directly from the people and presents an unfiltered representation of their living culture free of ideology. An appealing idea, but, as has been made clear in earlier chapters, the notion of metadata being devoid of ideology is a utopian one. Folksonomies are as ideological as any other form of metadata and what they present are beliefs about the world that are as value-laden as beliefs always are.

A core idea in social tagging is that the combined intelligence of a group of people will be more accurate than the knowledge of an individual, even an expert individual. This has undoubtful been the case, but as Gartner's quote said, it may still be as ideological as any other form of indexing. The most fruitful cases of social tagging may be

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

those, where a group of users are really experts in a field (e.g., fan-clubs for certain kinds of fiction). In such cases, users may know the content far better than any other kinds of indexers (such subject knowledge is opposed to formal subject knowledge in which you have a formal degree).

3.8.4 People with high formal level of subject knowledge

In research libraries, it often has been the norm that classification and subject indexing should be done by people with a formal degree in the subject (with or without additional formal training in LIS). Lancaster (2003, 88 emphasis original) wrote:

> Indexers should have some familiarity with the subject matter dealt with, and understand its terminology, although they need not necessarily be subject experts. Indeed, some organizations have had problems with indexers, who are too 'expert' — they tend to interpret too much and perhaps to go beyond the claims of the author (e.g., to index a possible application not specifically identified in the article) or even to exhibit prejudices by not indexing claims that they are unwilling to accept (see Intner, 1984, and Bell, 1991a, for discussion of bias and censorship in indexing). However, lack of subject knowledge may lead to overindexing. Unable to distinguish between two terms, perhaps, the indexer assigns both when only one is needed or only one is correct. Loukopoulos (1966) referred to this as *indecision*.

Lancaster's claim that too much subject knowledge may be problematic is unsupported (the references he gave (Intner, 1984, and Bell, 1991a) did discuss kinds of bias but not bias caused by formal subject knowledge). If Lancaster had provided real examples of indexing, it would have been interesting to examine them. As his claim stands, it seems problematic. If a true expert feels that a claim in a document is unsupported, it is a fine thing if that is reflected in the indexing (evidence based medicine, for example, is about how to evaluate claims by considering the research methods reported in documents). Again, Lancaster's view seems incoherent. On the one hand, he states that there is no one set of indexing terms that are the best for any target group—implying that indexing should not be just content-oriented but request-oriented. On the other hand, however, he claims that "too expert" indexers may use their knowledge to interpret what is and what is not fruitful to a given target group.

Swift, Winn and Bramer (1973) describe a research project investigating PRECIS' suitability for the indexing of documents within the sociology of education. The report concludes that PRECIS is not able to satisfy the requirements of professionals in respect to precision and validity of the indexing. One of the fundamental differences Swift et al. find between subject specialists and information specialists is that information specialists work on the assumption that efficient document retrieval presupposes the use of terms having an "agreed orientation" (12). As subject specialists in the field of sociology of education, Swift et al. maintain that such agreement does not exist at all. Instead, there are many different orientations with greater or smaller groups of adherents. In other words, the research and the documents within this research field is multi-paradigmatic. If, therefore, the indexing is made based on the assumption of some hypothetical agreed orientation (e.g., the commonsense conception of the indexers), a considerable distortion of the contents of the documents indexed will result. Swift et al. report that the research group originally worked on the assumption that the PRECIS system was a purely formal system with a developed syntax, appropriate for handling quite complex subject statements. Therefore, they endeavored to work out subject statements within the framework of this syntax. However, they arrived at the conclusion that it was PRECIS' own formal characteristics and presuppositions that prevented them from achieving satisfactory indexing results. Thus, the initial analysis and division of a subject into its basic components caused a large number of complex (many-worded) concepts to be dispersed and thereby lose their specific meaning. Similarly, they found the mechanics that reassemble the separate semantic components too restrictive, i.e., many of the semantic connections that had been severed in the analytical phase were not re-joined. For example, they missed the possibility of handling two-level structures of the type that allow one to specify the relation between a given subject and given theoretical structure, i.e., the possibility of discerning relations within a level from relations between levels.[60]

Krarup and Boserup (1982) is an empirical investigation where 100 sociological documents listed under the class 301 (sociology) in the British National Bibliography (BNB) from 1975 were selected at random and checked with the holdings of The Royal Library in Copenhagen until the number 100 was reached. Two sociologists independently assigned title-like phrases to each of these titles, based on the whole text of the documents. The 200 title-like phrases were then sent to two trained PRECIS-indexers without knowing which books they referred to or which sociologist had made them. The following datasets were compared:

B: The indexing string as it appears in BNB 1975 for the given document (The PRECIS string produced by indexers without specialized knowledge of sociology).
S1: The title-like phrase produced by sociologist 1 after full-text inspection

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

S2: The title-like phrase produced by sociologist 2 after full-text inspection
I1S1: The PRECIS-transformation of S1 made by indexer 1
I1S2: The PRECIS-transformation of S2 made by indexer 1
I2S1: The PRECIS-transformation of S1 made by indexer 2
I2S2: The PRECIS-transformation of S2 made by indexer 2

Krarup and Boserup found that here was less difference between the sociologists than between them and the BNB, and the analysis of the partial overlaps found that the percentage of inclusive overlap was considerably higher when the BNB-strings were compared to the strings based on the subject statements, than when these were compared with each other. Further it appeared that there was an unduly high amount of inclusive overlaps when one compared the indexers' interpretations of the sociologists' title-like phrases. This high amount of inclusive overlaps corresponds to the amount of generalizations in contents observed in connection with the comparisons.

The study aimed at providing an answer to the question: "Should subject specialists be employed for the indexing of social science literature?" And the answer was: "Yes, in view of the results obtained, this appears to be the case." This is one of very few investigations of the importance of formal subject-matter knowledge in indexing. Although its methodology may be questioned, as long as other studies do not exist, its findings are the best knowledge we have. Unfortunately, it has been overlooked by Lancaster (2003) and most of the literature on indexing.

3.8.5 People with a high degree of subject knowledge trained as professional indexers

Above, we saw that sociologists seem to perform better indexing compared to people without formal education in sociology. However, how can we know that other sociologists would not just index according to one narrow view of sociology? The best thing seems to be a combination of subject knowledge, adequate training in indexing theory and practical training. The demand of advanced medical indexing provides partly such a perspective (National Library of Medicine, 2018):

> Most MEDLINE indexers are either Federal employees or employees of firms that have contracts with NLM [the National Library of Medicine] for biomedical indexing. A prospective indexer must have no less than a bachelor's degree in a biomedical science. A reading knowledge of certain modern foreign languages is typically sought. An increasing number of recent recruits hold advanced degrees in biomedical sciences.
> …
> Indexers are trained in principles of MEDLINE indexing, using the Medical Subject Headings (MeSH) controlled vocabulary as part of individualized training. The initial part of the training is based on an online training module (partially available to the public at //www.nlm.nih.gov/bsd/indexing/index.html)[61], followed by a period of practice indexing. NLM does not accept other indexing training programs as a substitute."

If optimal indexing depends on criteria of relevance dependent on paradigmatic issues in the sciences, then research and education in this must lead the way in indexing theory and indexer education. Subject knowledge is highly relevant but must me combined with knowledge about epistemological issues in the domain. "The development of information services should be an active and on-going partnership between those who possess the skills of information science and those who understand the discipline within which the documents are written" (Watson, Gammage, Grayshon, Hockey et al. 1973, 273). Information science should aim at providing general knowledge that is found useful for high-level professional indexing, such as MEDLINE.

3.9 Algorithmic indexing (and algorithmic ideology)

Indexing today is very often large-scale indexing made by search-engines, algorithms and other forms of automatic indexing. They play very important roles today, not just because of problems of scale of available information, but also because of inherent in-human-based indexing as discussed in relation to the lack of inter-indexer consistency described in Section 3.2 (although it follows from the principle of request-oriented indexing that there is no one right way to index a given document or collection). However, this is not just solved by applying automatic indexing methods.

Because algorithms will provide the same results given the same input, they are sometimes wrongly understood to be "objective." However, repeatability should not be confused with objectivity. In the words of Dirk Lewandowski (2015):

> The presentation of a certain set of results is what I call an algorithmic interpretation of the world, that is, the web data. However, people often assume that search can produce right and wrong results. They think that, if a search engine has found the "magic formula," it can provide its users with the best possible results. But there is no such thing as a "right" results ranking (as opposed to a "wrong" results ranking). At least for informational queries there are

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> often hundreds if not thousands of relevant results. The goal of the search engines in these cases is not to provide a certain set of right/relevant results, but to list some of the potentially relevant results in the top few positions. (278)
> ...
> we can see that searches are always biased, and there is no such thing as an unbiased search engine. It would be impossible to construct such a search engine, because human beliefs and assumptions influence the design of algorithms, and they therefore prefer certain documents to others. It is even at the core of every idea of ranking that, based on certain technically mediated assumptions, certain items are preferred over others (279)

The bias of search engines is partly due to its indexing of the web, partly to its matching techniques. A search engine does not normally have a certain worldview or ideology. However, in the end, a given algorithm is based on a range of choices that have implications for what is made relatively more visible and what is made less visible; it has ideological implications (whether the programmer is aware of this or not). Therefore, it is fruitful to understand a search engine as a cultural-political agent and an epistemological agent. An example of a cultural-policy choice is the downgrading of pornographic results in later generations of search engines. An example of epistemological policy is information about rare diseases in which it has been found that Google is not performing in an optimal way, because Google is meant to serve common queries rather that rare ones (Dragusin et al. 2013a+b). There exists now a broader literature about algorithmic ideology and the political issues related to search engines, including Diaz (2008), Granka (2010), Hargittai (2007), Introna and Nissenbaum (2000), Mager (2012) and Pariser (2011).

It is important to know which kinds of queries are relatively bad served by a given algorithm, and such an analysis is a pragmatic analysis more than anything else. Correspondently, the analyses of goals, values and consequences is core issue for design of algorithmic ways of organizing knowledge.

3.10 Conclusion of section 3

No index can provide perfect subject retrieval (100% recall and 100% precision), but clearly some indexes are better than others. The most obvious quality criterion is whether all relevant concepts can be looked-up in the index (having to do with be the exhaustivity of indexing and the specificity of the index language, cf. Section 4.2 below). However, it is not just a matter of what percent of the queries can be answered in a satisfactory way—as the traditional thinking in information retrieval take as the basis for improvement. It is also a question of perspectivism[62]—which kinds of queries are relatively well served, and which kinds are relatively badly served—this being an implication of the view discussed above that there is, therefore, no single correct set of index terms for a document. No indexing can, therefore, be neutral in respect of perspective.

In the literature of indexing (and LIS in general), there has been a tendency to consider either the documents or the users. However, criteria for indexing are to be found in a third place: in epistemological theories. Take evidence-based medicine (EBM) as an example. The criteria of what counts as valid medical knowledge about medical interventions are connected to the methodology used in the documents (with randomized controlled trial as what is generally considered the most important). Given this theory, it provides criteria on which documents are most important to retrieve (i.e., those based on a solid methodology) and thus also criteria for how they should be indexed (partly by methodological criteria). The same argument is valid in all domains of knowledge, although they are seldom made explicit as in EBM (and even here, they are debatable). The uncertainty of guidelines for indexing are therefore primarily caused by uncertainties in epistemological theories about what counts as knowledge. The core of indexing is, as stated by Rowley & Farrow (2000, 99) to evaluate a paper's contribution to knowledge and index it accordingly[63]:

> In order to achieve good consistent indexing, the indexer must have a thorough appreciation of the structure of the subject and the nature of the contribution that the document is making to the advancement of knowledge.

But again, there often are different views of what a contribution to knowledge is, and in what way a given document contributes (or does not contribute). This does not, however, make everything as relevant as anything else; there are better or worse ways to index documents.

Another important aspect of indexing theory is the issue related to the users' selection power (versus the dominant tendency to transform queries automatically into ranked sets of relevant documents) (cf. Hjørland 2015b; Warner 2010). Algorithms and search engines are also based on the subjectivity of their programmers and the tools available; it is a myth that computers are objective whereas human indexing is subjective. All indexing is based on theoretical assumptions and interests, but the more selection power delegated to users, the better the chance of modifying search strategies to find the most important documents.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

4.0 Indexing languages (metadata systems and knowledge organization systems)

4.1 The concept "indexing language"[64]

The concept "indexing language" arose in information science in relation to representing documents in bibliographic databases and to evaluate the relative strengths of different kinds of systems (e.g., in Cleverdon and Mills 1963). It flourished in the 1970s (see, e.g., Foskett 1970; Soergel 1974; van Rijsbergen 1979), but it is still important by providing an overall perspective from which to consider indexing (although today the broader terms "metadata"[65] and "knowledge organization system" (KOS) (http://www.isko.org/cyclo/kos) are more used)). Soergel (1974, 27-28 emphasis original) defined:

> "*Indexing language*" (*documentary language*) as used in this book = any language (broadly defined) for the representation and/or for the arrangement of retrieval objects and/or their substitutes with the objective of making the items retrievable.

While van Rijsbergen (1979, 13) defined:

> An index language is the language used to describe documents and requests.

Soergel explicitly includes classification schemes as kinds of indexing languages.[66] The similarities between indexing and classification was developed by Lancaster (2003, 20-21)[67] as well as Anderson and Pérez-Carballo (2005, 413). The latter state:

> § 49 Definition of classification: Literally, classification simply means the creation of classes, and places objects or concepts into these classes. At this level, there is no difference between indexing on the one hand and classification or classifying on the other. In indexing, terms are extracted or assigned to a message, and in so doing, the indexer creates a class for the concept named by the term and links the message to this class. The process is exactly the same when a message is classified.
>
> §50 Classification versus indexing: In the concept of indexing, there is no clear indication of how the resulting index terms or headings should be arranged for consultation, but there is a common expectation that index terms should be arranged in some alphanumeric order. Similarly, there is a common expectation that the classes created in classification should be arranged in an order other than alphanumeric. The common dictionary definition of "classification" suggests a "systematic" arrangement (Webster's 1966).

Both these books considered classification schemes as kinds of controlled vocabularies and their notations as kinds of indexing languages. Figure 1 is a model of different indexing languages discussed in relation to classical bibliographical databases (including classification systems).

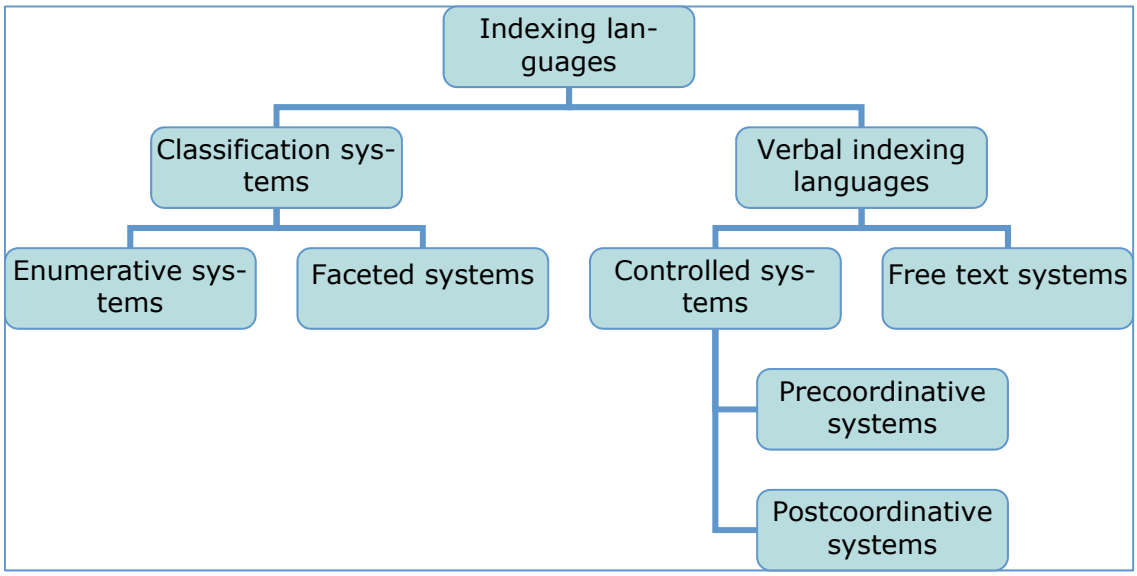


Figure 1. The traditional view of the kinds of indexing languages (after Hjørland 2012, 304; see also the more elaborated model provided by Chatterjee 2016, 138).

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

All kinds of indexing systems in Figure 1, apart from free text systems, represent different kinds of controlled vocabularies. A controlled vocabulary is a given set of terms or classes on which the indexer must base the indexing, whereas a free text system is one in which the indexer is free to use any term she prefers. A controlled vocabulary determines which term in a set of synonyms should be used ("preferred term"). Free text systems are today strongly represented in folksonomies (tagging systems) (see Rafferty 2017). The concept of controlled vocabularies is important and will be given an independent entry in this encyclopedia, but it should be said that research suggests that free-text and controlled vocabulary searches usually each provide unique hits and that each work better for different kinds of searches."[68]

It should be said that indexing often uses more than one indexing language. For example, in classical databases like PsycINFO, indexing is primarily done by using the PsycINFO thesaurus and add "descriptors" to a devoted field. Besides, uncontrolled terms, "identifiers" are added in another devoted field. In addition, all documents are indexed by "classification codes" and after about 2000 records are also indexed by their bibliographical references ("citation indexing").

Today the term "metadata" is much more popular than indexing language[69] but also much broader (including for example data for managing intellectual property rights). (See also Section 4.3 below). This encyclopedia will treat metadata in an independent article. The term knowledge organization system (KOS) (http://www.isko.org/cyclo/kos) (see Mazzocchi 2018) is also a broader term often today used as a near-synonym for indexing language.

4.2 Some characteristics of indexing languages and their use

Some important concepts related to indexing languages and their use are: depth of indexing, exhaustivity (density), specificity[70] and granularity. They are not always understood the same way in the literature. Browne and Jermey (2007, 29 emphasis original) suggested the following definitions:

- *Depth* of indexing is the degree to which a topic is represented in an index and depends on a combination of exhaustivity and specificity.
- Exhaustivity refers to the number of terms representing a document in an index. A fully exhaustive index includes entries for all of the concepts that have been identified, while a less exhaustive index only covers the main topics.[71]
- *Specificity* refers to the exactness of match between the indexing term and the concept being indexed. If you index the concept of 'image indexing' using the term *multimedia indexing*, you do not have specificity, as the term is broader than the concept. If you use the terms *picture indexing* or *graphics indexing* for the concept "image indexing," your specificity is closer. If you use the term *image indexing*, your specificity is perfect. Specificity can be achieved by using specific terms, general terms with subdivisions, or a number of controlled vocabulary terms in combination. When using a thesaurus to index the topic 'children's computer games' you may have to use the terms *games for children* and *computer games* to achieve specificity.[72]
- *Granularity*[73] is a measure of the depth of indexing, and refers specifically to the size of the indexable units. A book index is usually more granular than a periodical index, as it refers to small chunks of information the size of a paragraph or even a sentence. A periodical index, on the other hand, generally refers to a whole article or section.

Henttonen (2015) suggested five dimensions of classifications:

- Stability (need to change and update the classification);
- Generality (number of contexts it covers);
- Granularity (number of sub-divisions and subhierarchies);
- Specificity (exactness of description);
- Validity (classification's power to describe and predict features of context).

Such technical attributes are, however, only one aspect of indexing languages. Just as important is the underlying assumptions concerning terms having an "agreed orientation" versus representing different voices presented in section 3.8.4. The dominant view in indexing theory and practice seems to be associated with the idea of "ideal language," a neutral language free of ambiguity and able to represent all knowledge, even what has not yet been produced (cf. Laporte 2018). Such a view represents a rationalist view that is opposed to pragmatic and critical views of language.

4.3 Metalanguages versus object languages

Warner (2004, 112) wrote:

> The distinction meta:object has become increasingly familiar from the contrast between metadata and the objects described …, although the term object-data is not necessarily used. In that context, the term metadata is used to indicate a concise and deliberately

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> encoded description of a set of objects. In wider, although not necessarily ordinary usage, a metalanguage is congruently understood as a language that supplies terms for the analysis of an object-language or as a system of propositions about other propositions.

To exemplify (not necessarily in accordance with Warner), a given text may describe a bird (=object language). This text may itself be described in a bibliographical database (representing a description in a meta language). The first description may be termed the object description; the second may be described a meta-description. The object-language description may contain, among other attributes of birds:

- Body size
- Beaks and feet
- Skin and feathers
- Mouth and digestive system

The meta-language description may contain, among other attributes of documents about birds:

- The kind of document containing the description) (student essay, amateur ornithologists' publication, scientific journal); whether the document is illustrated or not and if the illustrations are in color or monochrome;
- A description of which elements were contained in the object description (body size, beaks and feet …).

It follows that a theory of indexing in LIS must be based on a theory of meta-descriptions. The first principle based on the theory of knowledge is that both kinds of descriptions are "subjective" in the sense that: 1) they are necessarily incomplete; and, 2) that they are based on the tradition and paradigm (for example, recently DNA-analysis have become an important part of the description of birds and made a revolution in biological classification). Both the object-language description and the meta-language description are based on underlying epistemological views (such as empiricism or historicism) whether this is recognized and made explicit or not. The nature of the object description in source documents is of course relevant for users to evaluate its relevance. Therefore, an important function of meta-descriptions is to provide information about the nature of the object description. This means, first of all, epistemological and methodological descriptions.[74]

We see from the example, that a full text representation of the object description does not necessarily contain the same information as required for the meta-language description. In other words, kinds of added information are necessary. This is obvious when the document is a picture, but in principle this is also relevant for texts. Examples of valuable meta-descriptions include genre-classifications and methodology classifications. Real-life information systems today often include object-language descriptions in the form of natural language representation in addition to meta-data representations. Object languages and meta-languages are often very similar—the same (natural) language may be used for both purposes.[75] However, the point made by Warner (2004) and this article, is that from the perspective of information retrieval and KO, it is fruitful to make this distinction. The study of meta-language is perhaps less related to first- and "second-order" logic as Warner suggested, and more to the studies of genres, epistemologies, methodologies etc. in different domains.

5.0 Conclusion

This article has argued that theories of knowledge (= epistemological theories)[76] provide a deep and systematic classification of indexing theories. It has also argued that social epistemological theories provide a better foundation for indexing compared to individualist epistemological theories. The term "social epistemology" was first used in a paper by Jesse Shera (1951), a fact that we in library and information science should be very proud of. Shera's concept has not yet been fully recognized and the term was (and still is) ambiguous. However, as applied by Shera (1951), the term seems fully in accordance with the way it has been used in the present paper: that indexing is always influenced by socio-cultural contexts and that these contexts need to be considered in indexing practice and theory, and that criteria for classification and indexing must be found in pragmatic criteria.

Notes

1. A term is a word or phrase used to refer to something in a particular branch of study. Examples of terms in library and information science are "index," "indexing language," "controlled vocabulary" and "knowledge organization."
2. Indexical by Peirce is a mode in which the signifier is not arbitrary but directly connected in some way (physically or causally) to the signified. Consider that most indexes use words (which are symbols, "not" indexes in the meaning given by Peirce, 1931-58, 2.306). Indexes in library and information science thus does not apply this term as understood it was used by Peirce.
3. Weinberg (2017) referred to other uses or misuses of the term "index" and criticized the tendency to use vogue words rather than scientific terminology (page 1980). She also wrote (1978): "In discussing the his-

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

tory of the word index, Wellisch [1996, 199–213] noted that the term was once used to refer to many types of information structures: abstracts, titles, etc. The word index is still encountered today as a heading for tables of contents, although these are far less sophisticated information structures than indexes. A table of contents lists the chapter titles and headings of a systematically organized work; such a list can be generated automatically. An index, in contrast, provides efficient access to the specific topics covered in a document. Compilation of an index requires far more cerebration than does preparing a table of contents."

4. The concept document has its own entry in this encyclopedia (Buckland 2018); see in particular Section 3: http://www.isko.org/cyclo/document#3.
5. Mulvany provided this link: http://www.asindexing.org/site/WilsonAward.shtml#awcrit (accessed July 2007). The link is now dead, but a stored version is available in WebCite: http://www.webcitation.org/6x3N0YiJ9
6. (The same definition was also given by Weinberg 2010, 2277). Consider, however, that in some cases, the index does not provide an alternative order. An encyclopedia, for example, may have an index in which the index terms are presented in the same alphabetical order as the articles in the encyclopedia. The function of the index is here to have a higher level of "indexing depth" (determined by "exhaustivity" of indexing and "specificity" of terms). Another example is biographical dictionaries, which presents named persons in alphabetical order. An index such as *Biography and Genealogy Master Index* provides alphabetical access to named persons in many biographical dictionaries. Here, the order is not new, but the index cumulates a long range of entries from different sources. By implication, it is, in principle, difficult to distinguish between an index and a table of content, although, usually, the index provides an alternative order in addition to a much greater depth of indexing.
7. The opposite of an embedded index has been termed "a stand-alone index" (Brenner and Rowland 2000).
8. From the users' perspective this relation may be reversed: an index is a source used to identify some target documents or information within one target document. However, because an index is in some way derived from source documents, the terminology applied in the article is preferred.
9. Niels Ole Finnemann suggested other metaphors for indexes: the hypertext concepts of "anchor" and "destination" and programming concept like "address system" and "go-to" function (and included "to do" function). He emphasized the dynamic nature of many e-sources. He wrote in an informal communication dated 2018-04-09 (here translated to English): "My point of view is that theoretically, more complex forms of knowledge should be the point of departure to understand less complex systems—an evolutionary, theoretical perspective. In that perspective, the kinds of indexes you describe can be understood as variants of proto hypertext-systems. They are based on a 1:1 correlation between a series of anchors and a series of destinations. When moving from proto hypertext to hypertext the complexity increases. It also increases when we mechanize not just the operation but also introduce editable instructions for action. It increases another level when we come to networks-based hypertext. There are a range of implications, the theories of indexing need to be thought in more dimensions, but also open the black boxes of search engines, e.g. the ranking principles of Google, the continuous modification and search in multisource knowledge systems" (see also Finnemann 2018).

   It should be said, however, that the present article emphasizes the knowledge, theories and interests (subjectivity) of the indexers and programmers to provide criteria on how to understand and evaluate indexes. In Finnemann's universe of dynamic documents, constantly rewritten by multiple actors, these are mixed in ways that seems to make such an analysis impossible. Also, the claim that the present article is based on a one-to-one relation between anchor and destination should be considered in the perspective of multi-paradigmatic fields, as discussed in, for example, Section 3.8.4.
10. Consider that the term "indexing" is also used about assigning classification codes to items.
11. See, for example, Mulvany (2010).
12. Fugmann (1997, abstract): "Traditionally, database indexing and book indexing have been looked upon a s being quite distinct and have been kept apart in textbooks and teaching. The traditional borderline between both variations of indexing, however, should not conceal fundamental commonalities of the two approaches. For example, *thesaurus construction and usage*, quite common in databases, has hardly been encountered in book indexing so far. Database indexing, all the other hand, has hardly made use of *subheadings* of the syntax- displaying type, quite common in book indexing. Most database users also prefer *precombining vocabulary units* and reject *concept analysis*. However, insisting on precombining descriptors in *a large database vocabulary* may, in the long run, well be destructive to the quality of indexing and of the searches. *A complementary approach* is conceivable which provides both precombinations and analyzed subjects, both index language syntax and subheadings, and provides access to

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

an information system via precombinations, without jeopardizing the manageability of the vocabulary. Such an approach causes considerable costs in input because it involves a great deal of intellectual work. On the other hand, much time and costs will be saved *in the use* of the system. In addition, such an approach would endow an information system with survival power" (italics in original).

13. About still image indexing, see, for example, Jörgensen (2017).
14. About moving image indexing see, e.g., Turner (James 2017).
15. See, for example, Keyser (2012, 113-120): "Chapter 6: Automatic Indexing of Music."
16. See, for example, Rafferty and Hidderley (2005) and Rasmussen Neal (2012).
17. About Web-indexes see Hedden (2007), Keyser (2012, 195-219: 11: Indexing the Web) and Lewandowski (2014). There is also a journal devoted to this subject: *Journal of Internet Cataloging* (1997 - 2007), from 2008 named *Journal of Library Metadata*. See also Hedden Information Management: Web Site Indexing. http://www.hedden-information.com/web-site-indexing.htm. WebCite archived version: http://www.webcitation.org/6wxGk9RCU
18. As an example of a controlled vocabulary developed for cataloging museum objects, see Bourcier, Dunn and the Nomenclature Task Force 2015.
19. See, for example, Towery (1998) about indexing in history.
20. See, for example, Wyman (1999) about indexing in medicine.
21. See, for example, Kendrick and Zafran (2001) about indexing law.
22. Anderson and Pérez-Carballo (2005, 111) defined the term: "Indexable matter is the part or portion of documentary units that is actually considered in the indexing process, whether that process is performed by humans through intellectual analysis of the message content and meaning or by machines through manipulation and analysis of textual symbols. Indexable matter is also called the 'analysis base' because it provides the base or basis for the analysis of a message and its text." Their chapter seven (111-116) is devoted to this concept, and they also present "complete texts versus partial texts as indexable matters" (but seems to underestimate the role of the first in modern computer-based indexing).
23. Hjørland and Kyllesbech Nielsen (2001) used the term "access point" (SAPs, search fields). This concept corresponds to the above definition of "indexing matter," the last term provides the indexers' perspective while SAP provide the searchers' perspective on the same parts of documents.
24. Concordances are kinds of full text indexes. *Oxford English Dictionary* defines "concordance" as "6b.: An alphabetical arrangement of the principal words contained in a book, with citations of the passages in which they occur." Wikipedia, however, claims "A concordance is more than an index; additional material make producing them a labor-intensive process, even when assisted by computers, such as commentary, definitions, and topical cross-indexing" (from https://en.wikipedia.org/wiki/Concordance_%28publishing%29 2018-01-06).
25. An N-gram or ngram is a sequence of tokens, usually words, but they can also be characters. N refers to the number of tokens. An N-gram is thus an N-character slice of a longer string. In general, a string of length k, padded with blanks, will have k+1 bi-grams, k+1tri-grams, k +1 quad-grams, and so on (cf., Cavnar and Trenkle 1994). An index based on ngrams will include misspellings and gibberish.
26. Anderson and Pérez-Carballo (2005, 546): "'Descriptive cataloging' is an old and honorable term that refers to the description and indexing of texts and documents with respect to features other than the content, purpose, or meaning of the text. Such features include the authors and other creators of texts (editors, composers, illustrators, translators, artists, etc.); the names or titles of texts (including subtitles, parallel titles, alternative titles, running titles etc.); the publisher or manufacturers and distributors of documents, the size and medium of documents; and the symbol set or code used to encode the text. Codes and symbols used to encode texts include natural language and their writing systems (French, German, Chinese), but also codes and symbols for music, dance, chemistry, mathematics, etc., and, at another level, codes for the representation of messages in digital media, Names and index terms are established for the most important of these features. Descriptive cataloging (along with subject cataloging) is part of the process of making a catalog. 'Descriptive indexing' is a rarely used term for the same process outside of the context of catalogs for particular collections of documents."
Wilson (1968) described two distinct kinds of bibliographic control. "Descriptive control" provides the means, traditionally by cataloging, to create lists that enable retrieval of all the entities characterized by certain attributes (e.g., written in the same language or by the same author). "Exploitative control," in contrast, is the ability to procure the best entities available serving a specific purpose. The first kind of power was regarded evaluative neutral (Wilson 1968, 22) while the second involves appraisal (Wilson, 1968, 22). Wilson considered exploitative control the most im-

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

portant form, but descriptive control being a precondition for achieving exploitative control; to identify the best entities, these entities must be known, and to be known, they must be described.

27. Klement (2002) made a distinction between "open-system" versus "closed-system" indexing (also discussed by Mulvany 2010, 485). A back-of-the-book index is considered an example of a closed system index while a journal index or an internet search engine are considered examples of open-system indexes. Mulvany (2010, 485) wrote: "Internet search engines are examples of open-system indexing. An index to a periodical such as Forbes is also an open-system index. The Internet grows every day; it also changes every day as Web pages are edited or removed. Forbes will continue to add new issues as long as the magazine is published. Open systems are in a state of flux. A book is a closed system. There is a beginning, a middle, and an end. The book and its audience is a self-contained universe. Unlike an open-system index that casts a wide, but shallow net, a book index deeply presents a systematic guide to the information contained in a text in a manner that enables readers to quickly find specific topics or concepts." One might say, of course, that a bound volume of a journal is a closed system like a book. However, this argument seems unimportant, because journals are mostly indexed continuously, not as bound volumes. Lancaster (2003, 37) briefly discuss this distinction and wrote: "When indexing applies to many items, and is continuous, the terms used in index entries must be standardized. Standardization is not really an issue in closed-systems indexing although it is obviously necessary to use consistent terminology throughout the single index. Closed system indexing may use terms that are non-continuous. "Leonardo da Vinci, dies" may be perfectly appropriate in such an index but is unlikely to appear in an open-system index (although ""Leonardo da Vinci" would)." An alternative terminology, cf. Finnemann (2018), is the differentiation between "finite texts" and "non-finite texts."
28. Systematic indexes are seldom in library classifications. The Danish Decimal Classification fifth edition (a Danish version of the *DDC*) has, however, a systematic index in a separate volume in which all subject terms associated with a particular class number are listed.
29. Chu (2001, 1011) wrote: "Two distinctive approaches, content-based and description-based, have been applied by researchers over the years for conducting studies in image indexing and retrieval. Simply speaking, the content-based approach refers to the techniques of indexing and retrieving images based on automatic processing of textual information, as well as of the image itself. Image properties analyzed via the content-based approach can be divided into three levels: 1) primitive features such as color, shape, and texture; 2) logical features such as the identity of objects shown; and 3) abstract attributes such as the significance of the scenes depicted [reference omitted]. The description-based method, on the other hand, manually employs captions, keywords, and other descriptions (e.g., artist and work size) of image data for indexing and retrieval purposes. Human beings are directly involved in the image indexing and retrieval process when using this approach. In addition, it appears that people in the field of computer science focus on the content-based approach, while the information science community, including library science, concentrates on the description-based method [references omitted]. However, is the division as evident and clear-cut as it has been perceived?"
    In this quote, "content-based" and "description-based" are considered the basic approaches. However, "content-based" also has another meaning and "description based" is sometimes referred to as "concept based," which may be a better term, as will now be explained. There are several issues to be distinguished: 1) the index may or may not use the same kinds of signs as the documents indexed (e.g., words or colors, sounds); and, 2) the signs used in the index may be derived from the source documents or they may be assigned. There are many kinds of both kinds. Derived indexing depends on signs in the documents, whereas assigned indexing may use the same kinds of signs or may use other kinds of signs. Assigned indexing of texts and pictures often uses words (or terms, including multiword strings); in the text case, this the same kind of signs as the indexed documents, but in the picture the verbal index these are different kind of signs. Assigned terms mostly represent concepts and, therefore, "concept based" should be preferred in the cases in which words or other signs are not just mechanically derived but represents the indexers conception of what is being indexed.
30. When no association exists between terms, information scientists speak of "bag of words." *Preserved Context Indexing* System (PRECIS) is an example of an indexing system based on syntax (see Austin 1974). Literature about syntax, roles and links include Spang-Hanssen (1976) and Svenonius (2003). Soergel (1974) suggested an alternative terminology in which "precombination" replaces "precoordination" and "postcombination" replaces "postcoordination."
31. String indexes are mentioned by Anderson and Pérez-Carballo (2005, 231, §118ff): §118: "String syntax is the

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

modern version of subject headings, inspired by the desire to take advantage of computer technology for the creation of headings. Because the instructions for the combination of terms into headings are programmed for the computer, string syntax tends to be much more regular than the idiosyncratic variety exhibited by subject heading syntax." §119: "The name "string syntax" or "string indexing" comes from the custom of displaying headings as "strings" of terms—terms strung together in various configurations. The variety of string systems approaches is mostly related to how terms are arranged in these strings." See also Craven (1986).

32. Bates (1988, 47; italics in original) wrote: "The idea [of a subject heading] is to describe the whole document in that one heading, or, at the most, in a handful of such headings" whereas "individual descriptors [in post-coordinate systems] were intended to describe *a single concept used within a document*, rather than the whole document ⊢ hence the phrase 'concept indexing.'"
However, both kinds of systems are designed to describe whole documents. In subject heading systems this is done by a string of terms pre-coordinated by the indexer. In the descriptor system this is done by a set of descriptors which together describe a whole document. (Parts of documents rather than whole documents were by Ranganathan (1963, 29) termed micro-documents: "322 Micro document-Document embodying micro thought, usually forming part of a host document." Today the term "passage retrieval" is the most common term for systems intended to retrieve parts of documents, cf. Kaszkiel, Zobel and Sacks-Davis 1999 and Shepherd 1981).
33. Anderson and Pérez-Carballo (2005), chapter fifteen, is devoted to the concept of locators. Their definition is (373): "Locators are devices that link or lead a user from a surrogate to a message, text, and documentary unit or to a larger surrogate. Locators locate the desired item. Locators are essential elements for all surrogates. Full surrogates were discussed in the previous chapter. The staged display of surrogates will be discussed in the next chapter."
Weinberg (2007) is about unusual locators in indexes.
34. A well-known example of a relative index is the one belonging to the *DDC* (see Miksa 2006).
35. Indexing software may accomplish many practical functions for indexers, kinds of specialized word processing tasks for entering index headings, subheadings and page references, and organize them in alphabetical order or other orders, and checking for cross-references. However, in this article, we are focusing on indexing theory. When software claims to analyze and index text automatically, without human intervention, its principles are identical with automatic indexing. Browne and Jermey (2007, 175) wrote: "These are structural features of the text (capital letters, words repeated several times) to try to make semantic judgements, identifying 'important' words and phrases and then listing these in alphabetical order with page numbers attached. The results are usually unimpressive for books and journals, but in large collections indexes automated programs may have a role to play." Browne and Jermey 2007 then mentions specific indexing software packages and their reviews.
36. Stock and Stock (2013, 760) use the term "the text-word method" for what is here called derived indexing.
37. The 20% rule is used by, for example, the Library of Congress (LC): "Assign to the work being cataloged one or more subject headings that best summarize the overall contents of the work and provide access to its most important topics." LC practice: "Assign headings only for topics that comprise at least 20% of the work." (Library of Congress. 2008; The Subject Headings Manual, sheet H 180). See also Mai Chan (2005, 188-189).
38. Lancaster (1991, 8) wrote: "Effective subject indexing involves deciding not only what a document is about but also why it is likely to be of interest to a particular group of users ... The same publication could be indexed rather differently in different information centers and should be indexed differently if the groups of users are interested in the item for different reasons."
39. Human based indexing is also called "intellectual indexing" or "manual indexing."
40. See further about IFLA's principles in Žumer (2017; this encyclopedia)
41. Logical positivism, for example, represents an important, but failed attempt to combine empiricism and rationalism. Smith (1986, 64) wrote "logical positivism arose as the joint product of two intellectual traditions that conflicted deeply with one another [empiricism and rationalism]: In attempting to unite these traditions, its adherents created an extremely influential approach to philosophy but one that embodied serious intellectual tensions from its dual ancestry."
42. Haider (2018) described Ranganathan's indexing method, "chain indexing": "Chain Indexing or Chain Procedure is a mechanical method to derive subject index entries or subject headings from the class number of the document. It was developed by Dr. S.R. Ranganathan. He first mentioned this in his book "Theory of Library Catalogue" in 1938. [Ranganathan 1938].
In Chain Procedure, the indexer or cataloguer is supposed to start from where the classifier has left. No duplication of work is to be done. He/she has to derive subject headings or class index entries from the

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

digit by digit interpretation of the class number of the document in the reverse direction, to provide the alphabetical approach to the subject of the document. Ranganathan designed this new method of deriving verbal subject heading in 1934 to provide the subject approach to documents through the alphabetical part of a classified catalogue. This method was distinctly different from the enumerated subject heading systems like Library of Congress Subject Headings (LCSH) or Sears List of Subject Headings (SLSH). He discerned that classification and subject indexing were two sides of the same coin. Classifying a document is the translation of its specific subject into an artificial language of ordinal numbers, which results in the formation of a class number linking together all the isolate ideas in the form of a chain. This chain of class numbers is retranslated into its verbal equivalent to formulate a subject heading that represents the subject contents of the document. The class number itself is the result of subject analysis of a document into its facet ideas and linked together by a set of indicator digits, particularly when a classification system like Colon Classification is used for the purpose. As this chain is used for deriving subject entries on the basis of a set of rules and procedures, this new system was called "Chain Procedure." This approach inspired in many other models of subject indexing developed afterward, based upon classificatory principles and postulates. Chain Indexing was originally intended for use with Colon Classification. However, it may be applied to any scheme of classification whose notation follows a hierarchical pattern."

43. "Essentialism is a standard philosophical view about natural kinds. It holds that each natural kind can be defined in terms of properties that are possessed by all and only members of that kind. All gold has atomic number 79, and only gold has that atomic number. It is true, as well, that all gold objects have mass, but having mass is not a property unique to gold. A natural kind is to be characterized by a property that is both necessary and sufficient for membership." (Sober 2000 , 148)
44. Fugmann's (1985) five axioms were: 1) Definability of topics for search and indexing in terms of concepts and concept relations; 2) order: any compilation of responses relevant to a topic is an order-creating process; 3) sufficient degree of order: the demands made on the degree of order increase as the size of the collection and/or the frequency of searches increases; 4) representational predictability: the accuracy of any directed search for relevant texts (especially the recall ratio) depends on the predictability of the modes of expression for concepts and concept relations in the search file; and, 5) representational fidelity: the accuracy of any directed search for relevant texts (especially the precision ratio) depends on the fidelity with which concepts and concept relations are expressed in the search file.
45. Hjørland (2017a, Section 4.2c) identified Ranganathan with rationalism, and Fugmann explicitly worked in the tradition of Ranganathan. A reviewer commented: "It is hard to connect determinate essence of documents with rationalism, it is to the same extent linked to empirical methods." However, because empirical methods are per definition a posteriori, an essence cannot be considered given. Also, different empirical methods may claim different essences and therefore the concept loses its meaning.
46. Anderson and Pérez-Carballo (2005) is a very valuable compendium on indexing research. Although the book does not develop or defend a theoretical view, it is clear that it is positive towards sociological, epistemological and critical approaches without neglecting the more technological, empirical and rationalist positions.
47. The view that no set of index terms can ever be considered the single correct set corresponds with "request-oriented indexing" (or, as suggested by Hjørland 2017, Section 2.4: "policy-based indexing"). The opposite view is "document-oriented indexing" (or content-oriented indexing) according to which a document contains a certain number of subjects, which should be represented by the index. According to the document-oriented view, a document contains one or more subjects, but according to the policy-based view, a document is attributed one or more subjects in order to facilitate certain uses of that document. A typical example of a document-oriented approach is the 20% rule used by Library of Congress "Assign headings only for topics that comprise at least 20% of the work" (see note 37). The policy-based principle, on the other hand would allow indexing of a topic that covers less than 20% if it is supposed that this would be relevant to meet needs, that would otherwise be difficult to satisfy. It is worth noting that automatic indexing mostly is less document oriented, because it selects the terms not just by their frequency in a given document but also in adverse relation to its occurrence in the selection as a whole (but this does not make automatic indexing policy-based).
48. In addition, Lancaster's criticism of not considering economic realities seems misplaced. Clearly, economic considerations are important for application, but a fundamental understanding of the nature of indexing and the criteria for optimal indexes, is a matter of basic research and is important, not just in itself, but

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

to evaluate different systems and to understand how to make compromises and economic feasible solutions. Fugmann's view were considered "extreme" by Lancaster (150), but they represent a well-developed and coherent theoretical view and should be met by alternative well-developed views.

49. Because there are many variations in how inter-indexer constancy is measured, results should be taken with much reservations.
50. Frohmann's article presented also other arguments: "Second, mentalism conceals legitimate rules formulated in disciplines outside the mentalist paradigm." "Third, mentalism conceals the text. Should mentalism even admit the independent existence of the text itself, it rarely considers it to be identical with what it takes to be the true object of inquiry, its representation in the mind of the reader." "Fourth, mentalism conceals relations between texts. Like intratextual criteria of significance, intertextual criteria may also reveal structures providing a basis for indexing rules." "Fifth, mentalism's focus on processes occurring in minds conceals the crucial social context of rules."
51. Nickles (2005) wrote: "In the twenty-first century nearly everyone is an empiricist in the everyday sense of taking experience seriously as a basis for knowledge claims about the natural world and human behavior, but most philosophers reject traditional, doctrinaire empiricism-the view that human sense experience provides a special connection of the knowing mind to the world and thus provides a foundation on which knowledge can build, step by step." Nickles lists the following challenges that changed or ousted classical empiricism:

    1) The linguistic turn;
    2) The holistic turn;
    3) Rejection of the analytic-synthetic distinction;
    4) Rejection of the scheme versus content distinction by Donald Davidson;
    5) Rejection of the correspondence theory of truth;
    6) Rejection of the linear-foundational model of justification;
    7) Anti-Kantian Kantianism;
    8) Rejection by Karl Popper (1902-1994) and the positivists of the traditional identification of empiricism with inductivism;
    9) Rejection of the imagist tradition that treats cognitive states or contents as little pictures before consciousness;
    10) Rejection of "the myth of the given," by Sellars and others, the idea that subjective experience provides a special, direct, infallible, nonnatural connection of knowing mind to known world;
    11) the failure of phenomenalism and sense datum theories of perception; and, more generally,
    12) rejection of the whole Cartesian-Lockean conception of cognition and language;
    13) The failure of attempts to define knowledge precisely as justified true belief; which inspired
    14) externalism versus internalism in epistemology;
    15) Recognition of the importance of tacit versus explicit knowledge (knowledge-how vs. knowledge-that) and of embodied knowledge, for example, skilled practices that we cannot fully articulate;
    16) The feminist introduction of gender variables into epistemology;
    17) Competing attempts to naturalize and socialize epistemology;
    18) The postmodern critique of empiricism. Postmodernists, including Richard Rorty and radical feminists and sociologists, regard empiricism, epistemology in general, and, indeed, the entire Enlightenment project to replace a tradition-bound life.

52. Among the significant theorists associated with historicism are Leopold von Ranke, Wilhelm Dilthey and G. W. H. Hegel. Historicism has influenced hermeneutical and phenomenological thinkers such as Martin Heidegger, Edmund Husserl and Hans-Georg Gadamer. It has also influenced pragmaticism as well as Marxist and critical positions.
53. The historicist approach regarding the object of classification is today the dominant principle in biological classification. It was clearly expressed by Charles Darwin (1859, 420): "all true classification is genealogical."
54. Pragmatism is a philosophical tradition founded by three American philosophers: Charles Sanders Peirce (1839-1914), William James (1842–1910) and John Dewey (1859–1952). The American social psychologist George Herbert Mead (1863-1931) and the American philosopher Clarence Irving Lewis (1883-1964) are also regarded as "classical" pragmatists. All three of the founding pragmatists combined a naturalistic, Darwinian view of human beings with a distrust of the problems that philosophy had inherited from Descartes, Hume and Kant. They hoped to save philosophy from metaphysical idealism. Their naturalism has been combined with an anti-foundationalist, holist account of meaning by Willard van Orman Quine, Hil-

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ary Putnam and Donald Davidson—philosophers of language who are often seen as belonging to the pragmatist tradition. That tradition also has affinities with the work of Thomas Kuhn and the later works of Ludwig Wittgenstein. One of the neo-pragmatic philosophers, Hilary Putnam, enumerates the most important pragmatic theses (Putnam, 1994, 152):

> What I find attractive in pragmatism is not a systematic theory in the usual sense at all. It is rather a certain group of theses… Cursorily summarized, those theses are
>
> 1) antiscepticism: pragmatists hold that doubt requires justification just as much as belief…
> 2) fallibilism: pragmatists hold that there is never a metaphysical guarantee to be had that such-and-such a belief will never need revision (that one can be both fallibilistic and antisceptical is perhaps the unique insight of American pragmatism);
> 3) the thesis that there is no fundamental dichotomy between "facts" and "values"; and
> 4) the thesis that, in a certain sense, practice is primary in philosophy.

55. Rothman (1974, 292) wrote: "The indexer has a choice of two basic approaches to the text to be indexed. He can use as index terms the vocabulary of the original document, or he can read the original document for content, assigning to the concepts discussed in it terms that seems most appropriate to him, whether or not they coincide with those used by the original author. (Word indexing is therefore often called derivative indexing; concept indexing is often called assignment indexing.)"
56. Consider, however, that Bernier (1980) is expressing a document-oriented view and not a request-oriented view. According to the request-oriented view, the subject of a document is what informs the users (see further in Hjørland 2017b (http://www.isko.org/cyclo/subject). Still, however, Bernier's distinction between concept indexing and subject indexing is extremely important.
57. See Gnoli (2018) for a comprehensive coverage of notational systems.
58. Swift, Winn and Bramer (1979, 218) wrote: "For instance, the same document might be viewed by one person as being about structural determinants of achievement in a capitalist society; by another as being about progressive versus traditional teaching styles and achievement in the primary school; and by a third as being about teaching working class children in inner urban areas. Each of these interpretations would be likely to find widespread support in social science."
59. Weinberg (2017, 1987) wrote: "Indexing may be slanted to the purpose of an organization, in which case it is called mission-oriented indexing." Remark, however, that from the pragmatic view, all indexing should be understood as slanted. The opposite, which has been termed "the view from nowhere" (Nagel 1986), is from the perspective of pragmatic and critical theory, an illusion.
60. Swift, Winn and Bramer (1977 and 1979) generalizes these perspectives for the theory of indexing and design of information systems, while Swift, Winn and Bramer (1978) contains a criticism of the concept aboutness; their papers seem extremely important. Unfortunately, Lancaster (2003) only discusses their 1978 paper on aboutness and thereby seems to have missed a very important perspective on indexing theory.
61. MEDLINE Online Indexing Training Module saved in WebCite 2018-03-03: http://www.webcitation.org/6xeIKGYRx
62. About perspectivism in knowledge organization see Mazzocchi (2018, 70-2, Section 5.3) also in IEKO: http://www.isko.org/cyclo/kos#5.3
63. Another way to express the same point is, with the words of Hjørland (1992, 1997), to index its informative potentials, i.e., its potential of informing users and advance the development of knowledge.
64. "Language" is here used in a broader meaning, including nonverbal languages (in the sense in which, for example, body language is understood as a language).
65. Weinberg (2017, 1980) found that "metadata" belongs to the "vogue words:" "The term metadata, once reserved for data about websites, has been applied retroactively to all cataloging and indexing data." However, such a generalized concept may be important from a theoretical perspective and ISKO Encyclopedia of Knowledge Organization is preparing a special article about this term.
66. The understanding of classification as indexing is in opposition to some ways of thinking. Weinberg (2017, 1978), for example, wrote: "Although Anderson [1989] believes that indexing and classification are the same thing, this entry treats index headings that are arranged alphabetically, while hierarchical display of content indicators is discussed in other articles."
67. Lancaster (2003, 20-21):

> Indexing as Classification.
> In the literature of library and information science, a distinction is sometimes made among

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

> the three terms *subject indexing*, *subject cataloging* and *classification*. … This distinction between the terms *subject cataloging* and *subject indexing* one referring to complete bibliographic items and the other to parts of items, is artificial, misleading, and inconsistent.
>
> …
>
> The situation is even more confusing when the term *classification* is considered. Librarians tend to use the word to the assignment of class numbers [drawn from some classification scheme—e.g., the Dewey Decimal (DDC), Universal Decimal (UDC), Library of Congress (LC)] to bibliographic items, especially for the purpose of arranging these items on the shelves of libraries, in filing cabinets, and so on. But the subject catalog of a library can be either alphabetically based (an *alphabetical subject catalog* or a *dictionary catalog*) or arranged according to the sequence of some classification (a *classified catalog*). Suppose that a librarian picks up a book and decides that it is about "birds." He or she might assign the subject heading *birds* to this item. Alternatively, the class number 598 may be assigned to it. Many people would refer to the first operation as *subject cataloging* and to the second as *classification*, a completely nonsensical distinction.
>
> …
>
> These terminological distinctions are quite meaningless and only serve to cause confusion … The fact is that *classification* in the broadest sense, permeates all of the activities associated with information storage and retrieval. Part of the terminological confusion is caused by the failure to distinguish between the *conceptual analysis* and the *translation* stages in indexing (italics in original).

68. Browne and Jermey (2007, 73) wrote: "Studies in the 1960s and 1970s suggested that free-text searching could provide results that were equal to or better than searches using human indexing based on a controlled vocabulary. Later studies using larger databases with realistic search queries have challenged these findings. Free-text and controlled vocabulary searches usually each provide unique his (that is, each search type finds some relevant items that the other does not find) and are therefore complementary. They also each work better for different kinds of searches."

    Gross, Taylor and Joudrey (2015) suggests that roughly 30% of relevant search results are lost by relying solely on "keyword" (free-text) searches. Their study is, however, restricted to library databases, not full-text databases. They wrote: "Research that looks at the effect of controlled subject vocabulary in discovery layers and Web-scale discovery tools has begun to appear, and in the near term, these rapidly changing environments are the domain in which the impact of subject headings needs to be investigated most urgently. In the long term, the ultimate test of the importance of controlled vocabulary will be its effect in full text environments. While most studies that have looked at the role of subject metadata in full text searching indicate that controlled vocabulary is needed in full text environments, research in this area needs to continue and expand as the extent and accessibility of full text resources increases." The authors have also reservations that their study did not consider the relevance of the found and missing documents for users.

69. In 2016, "metadata" was used in 300 titles indexed in Web of Science (on 2018-04-16), compared to one in 1982 (and zero before 1982); "indexing language*" or "index language*," on the other hand, occurred in one title in 2016 compared to five in 1982 (and the first record is from 1963: Cleverdon and Mills 1963).

70. Van Rijsbergen (1979, 13-4; italics in original) wrote about two of these concepts: "Traditionally the two most important factors governing the effectiveness of an index language have been thought to be the exhaustivity of indexing and the specificity of the index language. There has been much debate about the exact meaning of these two terms. Not wishing to enter into this controversy I shall follow Keen and Digger [1972] in giving a working definition of each.

    For any document, indexing exhaustivity is defined as the number of different topics indexed, and the index language specificity is the ability of the index language to describe topics precisely. Keen and Digger further define indexing specificity as the level of precision with which a document is actually indexed. It is very difficult to quantify these factors. Human indexers are able to rank their indexing approximately in order of increasing exhaustivity or specificity. However, the same is not easily done for automatic indexing.

    It is of some importance to be able to quantify the notions of indexing exhaustivity and specificity because of the predictable effect they have on retrieval effectiveness. It has been recognized (Lancaster [1968]) that a high level of exhaustivity of indexing leads to high recall* and low precision*. Conversely, a low level of exhaustivity leads to low recall and high precision. The converse is true for levels of indexing specificity, high specificity leads to high precision and low recall, etc. It would seem, therefore, that there is an optimum level of indexing exhaustivity and specificity for a given user population.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Quite a few people (Spärck Jones [1972, 1973], Salton and Yang [1973]), have attempted to relate these two factors to document collection statistics. For example, exhaustivity can be assumed to be related to the number of index terms assigned to a given document, and specificity related to the number of documents to which a given term is assigned in a given collection. The importance of this rather vague relationship is that the two factors are related to the distribution of index terms in the collection. The relationships postulated are consistent with the observed trade-off between precision and recall just mentioned. Changes in the number of index terms per document lead to corresponding changes in the number of documents per term and vice versa."

71. Anderson and Pérez-Carballo (2005, Chapter 9, 177-183) is about exhaustivity, recall and precision.
72. Anderson and Pérez-Carballo (2005, Chapter 10, 185-196) is about specificity.
73. Granularity as defined by Browne and Jermey (2007, 29) differ from the way the term is normally understood. Vickery (1997, 278), for example, defined: "the 'granularity' or 'grain size' of an ontology [means]—to what degree of specifity should the concept hierarchy be continued." Henttonen (2015) uses the term in the same meaning: how detailed an indexing language is.
74. In the context of the humanities, a document may itself be a meta-description of other documents (e.g., a history of American literature is a meta-description of the literature it refers to). Therefore, representations in LIS of, for example, histories of literature, becomes second-order meta-representations.
75. However, as stated by Chatterjee (2016, 137): "Whereas a natural language can function as its own meta language (i.e., a language used to talk about another language), an indexing language cannot."
76. Epistemology has generally been neglected in LIS. An important exception was Don R. Swanson, who, based on Popper's philosophy, concluded (1986, 114): "Any search function is necessarily no more than a conjecture and must remain so forever." The same is true also for the knowledge representation and indexing underlying a given search function.

References

American Society for Indexing. 2017. "Software." https://web.archive.org/web/20170902091621/https://www.asindexing.org/reference-shelf/software/

Andersen, Jack. 2004. "Analyzing the Role of Knowledge Organization in Scholarly Communication: An Inquiry into the Intellectual Foundation of Knowledge Organization." Ph.D. diss., Royal School of Library and Information Science. http://static-curis.ku.dk/portal/files/47069480/jack_andersen_phd.pdf

Anderson, James D. 1989. "Indexing and classification: File organization and display for information retrieval." In *Indexing: The State of Our Knowledge and the State of Our Ignorance. Proceedings of the 20th Annual Meeting of the American Society of Indexers, New York, May 13, 1998*, ed. Bella H. Weinberg. Medford, NJ.: Learned Information, 71-82.

Anderson, James D. 1994. Review of *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice,* by Robert Fugmann. *Library Quarterly* 64, no. 4: 475-7.

Anderson, James D. 1997. *Guidelines for Indexes and Related Information Retrieval Devices*. NISO Technical Report 2. Bethesda, MD: NISO Press. http://niso.kavi.com/publications/tr/tr02.pdf

Anderson, James D. and José Pérez-Carballo. 2001a. "The Nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part I: Research, and the Nature of Human Indexing." *Information Processing & Management* 37: 231-54.

Anderson, James D. and José Pérez-Carballo. 2001b. "The Nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part II: Machine Indexing, and the Allocation of Human Versus Machine Effort." *Information Processing & Management 37*: 255-77.

Anderson, James D. and José Pérez-Carballo. 2005. *Information Retrieval Design. Principles and Options for Information Description, Organization, Display and Access in Information Retrieval Databases, Digital Libraries, Catalogs, and Indexes*. St. Petersburg, Fl: Ometeca Institute.

Austin, Derek. 1974. *PRECIS: A Manual of Concept Analysis and Subject Indexing*. London: The Council of the British National Bibliography.

Barité, Mario. 2018. "Literary Warrant." *Knowledge Organization* 45: 517-36.

Bates, Marcia J. 1988. "How to Use Controlled Vocabularies More Effectively in Online Searching." *Online* 12, no. 6: 45-56.

Beghtol, Clare. 1986. "Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents." *Journal of Documentation* 42, no. 2: 84-113.

Bell, Hazel K. 1991. "Bias in Indexing and Loaded Language." *The Indexer* 17, no. 3: 173-7.

Bernier, Charles L. 1980. "Subject indexes." In *Encyclopedia of Library and Information Science*, ed. Allen Kent, Harold Lancour, and Jay E. Daily. New York, NY: Marcel Dekker, 29: 191-205.

Bourcier, Paul, Heather Dunn and the Nomenclature Task Force, eds. 2015. *Nomenclature 4.0 for Museum Cataloging*. 4th ed. Lanham, MD: Rowman & Littlefield. Fourth

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

edition of Robert G. Chenhall's system for classifying cultural objects

Borko, Harold. 1977. "Towards a Theory of Indexing." *Information Processing & Management* 13, no. 6: 355-65.

Borko, Harold and Charles L. Bernier. 1978. *Indexing Concepts and Methods*. Library and Information Science. New York: Academic Press.

Brenner, Diane and Marilyn Rowland, eds. 2000. *Beyond Book Indexing: How to Get Started in Web Indexing, and Other Computer-Based Media*. Phoenix, AZ: American Society of Indexers.

Browne, Glenda and Jon Jermey. 2007. *The Indexing Companion*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139168595

BSI (British Standards Institution). 1988. *British Standard Recommendations for Preparing Indexes to Books, Periodicals, and Other Documents*. 2nd revision. London: British Standards Institute.

Buckland, Michael. 2018. "Document Theory." *Knowledge Organization* 45: 425-36.

Cavnar, William B. and John M. Trenkle. 1994. "N-Gram-Based Text Categorization." In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. Las Vegas, NV: Information Science Research Institute, University of Nevada, 161-75.

Chan, Lois Mai. 1994. *Cataloging and Classification: An Introduction*. 2nd ed. Library Science Series. New York: McGraw-Hill.

Chatterjee, Amitabha. 2016. *Elements of Information Organization and Dissemination*. Cambridge, MA: Chandos.

Chu, Heting. 2001. "Research in Image Indexing and Retrieval as Reflected in the Literature." *Journal of the American Society for Information Science and Technology* 52, no. 12: 1011-18.

Cleverdon, Cyril W. and Jack Mills. 1963. "The Testing of Index Language Devices." *Aslib Proceedings* 15, no. 4: 106-30.

Cooper, William S. 1969. "Is Interindexer Consistency a Hobgoblin?" *American Documentation* 20, no. 3: 268-78.

Cooper, William S. 1978. "Indexing Documents by Gedanken Experimentation." *Journal of the American Society for Information Science* 29, no. 3: 107-19.

Craven, Timothy C. 1986. *String Indexing*. Library and Information Science. Orlando, FL: Academic Press.

Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection; or, the Preservation of Favoured Races in the Struggle for Life*. London: J. Murray.

David, Claire, L. Giroux, Suzanne Bertrand-Gastaldy, and D. Lanteigne 1995. "Indexing as Problem Solving: A Cognitive Approach to Consistency." In *ASIS '95: Proceedings of the 58th ASIS Annual Meeting, Chicago, Illinois, October 9-12, 1995; Converging Technologies; Forging New Partnerships in Information,* 32, ed. T. Kinney. Medford, NJ: Information Today for the American Society for Information Science, 49-55.

Diaz, Alejandro. 2008. "Through the Google Googles: Sociopolitical Bias in Search Engine Design." In *Web Search: Multidisciplinary Perspectives*, ed. Amanda Spink and Michael Zimmer. Information Science and Knowledge Management 14. Berlin: Springer, 11-34.

Diodato, Virgil and Gretchen Gandt. 1991. "Back of the Book Indexes and the Characteristics of Author and Nonauthor Indexing: Report of an Exploratory Study." *Journal of the American Society for Information Science* 42: 341-50.

Dousa, Thomas M. and Fidelia Ibekwe-Sanjuan. 2014. "Epistemological and Methodological Eclecticism in the Construction of Knowledge Organization Systems (KOSs): The Case of Analytico-synthetic KOSs." In *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects; Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014, Kraków, Poland*, ed. Wiesław Babik. Advances in Knowledge Organization 14. Würzburg: Ergon, 152-9. doi:10.1016/j.ijmedinf.2013.01.005

Dragusin, Radu, Paula Petcu, Christina Lioma, Birger Larsen, Henrik L. Jørgensen, Ingemar J. Cox, Lars Kai Hansen, Peter Ingwersen and Ole Winther. 2013a. "FindZebra: A Search Engine for Rare Diseases." *International Journal of Medical Informatics* 82, no. 6: 528-38.

Dragusin, Radu, Paula Petcu, Christina Lioma, Birger Larsen, Henrik L. Jørgensen, Ingemar J. Cox, Lars Kai Hansen, Peter Ingwersen and Ole Winther. 2013b. "Specialised Tools are Needed when Searching the Web for Rare Disease Diagnoses." *Rare Diseases* 1, no. 1. doi:10.4161/rdis.25001

Farradane, Jason. 1977. *Relational Indexing: Introduction and Indexing*. String Indexing. London, Ontario: University of Western Ontario, School of Library and Information Science.

Farradane, Jason. 1980. "Knowledge, Information, and Information Science." *Journal of Information Science* 2, no. 2: 75-80. doi:10.1177/016555158000200203

Farrow, John D. 1991. "A Cognitive Process Model of Document Indexing." *Journal of Documentation* 47: 149-66.

Farrow, John D. 1994. "Indexing as a Cognitive Process." In *Encyclopedia of Library and Information Science*, ed. Allen Kent, Harold Lancour, and Jay E. Daily. New York, NY: Marcel Dekker, 53, supp. 16: 155-71.

Farrow, John D. 1995. "All in the Mind: Concept Analysis in Indexing." *The Indexer* 19, no. 4: 243-47.

Finnemann, Niels Ole. 2018. "E-text." In *The Oxford Research Encyclopedia of Literature*, ed. Paula Rabinowitz. New York: Oxford University Press, 1-47. doi:10.1093/acrefore/9780190201098.013.272

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Foskett, D.J. 1970. *Classification for a General Indexing Language: A Review of Recent Research by the Classification Research Group*. London: The Chartered Institute of Library and Information Professionals.

Fraenkel, Carlos, Dario Perinetti, and Justin E.H. Smith. 2011. "Introduction." In *The Rationalists: Between Tradition and Innovation*, ed. Carlos Fraenkel, Dario Perinetti and Justin E.H. Smith. The New Synthese Historical Library 65. Dordrecht: Springer.

Frohmann, Bernd. 1990. "Rules of Indexing: A Critique of Mentalism in Information Retrieval Theory." *Journal of Documentation* 46, no. 2: 81-101.

Fugmann, Robert. 1979. "Toward a Theory of Information Supply and Indexing." *International Classification* 6, no.1: 3-15.

Fugmann, Robert. 1985. "The Five-Axiom Theory of Indexing and Information Supply." *Journal of the American Society for Information Science* 36, no. 2: 116-129.

Fugmann, Robert. 1992. "Indexing Quality - Predictability versus Consistency." *International Classification* 19, no. 1: 20-21.

Fugmann, Robert. 1993. *Subject Analysis and Indexing: Theoretical Foundations and Practical Advice*. Textbooks for Knowledge Organization 1, Frankfurt am Main: Indeks.

Fugmann, Robert. 1997. "Bridging the Gap between Database Indexing and Book Indexing." *Knowledge Organization* 24: 205 ·12. Corrections in *Knowledge Organization* 25: 35.

Furner, Jonathan. 2012. "FRSAD and the Ontology of Subjects of Works." *Cataloging & Classification Quarterly* 50, nos. 5-7: 494-516.

Gartner, Richard. 2016. *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*. Cham: Springer.

Gnoli, Claudio. 2-18. "Notation." *ISKO Encyclopedia of Knowledge Organization*. http://www.isko.org/cyclo/notation

Granka, Laura A. 2010. "The Politics of Search: A Decade Retrospective." *The Information Society* 26, no. 5: 364-74.

Gross, Tina, Arlene G. Taylor and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53, no. 1: 1-39.

Guimaräes, José Augusto Chaves. 2017. "Slanted Knowledge Organization as a New Ethical Perspective." In *The Organization of Knowledge: Caught Between Global Structures and Local Meaning*, ed. Jack Andersen and Laura Skouvig. Studies in Information 12. Bingley, UK: Emerald, 87-102.

Haider, Salman. 2018 "Chain Indexing." *Librarianship Studies & Information Technology* (blog). https://librarianshipstudies.blogspot.dk/2017/04/chain-indexing.html

Hanna, Robert. 1998. "Conceptual Analysis." *Routledge Encyclopedia of Philosophy*, ed. Edward Craig. Version 1.0. London: Routledge, 2: 518-22. doi:10.4324/9780415249126-U033-1

Hanna, Robert. 2007. "Kant, Wittgenstein and the Fate of Analysis." In *The Analytic Turn: Analysis in Early Analytic Philosophy and Phenomenology,* ed. Michael Beaney. Routledge Studies in Twentieth-Century Philosophy. London: Routledge, 142-63.

Hargittai, Eszter. 2007. "The Social, Political, Economic, and Cultural Dimensions of Search Engines: An Introduction." *Journal of Computer-Mediated Communication* 12, no. 3: 769-77.

Harper, Douglas. 2017. "Index." In *Online Etymological Dictionary*. https://www.etymonline.com/word/index

Hedden, Heather. 2007. *Indexing Specialties: Web Sites*. Medford, NJ: Information Today.

Henttonen, Pekka. 2015. "Dimensions of Contextual Records Management Classifications." *Knowledge Organization* 42: 477-85.

Hjørland, Birger. 1992. "The Concept of 'Subject' in Information Science." *Journal of Documentation* 48, no. 2: 172-200.

Hjørland, Birger. 1997. *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science*. New Directions in Information Management 34. Westport, CT: Greenwood Press.

Hjørland, Birger. 2011. "The Importance of Theories of Knowledge: Indexing and Information Retrieval as an Example." *Journal of the American Society for Information Science and Technology* 62, no. 1:72-7.

Hjørland, Birger. 2012. "Is Classification Necessary after Google?" *Journal of Documentation* 68, 3: 299-317.

Hjørland, Birger. 2013a. "Theories of Knowledge Organization: Theories of Knowledge." *Knowledge Organization* 40: 169−81.

Hjørland, Birger. 2013b. "User-based and Cognitive Approaches to Knowledge Organization: A Theoretical Analysis of the Research Literature." *Knowledge Organization* 40, no. 1: 11-27.

Hjørland, Birger. 2015a. "Are Relations in Thesauri 'Context Free, Definitional, and True in all Possible Worlds'?" *Journal of the Association for Information Science and Technology 66, no. 7: 1367-73.*

Hjørland, Birger. 2015b. "Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-Making During Searches." *Journal of the Association for Information Science and Technology* 66, no. 8: 1559-1575.

Hjørland, Birger 2017a. "Classification." *Knowledge Organization* 44: 97-128.

Hjørland, Birger. 2017b. "Subject (of Documents)." *Knowledge Organization* 44: 55-64.

Hjørland, Birger. 2018. "Library and Information Science (LIS)." *Knowledge Organization* 45: 232-54 and 319-38.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Hjørland, Birger and Lykke Kyllesbech Nielsen. 2001. "Subject Access Points in Electronic Retrieval." *Annual Review of Informaion Science and Technology* 35, 249-98.

Intner, Sheila S. 1984. "Censorship in Indexing." *The Indexer* 14, no. 2: 105-8.

Introna, Lucas D. and Helen Nissenbaum. 2000. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16, no. 3: 169-85.

ISO (International Organization for Standardization). 1996. *Information and Documentation: Guidelines for the Content, Organization and Presentation of Indexes*. 2nd ed. Ref no. ISO 999-1996. Geneva, Switzerland: International Organization for Standardization.

ISO (International Organization for Standardization).1985. *Documentation: Methods for Examining Documents, Determining Their Subjects, and Selecting Indexing Terms*. Ref. no. ISO 5963-1985. [Geneva]: International Organization for Standardization.

Jonker, Frederick. 1964. *Indexing Theory, Indexing Methods and Search Devices*. New York: Scarecrow Press.

Jörgensen, Corinne. 2017. "Still Image Indexing." In *Encyclopedia of Library and Information Sciences*, ed. John D. McDonald and Michael Levine-Clark. 4th ed. Boca Raton, FL: CRC Press, 7: 4407-16.

Kaszkiel, Marcin. Justin Zobel and Ron Sacks-Davis. 1999. "Efficient Passage Ranking for Document Databases." *ACM Transactions on Information Systems* 17, no. 4: 406-439.

Keen, E. Michael and Jeremy A. Digger. 1972. *Report of an Information Science Index Languages Test*. 2 vols. Aberystwyth: College of Librarianship, Department of Information Retrieval Studies.

Kendrick, Peter A. and Enid L. Zafran, eds. 2001. *Indexing Specialties: Law*. Medford, NJ: Information Today.

Keyser, Pierre de. 2012. *Indexing: From Thesauri to the Semantic Web*. Chandos Information Professional Series. Oxford: Chandos.

Klement, Susan. 2002. "Open-System versus Closed-System Indexing: A Vital Distinction." *The Indexer* 23, no.1: 23-24.

Knight, G. Norman. 1979. *Indexing, the Art of: A Guide to the Indexing of Books and Periodicals*. London: Allen & Unwin.

Krarup, Karl and Ivan Boserup. 1982. *Reader-Oriented Indexing: An Investigation into the Extent to which Subject Specialists should be used for the Indexing of Documents by and for Professional Readers, based on a Sample of Sociological Documents Indexed with the Help of the PRECIS Indexing System*. Automation and Documentation Pamphlets 2. Copenhagen: Royal Library.

Lancaster, F. Wilfrid. 1968. *Information Retrieval Systems: Characteristics, Testing and Evaluation. Information Sciences Series*. New York: Wiley.

Lancaster, F. Wilfrid. 1991. *Indexing and Abstracting in Theory and Practice*. Champaign, IL: University of Illinois, Graduate School of Library and Information Science.

Lancaster, F. Wilfrid. 1994. Review of *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice,* by Robert Fugmann. *Journal of Documentation* 50: 149-52.

Lancaster, F. Wilfrid. 2003. *Indexing and Abstracting in Theory and Practice*. 3rd ed. London: Facet Publishing.

Landry, Bertrand Clovis and James E. Rush. 1975. "Automatic Indexing: Progress and Prospects." In *Encyclopedia of Computer Science and Technology*, ed, Jack Baizer, Albert Holzman and Allen Kent. New York: Marcel Dekker, 2: 403-48.

Laporte, Steven. 2018. "Ideal Language." *Knowledge Organization* 45: 586-608.

Leinnger, Kurt. 2000. "Interindexer Consistency in PsycINFO." *Journal of Librarianship and Information Science* 32, no.1: 4-8.

Leonard, Lawrence E. 1977. *Inter-Indexer Consistency Studies, 1954-1975: A Review of the Literature and Summary of Study Results*. University of Illinois, Graduate School of Library and Information Science Occasional Papers 131. Champaign, IL: University of Illinois, Graduate School of Library Science.

Lewandowski, Dirk. 2014. "Why We Need an Independent Index of the Web." In *Society of the Query Reader: Reflections on Web Search*, ed. René Kõnig and Miriam Rasch. Amsterdam: Institute of Network Cultures, 49-58.

Lewandowski, Dirk. 2015. "Living in a World of Biased Search Engines." *Online Information Review* 39, no. 3: 278-80.

Library of Congress. 2008. *The Subject Headings Manual*, ed. Paul Weiss. 4 vols. Washington, DC: Library of Congress, Policy and Standards Division.

Loukopoulos, Loukas. 1966. "Indexing Problems and Some of Their Solutions." *American Documentation* 17, no. 1: 17-25.

Mager, Astrid. 2012. "Algorithmic Ideology. How Capitalist Society Shapes the Search Engines." *Information, Communication & Society* 15, no. 5: 769-87.

Markey, Karen. 1984. "Interindexer Consistency Tests: A Literature Review and a Report of Test of Consistency in Indexing Visual Materials." *Library and Information Science Research* 6, no. 2: 155-177.

Mai Chan, Lois. 2005. *Library of Congress Subject Headings: Principles and Application*. 4th ed. Library and Information Science Text Series. Westport, CT: Libraries Unlimited.

Mai, Jens-Erik. 2000. "The Subject Indexing Process: An Investigation of Problems in Knowledge Representation." PhD diss., University of Texas at Austin. http://jenserikmai.info/Papers/2000_PhDdiss.pdf

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Maron, Melvin Earl "Bill." 2008. "An Historical Note on the Origins of Probabilistic Indexing." *Information Processing & Management* 44, no. 2: 971-972.

Mazzocchi, Fulvio. 2018. "Knowledge Organization System (KOS): An Introductory Critical Account." *Knowledge Organization* 45: 54-78.

Miksa, Francis L. 2006. "The DDC Relative Index." In *Moving Beyond the Presentation Layer: Context and Content in the Dewey Decimal Classification (DDC) System*, ed. Joan S. Mitchell and Diane Vizine-Goetz. Binghamton, NY: Haworth Press, 65-95.

Milstead, Jessica L. 1984. *Subject Access Systems: Alternatives in Design*. Library and Information Science. Orlando, FL: Academic Press.

Morville, Peter and Louis Rosenfeld. 2007. *Information Architecture for the World Wide Web*. 3. edition. Sebastopol, CA: O´Reilly Media.

Mulvany, Nancy C. 1994. "The Author and the Index." *The Indexer* 19, no.1: 28-30.

Mulvany, Nancy C. 2005. *Indexing Books*. 2nd ed. Chicago, IL: University of Chicago Press.

Mulvany, Nancy C. 2010. "Back-of-the-Book Indexing." In *Encyclopedia of Library and Information Sciences*, ed. Marcia J. Bates and Mary Niles Maack. 3rd ed. Boca Raton, FL: CRC Press, 1: 485-491.

Mulvany, Nancy C. 2017. "Back-of-the-Book Indexing", In *Encyclopedia of Library and Information Sciences*, ed. John D. McDonald and Michael Levine-Clark. 4th ed. Boca Raton, FL: CRC Press, 1: 440-6.

National Library of Medicine. 2018. "Frequently Asked Questions About Indexing for MEDLINE: Who are the Indexers, and What are Their Qualifications?" https://www.nlm.nih.gov/bsd/indexfaq.html#qualifications

Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.

Nickles, Thomas. 2005. "Empiricism." In *New Dictionary of the History of Ideas*, ed. Maryanne Cline Horowitz. New York: Charles Scribners & Sons. http://www.encyclopedia.com/topic/empiricism.aspx

Olson, Hope A. 2002. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Pariser, Eli. 2011. *Filter Bubble: What the Internet is Hiding from You*. London: Viking.

Peirce, Charles Sanders.1931-58. *Collected Papers of Charles Sanders Peirce*, ed. Charles Hartshorne, Paul Weiss and Arthur W Burks. 8 vols. Cambridge, MA: Belknap Press.

Rafferty, Pauline. 2018. "Tagging." *Knowledge Organization* 45: 500-16.

Rafferty, Pauline and Rob Hidderley. 2005. *Indexing Multimedia and Creative Works: The Problems of Meaning and Interpretation*. Aldershot: Ashgate.

Ranganathan, S.R. 1938. *Theory of Library Catalogue*. Madras Library Association. Publication Series 7. Madras: The Madras Library Association.

Ranganathan, S.R., ed. 1963. *Documentation and Its Facets: Being A Symposium Of Seventy Papers By Thirty-Two Authors*. New York: Asia Publishing House.

Ranganathan, S.R. 1967. *Prolegomena to Library Classification*. Assisted by M.A. Gopinath. 3rd ed., Ranganathan Series in Library Science 20. London: Asia Publishing House.

Rasmussen Neal, Diane, ed. 2012. *Indexing and Retrieval of Non-Text Information*. Knowledge and Information. Berlin: De Gruyter Saur.

Rothman, John. 1974. "Index, Indexer, Indexing." In *Encyclopedia of Library and Information Science*, ed. Allen Kent, Harold Lancour, and Jay E. Daily. New York, NY: Marcel Dekker, 11: 286-99.

Salton, Gerard and C. S. Yang. 1973. "On the Specification of Term Values in Automatic Indexing." *Journal of Documentation* 29: 351-72.

Schroeder, Sandi, ed. 2003. *Software for Indexing*. Medford, NJ: Information Today.

Shepherd, M. A. 1981. "Text Passage Retrieval Based on Colon Classification: Retrieval Performance." *Journal of Documentation* 37: 25-35.

Shera, Jesse. H. 1951. Classification as the basis of bibliographic organization. In *Bibliographic Organization: Papers presented before the Fifteenth Annual Conference of the Graduate Library School July 24-29, 1950,* ed. Jesse H. Shera and Margaret E. Egan. Chicago: University of Chicago Press, 72-93.

Smiraglia, Richard P. and Xin Cai. 2017. "Tracking the Evolution of Clustering, Machine Learning, Automatic Indexing and Automatic Classification in Knowledge Organization." Knowledge Organization 44: 215-33.

Smith, Laurence D. 1986. *Behaviorism and Logical Positivism: A Reassessment of the Alliance*. Stanford, CA: Stanford University Press.

Sober, Elliott. 1988. *Reconstructing the Past: Parsimony, Evolution and Inference*. Cambridge, MA: MIT Press.

Soergel, Dagobert. 1974. *Indexing Languages and Thesauri: Construction and Maintenance*. Information Sciences Series. Los Angeles, CA: Melville.

Soergel, Dagobert. 1985. *Organizing Information: Principles of Data Base and Retrieval Systems*. Library and Information Science. Orlando, FL: Academic Press.

Soler Monreal, M. Concha and Isidoro Gil-Leiva. 2011. "Evaluation of Controlled Vocabularies by Inter-Indexer Consistency." *Information Research*, 16, no. 4, paper 502.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

*Spang-Hanssen, Henning. 1974. "Kunnskapsorganisasjon, informasjonsgjenfinning, automatisering og språk." In *Kunnskapsorganisasjon og informasjonsgjenfinning: Seminar arrangert 3.-7. desember 1973*. Skrifter fra Riksbibliotektjenesten 2. Oslo: Riksbibliotektjenesten, 11-61.

Spang-Hanssen, Henning. 1976. *Roles and Links Compared with Grammatical Relations in Natural Languages*. Dansk teknisk litteraturselskab 40. Lyngby, Denmark: Dansk Teknisk Litteraturselskab.

Spärck Jones, Karen. 1972. "A Statistical Interpretation of Term Specificity and Its Application in Retrieval." *Journal of Documentation* 28, 111-21.

Spärck Jones, Karen. 1973 "Does Indexing Exhaustivity Matter?" *Journal of the American Society for Information Science* 24: 313-6.

Stock, Wolfgang G. and Mechtild Stock. 2013. *Handbook of Information Science*, trans. Paul Becker. Berlin: De Gruyter Saur.

Suchting, Wal. 2012. "Empiricism." Translated by Peter Thomas. *Historical Materialism* 20, no. 3: 213-8.

Svenonius, Elaine. 2003. "Design of Controlled Vocabularies." In *Encyclopedia of Library and Information Science*. ed. Miriam A. Drake. 2nd ed. New York: Marcel Dekker, 2: 822-38.

Swanson, Don R. 1986. "Undiscovered Public Knowledge." *The Library Quarterly* 56, no. 2: 103-18.

Swift, Donald F., Viola A. Winn and Dawn A. Bramer. 1973. *A Case Study in Indexing and Classification in the Sociology of Education: Development of Ideas Concerning the Organisation Material for Literature Searching*, vol 1. OSTI Report 5171. Milton Keynes, Open University. https://files.eric.ed.gov/fulltext/ED086258.pdf

Swift, Donald F., Viola A. Winn and Dawn A. Bramer. 1977. "A Multi-Modal Approach to Indexing and Classification." *International Classification* 4, no. 2: 90-4.

Swift, Donald F., Viola A. Winn, and Dawn A. Bramer. 1978. "'Aboutness' as a Strategy for Retrieval in the Social Sciences." *Aslib Proceedings* 30, no. 5: 182-7.

Swift, Donald F., Viola A. Winn and Dawn A. Bramer. 1979. "A Sociological Approach to the Design of Information Systems." *Journal of the American Society for Information Science* 30, no. 4: 215-223.

Taube, Mortimer. 1953. *Studies in Coordinate Indexing*, vol. I. Washington, D.C.: Documentation. https://babel.hathitrust.org/cgi/pt?id=mdp.39015082966196;view=1up;seq=5

Towery, Margie, ed. 1998. *Indexing Specialties: History*. Medford, NJ: Information Today.

Turner, James M. 2017. "Moving Image Indexing." In *Encyclopedia of Library and Information Sciences*, ed. John D. McDonald and Michael Levine-Clark. 4th ed. Boca Raton, FL: CRC Press, 5: 3129-39.

University of Chicago Press. 2017. *Chicago Manual of Style*. 17th ed. Chicago: University of Chicago Press.

Van Rijsbergen, C.J. 1979. *Information Retrieval*. 2nd ed. London: Butterworths.

Vickery, Brian C. 1997. "Ontologies." *Journal of Information Science* 23, no. 4: 277-86.

Warner, Julian. 2004. "Meta- and Object-Language for Information Retrieval Research: Proposal for a Distinction." *Aslib Proceedings* 56, no. 2: 112-117.

Warner, Julian. 2010. *Human Information Retrieval*. History and Foundations of Information Science Series. Cambridge, MA: MIT Press.

Watson, L. E., P. Gammage, M. C. Grayshon, S. Hockey, R. K. Jones and D. Oldman. 1973. "Sociology and Information Science." *Journal of Librarianship* 5, no. 4: 270-83.

Webster, Noah. 1966. *Webster's New World Dictionary of the American Language*. College ed. Cleveland OH: World.

Weinberg, Bella Hass. 2007. "Known Orders: Unusual Locators in Indexes." *Indexer* 25, no. 4: 243-252.

Weinberg, Bella Hass. 2010. "Indexing: History and Theory", In *Encyclopedia of Library and Information Sciences*, ed. Marcia J. Bates and Mary Niles Maack. 3rd ed. Boca Raton, FL: CRC Press, 3: 2277-90.

Weinberg, Bella Hass. 2017. "Indexing: History and Theory." In *Encyclopedia of Library and Information Sciences*, ed. John D. McDonald and Michael Levine-Clark. 4th ed. Boca Raton, FL: CRC Press, 3: 1978-91.

Wellisch, Hans H. 1996. *Indexing from A to Z*. 2nd ed. New York: H.W. Wilson.

Wikipedia. 2017. "Concordance (Publishing)." https://en.wikipedia.org/wiki/Concordance%5F%28publishing%29

Wilson, Patrick. 1968. *Two Kinds of Power: An Essay on Bibliographical Control*. University of California Publications: Librarianship 5. Berkeley: University of California Press.

Wyman, Pilar, ed. 1999. *Indexing Specialties: Medicine*. Medford, NJ: Information Today.

Žumer, Maja. 2018. "IFLA Library Reference Model (LRM) Harmonisation of the FRBR Family." *Knowledge Organization 45: 310-8*.

https://doi.org/10.5771/0943-7444-2018-7-609 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY