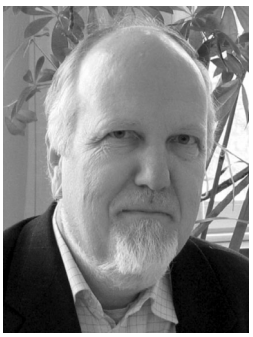
What is Knowledge Organization (KO)?

Birger Hjørland

Royal School of Library and Information Science,
6 Birketinget,DK-2300 Copenhagen S, Denmark,
<bh@db.dk>



Birger Hjørland, MA in psychology and PhD in Library and Information Science. Professor in Knowledge Organization (KO) at the Royal School of Library and Information Science in Copenhagen since 2001. Professor in KO at the University College in Borås, Sweden 2000-2001. Research librarian and coordinator of computer based information services at the Royal Library in Copenhagen 1978-1990. Taught information science at the Department of Mathematical and Applied Linguistics at the University of Copenhagen 1983-1986. 1974-1978 associate professor at RSLIS, and 1990-2000 head of department. He has published several papers on KO, including about 11 papers in the journal *Knowledge Organization* and the book *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science* (Greenwood Press 1997).

Hjørland, Birger. **What is Knowledge Organization (KO)?** *Knowledge Organization*, *35(3/2)*, 86-101. 45 references.

ABSTRACT: Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, databases, archives etc. These activities are done by librarians, archivists, subject specialists *as well as by computer algorithms*. KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations and concepts. There exist different historical and theoretical approaches to and theories about KO, which are related to different views of knowledge, cognition, language, and social organization. Each of these approaches tends to answer the question: "What is knowledge organization?" differently. LIS professionals have often concentrated on applying new technology and standards, and may not have seen their work as involving interpretation and analysis of meaning. That is why library classification has been criticized for a lack of substantive intellectual content. Traditional human-based activities are increasingly challenged by computer-based retrieval techniques. It is appropriate to investigate the relative contributions of different approaches; the current challenges make it imperative to reconsider this understanding. This paper offers an understanding of KO based on an explicit theory of knowledge.

1. Introduction: knowledge organization—the narrow and the broader meaning of the term

In the narrow meaning Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of "memory institutions" by librarians, archivists, information specialists, subject specialists, as well as by computer algorithms and laymen. KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations, works and concepts. Library and Information Science (LIS) is the central discipline of KO in this narrow sense (although seriously challenged by, among other fields, computer science).

In the broader meaning KO is about the social division of mental labor, i.e. the organization of universities and other institutions for research and higher education, the structure of disciplines and professions, the social organization of media, the production and dissemination of "knowledge" etc. A book such as Oleson & Voss (1979) *The Organization of knowledge in modern America, 1860-1920* is an example of the study of knowledge organization in the broad sense. We may distinguish between the social organization of knowledge on one hand, and on the other hand the intellectual or cognitive organization of knowledge. The broad sense is thus both about

how knowledge is socially organized and how reality is organized. The uncovering of structures of reality is done by the single sciences, e.g. chemistry, biology, geography and linguistics. Well known examples are the periodic system in chemistry and biological taxonomy. Generalized theories about the structure of reality, such as *the theory of integrative levels* first advanced by Auguste Comte belong to the philosophical disciplines "metaphysics" and "ontology."

While *Library and Information Science* (LIS) is the central discipline concerned with KO in the narrow sense of the word, other disciplines such as the sociology of knowledge, the single sciences and metaphysics are central disciplines concerned with KO in the broader sense of the word. The importance of regarding the broader field of KO is related to the question about how KO in the narrow sense can be developed. A central claim of this paper is that KO in the narrow sense cannot develop a fruitful body of knowledge without considering KO in the broader perspective. In other words: There exists no closed "universe of knowledge" that can be studied by KO in isolation from all the other sciences' study of reality.

Further description of the field of KO is dependent on the theoretical perspective, which is why we shall introduce the most important perspectives below.

2. Theoretical approaches to knowledge organization

KO has mainly been a practical activity without much theory. Miksa (1998, 49), for example, wrote:

> Now, we could simply conclude with Dolby and others that library classification continues mainly as a practical matter, that it is by and large devoid of substantive intellectual content, and that it continues merely because of inertia in a field in which classification schemes invented late in the nineteenth century continue to be used.

It has often been assumed that the practical organization of knowledge can be done by applying common sense or, in major research libraries and bibliographical databases, by employing subject specialists, who just apply their special knowledge. LIS professionals have often concentrated on applying new technology, software and standards. They have often seen themselves as applying standards for description of a relative objective nature. In other words practical KO may have been seen as a syntactic, rather than as a semantic activity as differentiated by Julian Warner (2007):

> Semantic labor is concerned with transformations motivated by the meaning or signified of symbols, while syntactic labor is determined by the form alone of symbols, operating on them in their aspect as signals. Semantic labor requires direct human involvement while originally human syntactic labor can be transferred to information technology, where it becomes a machine process.

Since the 1950s, computer scientists have been working with KO based on certain assumptions, mostly assuming that human classification and indexing will soon be made superfluous. A recent example (Sparck Jones 2005) is that automated systems based on relevance feedback from users might solve problems efficiently. Genuine theoretical contributions to KO are very rare, but seem mandatory in relation to the challenges with which this field is confronted. More and more people discuss the doomsday scenario for library and information science (cf., Bawden 2007). There exist many separated communities working with different technologies, but very little research about their basic assumptions and relative merits and weak sides. The problem is not just to formulate a theory, but to uncover theoretical assumptions in different practices, to formulate these assumptions as clearly as possible in order to make it possible to compare approaches.

A further problem is that the adherents of different approaches try to avoid criticism by incorporating ideas from competing approaches. The field cannot advance, however, without theoretical clarity, which is why it is important to describe different approaches in a way that they can be distinguished from each other and compared with each other. In other words: we have to examine and interpret different labels used for approaches very honestly and carefully. Otherwise we will stay in a very muddled field.

One way to classify approaches to KO was suggested by Broughton, Hansson, Hjørland and López-Huertas (2005):

1. The traditional approach to KO expressed by classification systems used in libraries and databases, including *DDC*, *LCC* and *UDC* (going back to about 1876).

2. The facet-analytical approach founded by Ranganathan about 1933 and further developed by the British Classification Research Group.
3. The information retrieval tradition (IR) founded in the 1950s.
4. User oriented and cognitive views gaining influence from the 1970s.
5. Bibliometric approaches following Garfield's construction of the *Science Citation Index* in 1963.
6. The domain analytic approach (first formulated about 1994).
7. Other approaches (among recent suggestions are semiotic approaches, "critical-hermeneutical" approaches discourse-analytic approaches and genre-based approaches. An important trend is also an emphasis on document representations, document typology and description, mark up languages, document architectures etc.).

Each of the 6 approaches (but not other approaches) will be presented and discussed below.

2.1 The traditional approach

It is difficult to define "the traditional approach" because there is no united theory that corresponds to this concept. If we disregard the other approaches to be introduced, what exist are mostly various different practices and some scattered suggestions on how to organize knowledge. Even a single system such as the *Dewey Decimal Classification* (*DDC*) has used quite different principles in various editions (cf., Miksa 1998). The classification researcher Vanda Broughton (2004, 143) wrote about one of the old established systems: "It is quite hard to discern any strong theoretical principles underlying *LCC* [*Library of Congress Classification*]." Also some formulations by S. R. Ranganathan (e.g., 1951) suggest that "traditional" systems seem to lack a theoretical foundation (in his eyes as opposed to his own approach).

Among the major figures in the history of KO, which can be classified as "traditional," are Melvil Dewey (1851-1931) and Henry Bliss (1870-1955). Eugene Garfield wrote about Bliss (1975, 252): "His goals and aspirations were different from those of Melvil Dewey, whom he certainly surpassed in intellectual ability, but by whom he was dwarfed in organizational ability and drive. Dewey was a businessman, but he was in no sense as profound in his accomplishments." This difference in the character of the two men is reflected in their approach to knowledge organization as also reflected by Miksa's (1998, 42-45) presentation of Melvil Dewey's business perspective. Dewey's business approach is hardly an intellectual approach on which the field can find a theoretical foundation for KO understood as an academic discipline. His interest was not to find an optimal system to support users of libraries, but rather to find an efficient way to manage library collections. He was interested in developing a system which could be used in many libraries, a standardized way to manage library collections.

DDC should thus be seen as the dream of the library administrator rather than the dream of the library user. It is not designed for any specific collection and must be seen as a compromise between different collections and corresponding scholarly interests. In order to minimize the workload in libraries, the system is conservative in the sense that it often prefers to avoid structural change. In other words, internal consistency over different editions has often taken priority compared to updating the system in order to make it more in accordance with the surrounding society. The user does not get a detailed, realistic view about relations between disciplines and fields of knowledge, but the library administrator gets a system in which most of the books are already classified by other libraries or agencies and which is used for both shelf arrangement and catalog searching. The library administrator may hire people from library schools, who know the system and may apply this knowledge in all the libraries using *DDC*. The system is thus also supporting professional interests. It probably represents a rationalization of library work more than anything else. Its main quality may be that it represents a standard and not a system optimized for browsing or retrieval for any particular interest. It should be added that what is today called *Library and Information Science*, LIS, was termed *library economy* in 1876 when the system was first published, which is also an indication of the administrative rather than the academic goals of the system. This may also explain why systems designed on the basis of more modern principles have not succeeded in influencing practice in libraries.

Among the critics of the *DDC* is Bernd Frohmann, who wrote (1994, 112-13):

> Dewey's subjects were elements of a semiological system of standardized, techno-bureaucratic administrative software for the library in its corporate, rather than high culture, incarnation.... Dewey emphasized more than once that his system maps no structure beyond its

> own; there is neither a "transcendental deduction" of its categories nor any reference to Cutter's objective structure of social consensus. It is content-free: Dewey disdained any philosophical excogitation of the meaning of his class symbols, leaving the job of finding verbal equivalents to others. His innovation and the essence of the system lay in the notation. The *DDC* is a poorly semiotic system of expanding nests of ten digits, lacking any referent beyond itself.... The conflict of interpretations over "subjects" became explicit in the battles between "bibliography" (an approach to subjects having much in common with Cutter's) and Dewey's "close classification." William Fletcher spoke for the scholarly bibliographer.... Fletcher's "subjects," like Cutter's, referred to the categories of a fantasized, stable social order, whereas Dewey's subjects were elements of a semiological system of standardized, technobureaucratic administrative software for the library in its corporate, rather than high culture, incarnation.

The quote from Frohmann shows that already when Melvil Dewey published his system there was a critique of the *DDC* as being empty and rather non-academic. Dewey's attitude may have influenced library philosophy and practice. LIS professionals may have seen their work more like a syntactical activity that an activity involving interpretation and analysis of meaning.

In order to identify an approach to KO which may deserve the label "the traditional approach," we shall turn to other scholars, including Henry Bliss. An important characteristic in his (and many contemporary thinkers of KO) was that the sciences tend to reflect the order of Nature and that library classification should reflect the order of knowledge as uncovered by science:

> Natural order → Scientific Classification → Library classification (KO)

The implication is that librarians, in order to classify books, should know about scientific developments. This should also be reflected in their education (Ernest Cushing Richardson, quoted from Bliss 1935, 2):

> Again from the standpoint of the higher education of librarians, the teaching of systems of classification ... would be perhaps better conducted by including courses in the systematic encyclopedia and methodology of all the sciences, that is to say, outlines which try to summarize the most recent results in the relation to one another in which they are now studied together.

This important principle has been implicit in the management of research libraries and bibliographic databases such as MEDLINE, in which subjects specialists are often hired to do the work in KO. The importance of subject knowledge has not been explicit in the following approaches to KO except in domain analysis (and outside LIS in certain computer approaches).

Among the other principles which may be attributed to the traditional approach to KO are:

- Principle of controlled vocabulary
- Cutter's rule about specificity
- Hulme's principle of literary warrant (1911)
- Principle of organizing from the general to the specific.

The principle of controlled vocabulary is essentially a way of avoiding synonyms and homonyms as indexing terms by using standardized vocabulary. Cutter's rule states that it is always the most specific, most appropriate expressions that should be looked up in the vocabulary of notations and assigned to documents. In this way the expressions for the topics to be made retrievable are rendered most predictable. The term "literary warrant" as well as the basic principle underlying this expression was introduced by E. Wyndham Hulme (1911, 447). Hulme discusses whether, for example, the periodic system of chemistry should be used for book classification. He writes (1911, 46-47):

> In Inorganic Chemistry what has philosophy to offer? [Philosophy here meaning science, which produced the periodic system]. Merely a classification by the names of the elements for which practically no literature in book form exists. No monograph, for instance, has yet been published on the Chemistry of Iron or Gold.... Hence we must turn to our second alternative which bases definition upon a purely literary warrant. According to this principle definition is merely the result of an accurate survey and measurement of classes in literature. A class

> heading is warranted only when a literature in book form has been shown to exist, and the test of the validity of a heading is the degree of accuracy with which it describes the area of subject matter common to the class. Definition [of classes or subject headings], therefore, may be described as the plotting of areas pre-existing in literature. To this literary warrant a quantitative value can be assigned so soon as the bibliography of a subject has been definitely compiled. The real classifier of literature is the book-wright, the so-called book classifier is merely the recorder.

The principle of ordering from general subjects to specific subjects is generally acknowledged and may be related to an essentialist way of understanding.

Today, after more than 100 years of research and development in LIS, the "traditional" approach still has a strong position in KO and in many ways its principles still dominate. The traditional approach, however, shows signs of a certain vagueness in its theoretical and methodological basis. Is it subject knowledge rather than competency in KO that marks the construction and administration of knowledge organizing systems? Often it seems to be assumed that that the organization of knowledge is just a matter of "reading" *the* correct relations between concepts. There is not much indication of how this is done. Although debates about the philosophy of science, e.g. in relation to positivism, was not unknown among the founding fathers of knowledge organization, they were not particularly clear on this point and the same is also the case with the ordinary practice of KO. It is with the development of the domain-analytic approach that the question about the subjectivity and objectivity of KO in a systematic way is first built into the methodological foundation of KO.

2.2 The facet-analytical approach

The date of the foundation of this approach may be chosen, for example, as the publication of S. R. Ranganathan's *Colon Classification* in 1933. The approach has been further developed by, in particular, the British Classification Research Group. In many ways this approach has dominated what might be termed "modern classification theory." The BC2 system is probably today the theoretically most advanced system based on this theory (and has also contributed to the further development of this approach).

The best way to explain this approach is probably to explain its analytico-synthetic methodology. The meaning of the term "analysis" is: breaking down each subject into its basic concepts. The meaning of the term synthesis is: combining the relevant units and concepts to describe the subject matter of the information package in hand. Given subjects (as they appear in, for example, book titles) are first analyzed into a few common categories, which are termed "facets." Ranganathan proposed his PMEST formula: Personality, Matter, Energy, Space and Time:

- Personality is the distinguishing characteristic of a subject
- Matter is the physical material of which a subject may be composed
- Energy is any action that occurs with respect to the subject
- Space is the geographic component of the location of a subject.
- Time is the period associated with a subject.

The British Classification Research Group (CRG) expanded this list, but here we shall only consider the original one. The first assumption is that all subjects can be analyzed in a way that fits into these five categories. Those categories have been developed before the books have been written and arrived in the library. In other words, they are neither dynamically developed nor empirically given: they are logical, a priori categories. Each category (facet) has in principle its own classification or lists of symbols. A given document is classified by taking one or more symbols from the appropriate facets and combining them according to certain rules. This combination is called notational synthesis. The idea is that the same building blocks can be used for all purposes. The underlying philosophical assumption is that elements do not change their meaning in different contexts. This assumption has never, as far as I know, been discussed in the literature. According to modern theories of meaning it is a rather problematic assumption.

Ranganathan has had many followers in LIS. It has however, been extremely difficult to trace critical examinations of this approach. Very few researchers have had the broader knowledge which enabled them to consider this approach in relation to fields like philosophy and linguistics. Among the few who have done this is Moss (1964) who found that Ranganathan based his system of five categories on that of Aristotle without recognizing this. Another critical voice is Francis Miksa, who, for example, wrote (1998, 73):

> In the end, there is strong indication that Ranganathan's use of faceted structure of subjects may well have represented his need to find more order and regularity, in the realm of subjects, than actually exist.… Ranganathan vigorously pursued the goal of finding *one* best subject classification system.

Hjørland (2007b, 382-84) related the basic philosophy of facet analysis to the philosophy of semantic primitives and thus to a broader theory of semantics. According to his analysis, semantic elements are not direct attributes of language, but are related to models of reality, which are then expressed in language. Chemical compounds may, for example, be expressed in chemical formulae by chemical elements. Chemical elements are discovered and named by chemists; they are not given elements in natural languages. The names of the chemical elements are in this case the semantic primitives. Semantic relations, including the relation between elements and composed expressions, are thus connected to theories of reality.

S. R. Ranganathan wrote in his 'Philosophy of Library Classification' (1951, 87ff.):

> An enumerative scheme with a superficial foundation can be suitable and even economical for a closed system of knowledge.... What distinguishes the universe of current knowledge is that it is a dynamical continuum. It is ever growing; new branches may stem from any of its infinity of points at any time; they are unknowable at present. They can not [sic] therefore be enumerated here and now; nor can they be anticipated, their filiations can be determined only after they appear.

Ranganathan thus expresses the views:

1. That enumerative systems have a superficial foundation;
2. That the discovery of new knowledge cannot be anticipated in an enumerative system; and,
3. That the discovery of new knowledge can be anticipated in a faceted system (based on the view that new knowledge is formed by combination of a priori existing categories).

These views reveal some basic assumptions in the facet-analytic approach. The difference between the theoretical foundations of enumerative systems compared to faceted systems is not that the former have a superficial foundation while the latter have a profound foundation. The basic questions in knowledge organization are shared by both approaches: How terms are selected and defined and their semantic relations established. This is not a purely logical matter, but largely an empirical question. While it is correct that it may be easier to combine existing elements to form new classes and thus easier to place new subjects in faceted systems, it is of course impossible for any system to anticipate the discovery of new knowledge. The belief that this should be possible reveals that part of the philosophy of facet analysis is without contact with the real world.

La Barre (2006) found that faceted techniques are increasingly being used in the design of web-pages. A specific format, XFML, a simple XML format for exchanging metadata in the form of faceted hierarchies has been developed (Van Dijck 2003). The technique is thus very alive and in use.

2.3 The information retrieval tradition (IR)

Information retrieval (IR) and knowledge organization (KO) are normally considered two different—although strongly related—subfields within Library and Information Science (LIS)—related respectively to search labor and description labor (Warner 2002). They are, however, trying to solve the same kind of problems: enabling users to find relevant information. For this reason we have to consider them competing approaches, and thus try to evaluate their relative strengths and weaknesses. The question then becomes: how can IR be characterized as an approach relative to the other approaches discussed?

One way to do this has been to make a distinction between the "physical paradigm" (or "system-driven paradigm") on one side and "user-oriented" or "cognitive paradigm" on the other. The IR tradition has been understood as "systems driven" as if the system makes a decision of what to present for the users (Gruzd 2007, 758).

> In the conventional system-oriented view, a "perfect" system is defined as one that finds the best match between a user's stated request and documents from a collection. This view has proven to be very limiting. It has led many researchers to focus only on how to improve various aspects of document representations and the matching algorithms. As a result, the system-oriented approach to IR tends to disre-

> gard users' cognitive behaviors as well as the problem-solving context in which an IR process is being carried out. It has become evident that to succeed, IR researchers need to look beyond machine algorithms.

This distinction between "the system-oriented view" and "the user-oriented view" may, however, represent a misinterpretation. The difference between the Cranfield experiments and user-oriented views is first and foremost that the Cranfield experiments are based on expert evaluations of recall and precision, while the user-oriented views are based on users' evaluation. It is never the technology that makes the decision of what is relevant. The technology is just constructed on the basis of some views of what is relevant and how this can be measured. Neither the system-oriented view nor the user-oriented view has considered the epistemological problem: How are answers to queries related to different theories or views?

Important in the IR tradition have been, among others, the Cranfield experiments, which were founded in the 1950s, and the TREC experiments (Text Retrieval Conferences) starting in 1992. It was the Cranfield experiments which introduced the famous measures "recall" and "precision" as evaluation criteria for systems efficiency. The Cranfield experiments found that classification systems like *UDC* and facet-analytic systems were less efficient compared to free-text searches or low level indexing systems ("UNITERM"). The Cranfield I test found according to Ellis (1996, 3-6) the following results:

UNITERM	82.0% recall
Alphabetical subject headings	81.5% recall
UDC	75.6% recall
Facet classification scheme	73.8% recall.

Although these results have been criticized and questioned, the IR tradition became much more influential while library classification research lost influence. The dominant trend has been to regard only statistical averages. What has largely been neglected is to ask: Are there certain kinds of questions in relation to which other kinds of representation, for example, controlled vocabularies, may improve recall and precision?

Julian Warner (2002) has characterized the dominant IR tradition with the word "query transformation" meaning that systems automatically transform a query to a set of relevant references. He contrasts this principle by what he terms "selection power," a principle that, according to him has been valued in traditional library work.

Although thesauri were developed in the IR tradition, this is the exception that confirms the rule: The IR approach may be characterized as generally sceptical of all forms of human interpretation, indexing and classification. Its focus has clearly been on free-text retrieval: the assumption that texts contain all necessary information needed to retrieve them. Recently Karen Sparck Jones (2005) wrote that traditional (pre-)classification probably is obsolete and may be replaced by new promising techniques such as relevance feedback. If Sparck Jones' view is typical of the IR approach, then a criticism of this view may provide the basis of an alternative to the IR approach. In fact, two basic criticisms of relevance feedback can be summarized:

1. Relevance feedback is based on certain premises about users' knowledge that are largely unexplored and may turn out to be highly unrealistic: If users do not have the necessary knowledge to classify a domain, they cannot distinguish relevant and non-relevant documents and are thus unable to provide useful feedback.
2. Relevance feedback represents unspecified and unclear semantic relations between documents considered relevant. Why prefer a kind of system implying unspecified relations rather than specified and user-controlled relations?

In conclusion: The IR tradition has generally been based on positivist assumptions: that optimal retrieval can be determined by retrieval tests without considering different views or "paradigms" and without considering text corpora as a merging of different views each putting different meanings to terms. In other words, it has mainly been based on statistical averages, and has neglected to investigate how different kinds of representation and algorithms may serve different views and interests.

2.4 User-oriented views

In some sense, all approaches to KO may agree about the goal that systems and processes are aimed at fulfilling users' "information needs." For example, facet analytic researchers may rightly claim that users benefit from well structured systems, which is why this approach is "user-oriented" or "user friendly." If the term "user-oriented" is to be a meaningful label

for an approach, it needs to be defined in a more precise way. We need to distinguish at least the following meanings:

- User-friendly knowledge organization
- Market-oriented knowledge organization
- Knowledge organization based on empirical studies of users
- Knowledge organization done by users (e.g. the recent trend in folksonomies).

The best way to define this approach is probably by method: Systems based upon user-oriented approaches must specify how the design of a system is made on the basis of empirical studies of users.

User studies demonstrated very early that users prefer verbal search systems as opposed to systems based on classification notations. This is one example of a principle derived from empirical studies of users. Adherents of classification notations may, of course, still have an argument: That notations are well-defined and that users may miss important information by not considering them.

In order to consider the function of empirical user studies it might be fruitful to consider the development of a field such as biological systematics. Table 1 shows Mishler's (2000) historical outline of this domain:

Historical periods in biological systematics (after Mishler, 2000)
1) Pre-history. Folk classifications
* 2) Ancient Greeks through Linneaeus: Essentialism
* 3) Natural system. Overall resemblance; "importance."
4) Darwin. Evolutionary language added (Only a superficial effect for a long time, cf. 6)
5) Numerical Phenetics. Computers added. (Only a superficial effect)
* 6) Phylogenetic systematics (Cladistics). [A late Darwinian approach]
[*7) Systematics based on DNA-analysis]
* argued by Mishler (2000) to be the only true revolutions in the conceptual bases of systematics

Table 1. *Mishler's (2000) outline of the domain of biological systematics*

The table shows how "folk classification" was succeeded by an essentialist classification from Aristotle to Linné, then by a natural classification [founded by de Jussieu] and later by phylogenetic systematics and DNA-analysis. Thus, according to this outline folk classification represented a pre-scientific period. One might ask: Are classifications based on empirical information from users to enjoy the same status as folk classifications (i.e., to represent a pre-scientific form of knowledge organization)? Do adherents of user-oriented views find that it is better to base classification systems for libraries and bibliographical databases on folk classifications and user studies rather than on scientific methods?

It is strange that somebody seems to believe so. Are amateurs supposed to know better? In some cases, of course, it may be hard to find experts among established researchers. In the case of music, established researchers have not until recently regarded popular music and experts have had to be found in other circles, for example, among journalists and the users themselves. Even in that case, it is probably not the average user who knows about relevant genre concepts, but some experts among the users. That being said, it must be admitted that some serious researchers do regard biological folk-classification equal to scientific classification (Dupre 2006).

Hjørland (2007a) found that user-oriented views seem to have driven out the study of documents and that they have made some problematic critiques of "the bibliographical paradigm." User-oriented views are often contrasted with "the systems driven approach" which is again associated with the Cranfield experiments (Hildreth 2001):

> Theoretically, the Cranfield model relies almost entirely on the attractive, but troublesome concept of relevance. Furthermore, two key assumptions underlie the Cranfield model: users desire to retrieve documents relevant to their search queries and don't want to see documents not relevant to their queries, and document relevance to a query is an objectively discernible property of the document. Neither of these two assumptions has stood the test of time, experience and astute analysis.

The question whether a 'document relevance to a query is an objectively discernible property of the document' is an epistemological issue, which, according to Hildreth (2001), is differently perceived in the Cranfield experiments and in the user-oriented tradi-

tion. Both traditions have, however, almost totally neglected epistemological theories and thus confused the concept of 'users' and the concept of 'subjectivity.' Studying users and their psychology is in user studies mixed up with studying subjectivity in different views on knowledge. In the Cranfield experiments relevance was evaluated by subject experts, while the user-oriented approach used users for evaluation (often using the same measures of recall and precision). It is correct that Cranfield by applying expert evaluations expected the system to provide relevant references for all users, i.e. assuming a kind of a standard user. However, in the user-oriented framework this is not very different. Algorithms are often constructed on the basis of an average of users' evaluations. What has been neglected in both traditions is to develop different representations of the same documents to serve different users. Both traditions are rooted in the positivist understanding that a representation is objective and neutral and that "one size fits all."

2.5 Bibliometric approaches

These approaches are primarily based on using bibliographical references to organize networks of papers, mainly by bibliographic coupling (introduced by Kessler 1963) or co-citation analysis (independently suggested by Marshakova 1973 and Small 1973). In recent years it has become a popular activity to construe bibliometric maps as structures of research fields.

Two considerations are important in considering bibliometric approaches to KO:

1. The level of indexing depth is partly determined by the number of terms assigned to each document. In citation indexing this corresponds to the number of references in a given paper. On the average, scientific papers contain 10-15 references, which provide quite a high level of depth.
2. The references, which function as access points, are provided by the highest subject-expertise—the experts writing in the leading journals. This expertise is much higher than that which library catalogs or bibliographical databases typically are able to draw on.

The main advantages and disadvantages in this approach are summarized in Table 2.

Advantages	Disadvantages
– Citations are provided by highly qualified subject specialists – The number of references reflect the indexing depth and specificity (average in scientific papers is about 10 references per article) – Citation indexing is a highly dynamic form of subject representation – References are distributed in papers which allows the utilization of paper structure in the contextual interpretation of citations – Scientific papers form a kind of self-organizing system	– The relation between citations and subject relatedness is indirect and somewhat unclear (related to the difference between social organization of knowledge and intellectual organization of knowledge) – Does not provide clear logical structure with mutually exclusive and collectively exhaustive classes – Explicit semantic relations are not provided – Namedropping and other forms of imprecise citations may cause noise

Table 2. *Advantages and disadvantages of the bibliometric approach to KO*

Data coverage is an important problem in the bibliometric approach. Bibliometric maps are extremely vulnerable to how journals are selected. There is no objective and neutral way to select journals as data for bibliometric analysis. If, for example, *Knowledge Organization* is excluded from LIS, then classification researchers like Ranganathan will be relatively underrepresented, because they are more often cited in this journal. This does not, however, imply, that bibliometrics is totally subjective and arbitrary. By working with different methods and by doing iterative investigations strong arguments may be made concerning data coverage.

Schneider (2004) found that bibliometric methods can be used to provide candidate terms for thesauri. Bibliometric maps may, however, be considered a knowledge organizing tool in their own right, one that can supplement thesauri, whether or not they can be "verified" by thesauri. Typically bibliometric maps show networks of cooperating authors, while thesauri show ontological links. Analytically we may make a distinction between the intellectual organization of knowledge and the social organization of knowledge and it may be argued that bibliometrics is closer to the social pole. Bibliometric methods may thus provide supplementary information that is useful in their own right.

2.6 The domain analytic approach (DA)

The domain analytic approach is an approach formulated at the beginning of the 1990s as an alternative to the dominant cognitive view in LIS. Here, it will be presented more specifically as an alternative to the other approaches to KO previously discussed.

Domain analysis is a sociological-epistemological standpoint. The indexing of a given document should reflect the needs of a given group of users or a given ideal purpose. In other words, any description or representation of a given document is more or less suited to the fulfillment of certain tasks. A description is never objective or neutral, and the goal is not to standardize descriptions or make one description once and for all for different target groups.

The development of the Danish library "KVINFO" may serve as an example that explains the domain-analytic point of view. KVINFO was founded by the librarian and writer Nynne Koch and its history goes back to 1965. Nynne Koch was employed at the Royal Library in Copenhagen in a position without influence on book selection. She was interested in women's studies and began personally to collect printed catalog cards of books in the Royal Library, which were considered relevant for women's studies. She developed a classification system for this subject. Later she became the head of KVINFO and got a budget for buying books and journals, and still later, KVINFO became an independent library. The important theoretical point of view is that the Royal Library had an official systematic catalog of a high standard. Normally it is assumed that such a catalog is able to identify relevant books for users whatever their theoretical orientation. This example demonstrates, however, that for a specific user group (feminist scholars), an alternative way of organizing catalog cards was important. In other words: Different points of view need different systems of organization.

DA is the only approach to KO which has seriously examined epistemological issues in the field, i.e. comparing the assumptions made in different approaches to KO and examining the questions regarding subjectivity and objectivity in KO. Subjectivity is not just about individual differences. Such differences are of minor interest because they cannot be used as guidelines for KO. What seems important are collective views shared by many users. A kind of subjectivity about many users is related to philosophical positions. In any field of knowledge different views are always at play. In arts, for example, different views of art are always present. Such views determine views on art works, writing on art works, how art works are organized in exhibitions and how writings on art are organized in libraries (see Ørom 2003). In general it can be stated that different philosophical positions on any issue have implications for relevance criteria, information needs and for criteria of organizing knowledge.

The representation of a document is made in order to enable users to make relevant discriminations. The document should be looked upon with the eyes of potential users. In a feminist library, for example, a book should be indexed by anticipating what it might contribute to feminist scholarship. This may sound strange, but in many situations this is obvious and the natural thing to do. This view is known in the literature as "request oriented indexing." The core of indexing is, as stated by Rowley & Farrow to evaluate a paper's contribution to knowledge and index it accordingly (2000, 99):

> 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.

Or, with the words of Hjørland (1992, 1997): "the subjects of a document are its informative potentials." A more simple way to put it: the indexer should ask "what use can be made of this particular document—relative to other documents?"

The kind of information which is judged relevant for a given task depends on the theory of the person doing the judgment. If one believes that schizophrenia is caused by a problematic communication between mother and child, then studies of family interaction are evaluated as relevant. If, on the other hand, one believes schizophrenia is caused by genetic factors, then the study of genes becomes most relevant. The criteria used to represent documents are thus in principle the same criteria that are implied by current scientific theories. (This is why citation indexes have an advantage by their extremely dynamic way of indexing).

The facet analytic point of view takes as the point of departure the terminology of a given field; little is said, however, about how the terminology is to be selected. Domain analysis acknowledges a dilemma, a kind of chicken-and-egg problem, and a hermeneutic circle: In order to select the terminology, one needs to have an understanding of the field. But in order to get an understanding of a field, one needs to know

about its concepts. The way this has to be solved is by using iterative methods. DA assumes that different approaches (or "paradigms") exist in all domains of knowledge and have to be identified. They are not equally distributed in the literature or among the users, which is why so-called representative samples cannot be used. (If they were used some important views would not be properly represented). Different approaches in a given domain have to be actively searched for. Any system of knowledge organization is always biased toward some philosophical position. There is no neutral platform from which knowledge can be organized. The task is to mediate between different views and to develop arguments for a point of view that is in accordance with the goals and values of the organization for which the system is developed.

3. Some concepts considered units in KO: "document," "information," and "knowledge"

The field of knowledge organization consists of some units, elements or entities to be organized and some relations between those units (e.g., semantic relations and bibliographic relationships). If we look at an introductory paper on knowledge organization such as Anderson (2003), many different suggestions about what is organized in KO is given (Anderson 2003, 471 emphasis added):

> The description (indexing) and organization (classification) for retrieval of messages representing *knowledge*, *texts* by which knowledge is recorded and *documents* in which texts are embedded. Knowledge itself resides in minds and brains of living creatures…. Its organization for retrieval via short- and long-term memory is a principal topic of cognitive science. Library and information science deals with the description and organization of the *artifacts* (messages, texts, documents) by which knowledge (including feelings, emotions, desires) is represented and shared with others. These knowledge resources are often called information resources as well. Thus 'knowledge organization' in the context of library and information science is a short form of '*knowledge resources* organization'. This is often called '*information* organization.'

This quotation provided six different terms (emphasized) for consideration as candidate terms for the units in KO. Other views may be found scattered in different literatures. On the basis of the literature, many candidate terms may be considered. In this paper, only three of those terms will be briefly discussed: Document, information and knowledge.

3.1 Document

Library science was mainly about the organization of books and book representations on shelves and in catalogs. Bibliography included articles and other kinds of documents listed in bibliographies. Archives organize "records," while museums organize physical objects. The documentalists made a generic concept "document" to include not just books, articles, "records" and objects such as globes, but any kind of material indexed to serve as some kind of documentation, including pictures, maps and globes. Even animals were considered documents (if captured and kept in a zoo). The concept of document is important but lost much influence with the entrance of computers in the 1950s, but has recently had an important renaissance.

3.2 Information

Computer scientists ignored earlier conceptual work in the fields of library science and documentation and just talked about "information storage and retrieval." To talk about information rather than documents may have raised the status of the dusty profession of library science/documentation, as suggested by Spang-Hanssen (2001). Intellectually, however, it has brought much confusion and may have misled KO from its proper theoretical basis. Experiments with "information retrieval" in the 1950s-1960s were mainly based on bibliographical databases. The transformation to electronic media did not change the nature of what was represented. The use of the term "information" was associated with the belief that Shannon's "information theory" was a long-needed answer to a theory also about libraries and scholarly communication. The expectations were never met, however, and the talk about information rather than documents has not strengthened the theoretical basis of the field (although, of course information theory is valuable in computer science for technical problems such as measuring the storage capacity of disks). Documents are more related to the concept and theory of semiotics (the field about signs), which may turn out to be a more fruitful theoretical frame for KO.

3.3 Knowledge

The term KO originated in the library field. It seems to have been established around 1900 by people like Charles A. Cutter and Ernest Cushington Richardson and stabilized by W. C. Berwick Sayers and Henry Bliss. Bliss' book (1929) *The organization of knowledge and the system of the sciences* represents one of the main intellectual contributions in the field. All of these authors argued that book classification is based on knowledge organization as it appears in science and scholarship. The best way to organize books in libraries (and document representations in bibliographies) was to make the library classification reflect a scientific classification which, in turn, was supposed to reflect the nature of reality.

Cutter, Bliss, and other important classification researchers from the period of the second half of the 19th century and the first half of the 20th century, realized that what is organized cannot be taken as absolute truth. However, Bliss believed that knowledge was relatively safe and true, which is why a kind of consensus could be established. Because of this, Bliss and his contemporary chose the term "knowledge organization," "knowledge" understood in the Platonic tradition as "verified, true belief."

In his preface to Bliss (1929), the philosopher John Dewey wrote (Dewey 1929, viii):

> A classification of books to be effective on the practical side must correspond to the relationships of subject-matters, and this correspondence can be secured only as the intellectual, or conceptual, organization is based upon the order inherent in the fields of knowledge, which in turn mirrors the order of nature.

This quotation is in accordance with the traditional view of knowledge as a neutral and objective reflection of reality. It is, however, a bad representation of John Dewey's pragmatic view of knowledge and of classification, as demonstrated by another quotation (Dewey 1920/1948, 151-54):

> No sensible person tries to do everything. He has certain main interests and leading aims by which he makes his behavior coherent and effective. To have an aim is to limit, select, concentrate, group. Thus a basis is furnished for selecting and organizing things according as their ways of acting are related to carrying forward pursuit. Cherry trees will be differently grouped by woodworkers, orchardists, artists, scientists and merry-makers. To the execution of different purposes different ways of acting and re-acting on the part of trees are important. Each classification may be equally sound when the difference of ends is borne in mind.
>
> Nevertheless there is a genuine objective standard for the goodness of special classifications. One will further the cabinetmaker in reaching his end while another will hamper him. One classification will assist the botanist in carrying on fruitfully his work of inquiry, and another will retard and confuse him. The teleological theory of classification does not therefore commit us to the notion that classes are purely verbal or purely mental. Organization is no more merely nominal or mental in any art, including the art of inquiry, than it is in a department store or railway system. The necessity of execution supplies objective criteria. Things have to be sorted out and arranged so that their grouping will promote successful action for ends. Convenience, economy and efficiency are the bases of classification, but these things are not restricted to verbal communication with others nor to inner consciousness; they concern objective action. They must take effect in the world.
>
> At the same time, a classification is not a bare transcript or duplicate of some finished and done-for arrangement pre-existing in nature. It is rather a repertory of weapons for attack upon the future and the unknown. For success, the details of past knowledge must be reduced from bare facts to meanings, the fewer, simpler and more extensive the better.

This quotation clearly demonstrates that John Dewey did not accept the mirror metaphor of knowledge, or, as he expressed it: "a bare transcript or duplicate of some finished and done-for arrangement pre-existing in nature."

For KO this issue is important. Two different views of knowledge can be contrasted:

1. "Positivist view": Knowledge and KO as "a bare transcript or duplicate of some finished and done-for arrangement pre-existing in nature."
2. "Pragmatic view": Knowledge and KO as something constructed to deal with some human needs and interests.

The pragmatist view of knowledge is also connected with "fallibilism," the view that scientific research is never to be taken finally proved, that new evidence may change scientific beliefs. The implication of fallibilism is that we cannot understand the documents as representing knowledge, as traditionally understood. We should not talk about knowledge or knowledge organization, but about knowledge claims and the organization of knowledge claims. The implication is that each knowledge claim is supported by and connected with arguments, theories and world views. If this is recognized by the people performing KO, then the activity is not based on "positivism."

4. Fields contributing to knowledge organization

Knowledge Organization is not just something the LIS-profession can do without considering research in other domains, for example, computer science, linguistics and natural language processing, theory of knowledge, theory of social organization etc. In particular an understanding of the nature of knowledge, cognition, language and social organization is decisive for the understanding of KO and thus for the ability to design, evaluate and use knowledge organizing processes and knowledge organizing systems. Many fields may have an interest in the defining questions of knowledge organization or may be considered related disciplines. This issue has already been introduced above, for example, the role of the sociology of knowledge, the single sciences and metaphysics/ontology.

A few words about the concept of discipline in relation to this issue: much knowledge is today scattered in different disciplines. Library schools have traditionally educated librarians and information specialists, schools of language for special purposes have educated translators, business schools have educated information managers, schools of computer science have educated software engineers etc. In many ways much of what they have been working with is based on the same kind of theoretical knowledge. Their separation has posed a problem rather than provided a fruitful development of separate fields. This journal (*Knowledge Organization*) sometimes publishes information related to the field of terminology, but this is an exception that confirms the rule that the two fields are separated. In each discipline, there is a need for theoretical clarification about the fundamental problems in knowledge, cognition, communication, language and social organization, which are common to all these disciplines.

Our journal, *Knowledge Organization*, has the subtitle: *International Journal. Devoted to Concept Theory, Classification, Indexing, and Knowledge Representation*. Each of these fields may be studied from different perspectives. First, they may be studied from different disciplinary perspectives. Concepts, for example, may be studied by psychology, by linguistics, by philosophy, by sociology, by artificial intelligence and so on. Each of these fields tends to emphasize different aspects of concepts. At the same time, however, each of those fields struggles with the same fundamental problems regarding the nature of concepts. Second, there are basic (epistemological) theories of concepts that are common to all those fields and within each field competing for attention. It is this epistemological level that is most important. If a strong theory is developed at this level, all the involved disciplines will benefit in a very important way.

Let us consider linguistics_as an example. First, linguistics is a discipline (studying language) but language is also studied by, for example psychology, and sociology. Linguistics should be extremely important for LIS and KO because of the dominance of texts in libraries and because most intermediating activity is based on language. The case is, however, that linguistic research is very seldom cited in the literature of LIS (cf. Warner 1991). Why is this?

The influential computer scientist Gerald Salton expressed pessimism concerning the usefulness of linguistics in information science. In the words of the Danish linguist and information scientist Henning Spang-Hanssen (1974, 17, translated by BH):

> In this connection it is important to realize that the points of view, which have been dominant within linguistics in the last 10-15 years, in particular in the USA (i.e. Noam Chomsky's school of generative grammar) have not had practical influence worth mentioning in relation to natural language processing. In its theoretical foundation and in the technicalities (such as the writing of rules in algorithmic form) exist important similarities between generative grammar and electronic data processing. Natural language processing seems, however, in practice still to depend on traditional categories of grammar and traditionally formed dictionaries. This demonstrates in my opinion the problems related to automation of text—as opposed to problems related to automation of mathematical computations—are fundamental and thus can-

not be eliminated just by computer-oriented versions of linguistics.

I thus share with Gerald Salton his pessimism about the usefulness of recent linguistics in relation to automated documentation. However, Salton seems to identify linguistics with modern American linguistics and thus to miss the knowledge, which was gained before generative grammar evolved or which was gained in other countries such as Scandinavia.

In order to understand the relation between linguistics and LIS it is thus important to understand that both fields are influenced by changing epistemological views and interdisciplinary trends. Epistemology is simply a deeper way to understand both fields. This situation unfortunately makes it more difficult for all parties, including knowledge organization. In order to draw from related fields such as linguistics, we simply have to find a satisfactory metatheory before we can do so. In line with what is written earlier in this paper, I find that such a metatheory must be related to pragmatism.

5. Conclusion

Knowledge Organization is one among many contemporary fields which try to play a role in the future environments of communicating and exchanging knowledge. Among the competitors are Knowledge Management and Computer Science. Much knowledge may be shared among such fields, but is important for each field to develop a clear identity and a history of its own. KO has in particular been connected with LIS and has aimed at supporting learning and research activities, which may be one of the important pillars on which to base the field. Another related pillar is the concept of knowledge and theories of knowledge. Knowledge Organization may have a valuable theoretical base in theory of knowledge, which may be the reason why we should stick to this label as the name of our field.

References

Anderson, J.D. 2003. Organization of knowledge. In Feather, John and Sturges, Paul, eds., *International encyclopedia of information and library science.* 2nd ed. London: Routledge, 471-90.

Bawden, David. 2007. The doomsday of documentation? *Journal of documentation* 63:2, (editorial).

Bliss, Henry Evelyn. 1929. *The organization of knowledge and the system of the sciences*. With an introduction by John Dewey. New York: Henry Holt and Co.

Bliss, Henry Evelyn. 1935. *A system of bibliographic classification*. New York: H. W. Wilson.

Broughton, Vanda. 2004. *Essential classification*. London : Facet Publishing.

Dupré, J. 2006. Scientific classification. *Theory, culture & society* 23:2-3, 30-32.

Broughton, Vanda, Hansson, Joacim, Hjørland, Birger and López-Huertas, Maria J. 2005, Knowledge organisation: Report of working group 7. In Kajberg, L. and Lørring, L., eds., *European curriculum reflections on education in library and information science*, Royal School of Library and Information Science, Copenhagen. Available at http://www.db.dk/LIS-EU/workshop.asp

Cole, Jonathan R. and Cole, Stephen. 1973. *Social stratification in science.* Chicago, IL: University of Chicago Press.

Dewey, John. 1929. Introduction. In Bliss, Henry Evelyn, *The organization of knowledge and the system of the sciences*. New York: Henry Holt and Company.

Dewey, John. 1920/1948. *Reconstruction in philosophy.* Enlarged edition. New York: Beacon, 1948. (Original work published 1920).

Dolby, R. G. Alex. 1979. Classification of the sciences: the nineteenth century tradition. In Ellen, R. F. & Reason, D., eds., *Classifications in their social contexts*. New York: Academic Press, 167-93.

Dupré, John. 2006. Scientific classification. *Theory, culture & society* 23:2-3, 30-32.

Ellis, David. 1996. *Progress and problems in information retrieval*. London: Library Association Publishing.

Ereshefsky, Marc. 2000. The poverty of the Linnaean hierarchy: a philosophical study of biological taxonomy. Cambridge: Cambridge University Press.

Frohmann, Bernd. 1994. The social construction of knowledge organization: the case of Melvil Dewey. In Albrechtsen, Hanne and Oernager, Susanne, eds., *Knowledge organization and quality management: Proceedings of the Third international ISKO Conference 20-24 June 1994 Copenhagen.* Frankfurt/Main: INDEKS Verlag, 109-17.

Garfield, Eugene. 1975. The "other" immortal: a memorable day with Henry E. Bliss. *Current contents* 15, 7-8. Reprinted in: *Essays of an information scientist* 2 (1974-76), 250-251. Available at

http://www.garfield.library.upenn.edu/essays/v2p250y1974-76.pdf

Gruzd, Anatoliy. 2007. Book review of 'New Directions in Cognitive Information Retrieval'. *Journal of the American Society for Information Science and Technology* 58:5, 758-60.

Hildreth, Charles R. 2001. Accounting for users' inflated assessments of on-line catalogue search performance and usefulness: an experimental study. *Information research* 6:2. Available at http://InformationR.net/ir/6-2/paper101.html

Hjørland, Birger. 1992. The concept of "subject" in information science. *Journal of documentation* 48:2, 172-200. Available at http://www.db.dk/bh/Core%20Concepts%20in%20LIS/1992JDOC_Subject.PDF

Hjørland, Birger. 1997. *Information seeking and subject representation. An activity-theoretical approach to information science*. Westport & London: Greenwood Press.

Hjørland, Birger. 2005ff. Lifeboat for Knowledge Organization. Available at http://www.db.dk/bh/lifeboat%5Fko/home.htm

Hjørland, Birger. 2007a. Arguments for 'the bibliographical paradigm'. Some thoughts inspired by the new English edition of the UDC. *Information research* 12:4, paper colis06. Available at http://informationr.net/ir/12-4/colis/colis06.html

Hjørland, Birger. 2007b. Semantics and knowledge organization. *Annual review of information science and technology* 41, 367-405.

Hjørland, Birger & Nissen Pedersen, Karsten. 2005. A substantive theory of classification for information retrieval. *Journal of documentation* 61:5, 582-97. Available at http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Hjorland%20&%20Nissen.pdf

Hulme, E. Wyndam 1911. Principles of book classification. *Library Association record* 13: Oct. 1911, 354-58; Nov. 1911, 389-94; and Dec. 1911, 444-49.

Kessler, Myer Mike. 1963. Bibliographic coupling between scientific papers. *American documentation* 14, 10-25.

La Barre, Kathryn. 2006. *The use of faceted analytico-synthetic theory as revealed in the practice of website construction and design*. Ph.D. dissertation, Indiana University.

Marshakova, I. V. 1973. A system of document connection based on references. *Scientific and technical information serial of VINITI*, 6:2, 3-8.

Martyn, J. 1964. Bibliographic coupling. *Journal of documentation* 20:4, 236.

Mayr, Ernst. 1982. *The growth of biological thought: diversity, evolution, and inheritance.* Cambridge, Mass.: The Belknap Press of Harvard University Press.

Miksa, Francis. 1998. *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.

Mishler, Brent D. 2000. Deep phylogenetic relationships among "plants" and their implications for classification. *Taxon* 49:4, 661-683.

Moss, R. 1964. Categories and relations: origins of two classification theories. *American documentation*, 296-301.

Oleson, Alexandra & Voss, John, eds. 1979. *The organization of knowledge in modern America, 1860-1920*. Baltimore: Johns Hopkins University Press.

Ranganathan, Shiyali Ramamrita. 1951. *Philosophy of library classification*. Copenhagen: E. Munksgaard.

Rowley, Jennifer E. & Farrow, John. 2000. *Organizing knowledge: an introduction to managing access to information.* 3rd ed. Alderstot: Gower Publishing Company.

Schneider, Jesper W. 2004. *Verification of bibliometric methods' applicability for thesaurus construction*. Ph.D. dissertation, Royal School of Library and Information Science, Aalborg. Available at http://biblis.db.dk/archimages/199.pdf

Small, Henry. 1973. Co-citation in the scientific literature: a new measurement of the relationship between two documents. *Journal of the American Society of Information Science*, 24:4, 265-269.

Spang-Hanssen, Henning. 1974. Kunnskapsorganisasjon, informasjonsgjenfinning, automatisering og språk. In *Kunnskapsorganisasjon og informasjonsgjenfinning*. Oslo: Riksbibliotektjenesten, 11–61. Available at http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Spang%5FHanssen%5F1974.pdf

Spang-Hanssen, Henning. 2001 [written 1970]. How to teach about information as related to documentation. *Human IT* 1, 125-143. Available at http://www.hb.se/bhs/ith/1-01/hsh.htm

Sparck Jones, Karen 2005. Revisiting classification for retrieval. *Journal of documentation* 61:5, 598-601. [Reply to Hjørland & Nissen Pedersen, 2005]. Available http://www.db.dk/bh/Core%20Concepts%20in%20LIS/Sparck%20Jones%5Freply%20to%20Hjorland%20&%20Nissen.pdf

Warner, A. J. 1991. Quantitative and qualitative assessments of the impact of linguistic theory on information science. *Journal of the American Society for Information Science* 42:1, 64-71.

Warner, Julian 2002. Forms of labour in information systems. *Information research* 7:4. Available at http://informationr.net/ir/7-4/paper135.html

Warner, Julian 2007. Description and search labor for information retrieval. *Journal of the American Society of Information Science and Technology* 58:12, 1783–90.

Ørom, Anders 2003. Knowledge organization in the domain of art studies: History, transition and conceptual changes. *Knowledge organization* 30: 3/4, 128-43.