



Is classification necessary after Google?

Is classification
necessary after
Google?

Birger Hjørland

Royal School of Library and Information Science, Copenhagen, Denmark

299

Abstract

Purpose – The purpose of this paper is to examine challenges facing bibliographic classification at both the practical and theoretical levels. At the practical level, libraries are increasingly dispensing with classifying books. At the theoretical level, many researchers, managers, and users believe that the activity of “classification” is not worth the effort, as search engines can be improved without the heavy cost of providing metadata.

Design/methodology/approach – The basic issue in classification is seen as providing criteria for deciding whether A should be classified as X. Such decisions are considered to be dependent on the purpose and values inherent in the specific classification process. These decisions are not independent of theories and values in the document being classified, but are dependent on an interpretation of the discourses within those documents.

Findings – At the practical level, there is a need to provide high-quality control mechanisms. At the theoretical level, there is a need to establish the basis of each decision, and to change the philosophy of classification from being based on “standardisation” to being based on classifications tailored to different domains and purposes. Evidence-based practice provides an example of the importance of classifying documents according to research methods.

Originality/value – Solving both the practical (organisational) and the theoretical problems facing classification is necessary if the field is to survive both as a practice and as an academic subject within library and information science. This article presents strategies designed to tackle these challenges.

Keywords Classification, Epistemology, Library practice, Knowledge organisation

Paper type Research paper

Received 17 May 2011
Revised 5 September 2011
Accepted 6 September 2011

Introduction

Classification (together with indexing, document description, and metadata assignment) forms the basis of knowledge organization (KO)[1], both a practical activity and a main sub-discipline of library and information science (LIS), which focuses on improving this activity. Many consider KO to be the core of LIS, and KO has been institutionalized with professors, journals, conferences, educational programs etc. (cf. Pattuelli, 2010). The practice of classification and KO has been carried out in libraries for more than 100 years. Formally speaking, it has also been an academic subject in LIS programs since Melvil Dewey (1851-1931) established the first school of “library economy” in the USA in 1876. The outlook for the future is, however, fundamentally challenged by digital technologies at both the practical and at the theoretical levels. In hindsight, the question arises as to whether KO has ever had a sound theoretical basis because, as discussed below, questions such as: “How do we decide whether A is a kind of B?” have not been properly addressed in the field of KO.

At the practical level, we are facing the following challenge: the two largest Danish libraries, The Royal Library in Copenhagen and the State Library in Aarhus, have both almost entirely ceased using their own classification systems as well as classifying



their books themselves. The decision to do this may have been made based on the following considerations:

- Many libraries now rely mainly on the Dewey Decimal Classification (DDC) made by the Library of Congress (LC)[2] and disseminated in the MARC records[3] rather than making their own classification of each document.
- Many library directors expect that, in the future, large scanning projects (such as that which is being conducted by Google) may enable full text searches to be carried out of all available content. For this reason, they may consider it a waste of resources to classify or index books.
- Many libraries, including The Royal Library in Copenhagen, now also rely on user tagging and may perhaps expect that this will somehow act as a substitute for professional indexing and classification.
- Users mostly find the books they need using tools other than the library online public access catalog (OPAC)[4].

This means that we must distinguish between libraries functioning as tools for finding documents, as document-delivery services[5], and as reference (or other) services[6]. Both the finding function and the delivery function of libraries are seriously threatened by the so-called “library bypass”, which is the result of publishers’ digital libraries as well as possible tendencies towards open-access publishing. Here, however, we are concerned with KO, and therefore we will only consider the KO carried out by libraries and bibliographical databases from the perspective of the competing systems of KO available to users.

Our strategy regarding how to develop KO should be based on the premise that users today have access to the internet[7]. This means that every specific system or classified collection is facing the challenges which arise from competing systems and services. For a user who is interested in a specific issue, it does not matter whether or not the classification/KO is made by a local library. As an aid to finding information, the best[8] KO is only one click away (and there is no need for the second-best KO or for the hundreds of lesser KO systems used in parallel by libraries, publishers, or bibliographical databases). By implication, the theory and practice of classification and KO should aim to provide high-quality services on a global scale, for example, in subject-specific bibliographical databases, in WorldCat, and in related bibliographies and catalogs.

Given the significant reliance on centralized classification and the challenges posed by smart technologies, it follows that there is a need for control mechanisms to check the quality of the classifications made, for example, by LC. There is far too little research about the quality of indexing and classification today, and particular decisions are not documented and made available for research. If the centralized classifications are not of a very high quality, users may not find them to be viable, given the many alternatives. In his study, Larson (1991) indicated that many users did not find Library of Congress Subject Headings (LCSH) useful, and even found them to be harmful in the online context. He wrote:

Experience in catalog use may not necessarily imply that users have been “conditioned” to avoid subject searches [i.e. use of controlled vocabulary], although such conditioning appears to be a likely result of gaining experience in catalog use, whether card or online catalog. We

would suggest, as a hypothesis for further study, that individual users' experiences of subject search failure and information overload lead them to reduce their use of the subject index and to increase their use of alternate means of subject access, such as title keyword searching and shelf browsing following a known item search (Larson, 1991, p. 211).

A more recent study came to a similar conclusion:

These results point to a characteristic use in the University of Granada of a strong preference for searching by title (49%), followed by searches by author (37%), and finally, by subject search (14%). Our findings come to support the results of previous authors, highlighting the inherent difficulties and reduced interest surrounding searches by subject [i.e. controlled vocabulary] (Villen-Rueda *et al.*, 2007, p. 336).

These two studies may or may not reveal the true picture (for a different conclusion, see Gross and Taylor, 2005), and there is a desperate need for more studies. Another indication of the crisis of bibliographic classification is that the citation databases from Thomson Reuters only have a rough classification of the journals they index based on an "intuitive" rather than any kind of scholarly methods (cf. Leydesdorff, 2006, p. 602). That these very successful databases have not found it worth investing in the classification of their articles is also an indication of a crisis of classification as ordinarily understood. We shall not go further into the actual situation here; at this point, I simply want to point to a possible problem which is concerned with the principles and quality of subject analysis, to which we shall return below.

The challenge on the practical level can therefore be formulated this way: how can LIS professionals contribute to the findability of documents, given the availability of many competing services in the "information ecology"? The answer involves, among other things, a strongly coordinated organization of efforts and a strengthening of the connection between theory and practice in KO.

At the theoretical level, we are facing other challenges. We now have Google, for example, which students use far more than they use library catalogs in order to find what they need (De Rosa *et al.*, 2005, 2006; Pors, 2005). This triggers the question: can information retrieval (IR) theoretically be carried out perfectly without any kind of "classification"? The well-known computer scientist and information scientist Karen Sparck Jones (2005) argued that techniques such as "relevance feedback" would remove the need for classification as it is commonly understood. In previous years, the famous computer scientist and information scientist, Gerard Salton, has also argued:

Acting as if we were stuck in the nineteenth century with controlled vocabularies, thesaurus control, and all the attendant miseries, will surely not contribute to a proper understanding and appreciation of the modern information science field (Salton, 1996, p. 333).

Is classification – even at the theoretical level – still needed in the post-Google era? Or are computer algorithms able to do a 100% satisfactory job without the need for classification[9]?

This theoretical challenge constitutes a serious threat to the justification of classification, KO, and LIS as fields of both research and practice (cf. Bawden, 2007). If we believe that we in the field of knowledge organization are entitled to a place in the academic world as well as in the practice of KO, we have to be able to provide both academic and practical justification for classification and in other forms of KO. This article is an attempt to contribute to the development of such an argument.

Digression: some comments about the UDC

The UDC system is a classification system that was proudly developed by the documentation movement (which I consider to be a part of LIS). First of all, it reflected a high level of subject knowledge and corresponded well to the academic discourses in various fields. Second, it was originally designed not just for monographs, but also for articles and other kinds of documents. It was therefore much more detailed than, for instance, DDC. I say “*was* proudly developed” because, although it still exists and some people still contribute to its development, I do not think that we can take pride in it any longer (although I have great respect for the small group of researchers who are still working on it, such as Ian McIlwaine; see, for example, McIlwaine, 2010). There are various reasons for this decline. In a previous study (Hjørland, 2007a), I argued that our community has not been able to maintain and update this system properly, and I find it almost scandalous that the new edition has so many totally obsolete sections.

The UDC was once maintained by the International Federation for Information and Documentation (FID), which, after a period of crisis, was dissolved in 2002. Before that time, there were many committees (international as well as national, including a Danish committee) working together for the improvement of this system. Today, to my knowledge, nothing similar exists, and there are no large-scale groups, committees, or organizations working together to produce high-level classifications (or thesauri, ontologies, or related tools)[10]. I believe that this is becoming a problem in the digital environment because the development of quality classification systems or KO tools is much too big a job for a single library or for small groups of experts such as those who are maintaining the DDC in LC. If our community is to be able to produce something which would be able to make LIS more visible and improve access to information in the post-Google Era, this would probably need to be based on large-scale international cooperation involving LIS practitioners, LIS researchers, and (other kinds of) specialists, including subject specialists in all major fields. The practical work of maintaining and updating a given system (such as the UDC) is not, however, research in itself, but should be based on research. What could be recognized as research would be the publication of articles which carefully argue for the specific decisions made, considering empirical, logical, historical and pragmatic issues (e.g. why A should be considered as part of B in a specific context).

The UDC – as well as other traditional library classification systems – is also designed in such a way that makes it suitable for arranging books on shelves[11]. This implies constraints that are unnecessary in online searches, and as such the problems of shelving and retrieval functions should be considered separately. This may also form part of the explanation of why the UDC is no longer playing its former role.

It should also be said that systems such as the UDC may, in the past, have been based on the idea that classifications are neutral, objective, and content-independent decisions (e.g. that concept A is related to concept B in specific ways, regardless of domain or perspective). I believe that these assumptions are wrong and should be replaced by principles that are more in accordance with alternative views. This would, in my opinion, imply the need for specific KO tools aimed at different subjects and paradigms. The widespread philosophy that classification can be standardized and thus reused in different contexts seems problematic because different discourse communities develop their own terminology, meanings, and relevance criteria. Strong

arguments can therefore be put forward in accordance with the view that classification should be tailored towards different domains, epistemic communities, and user groups.

The task at hand is therefore not just to update systems such as the UDC or to create a set of thesauri spanning all disciplines or other kinds of knowledge organizing systems (KOS; of which the most advanced[12] are ontologies; cf., Hodge, 2000). Instead, the task is to provide an overall framework for developing and discussing different alternatives for ways in which users will be enabled to make informed decisions during IR, along with a broad-based development of specific tools. However, large international projects aimed at classifying the world's knowledge based on literary warrant (i.e. through an examination of concepts in the literature) still seem to be important and are probably necessary for the survival of KO. The main difference between the former understanding of the UDC and an approach based on today's outlook would be that the new approach should be less prescriptive. Instead, it should be more descriptive and contextual (and based on, among other things, bibliometric studies, historical dictionaries, and other forms of domain analysis).

What is classification?

We could say that classification is the interdependent processes of:

- defining classes;
- determining relationships between classes (such as hierarchical relations, among others), i.e. making a classification system; and
- assigning elements (in LIS, documents) to a class in a given classification system.

This is equivalent to the interdependent processes of:

- defining concepts (see Hjørland, 2009);
- determining semantic relations between concepts (see Hjørland, 2007b); and
- determining which elements fall under a given concept (to assign a given “thing” to a concept).

Example

To say that the concept of “the Müller-Lyer illusion” is a kind of “optical illusion”, which is a kind of “perceptual phenomenon”, which is a kind of “psychological phenomenon”, is the same as classifying “the Müller-Lyer illusion” within a class termed “optical illusions”, which is part of the broader class of “perceptual phenomena”, which is part of the broader class of “psychological phenomena”.

A recent dissertation wrote:

Ingetraut Dahlberg, another influential classificationist, states that “the elements” of classification schemes are “concepts or representations of concepts” (Dahlberg, 1978, p. 9). One could thereby conclude that when a concept is chosen for a class, the concept in question refers to something which should be shared by all documents of that class, and that this concept is treated by the documents. However, taking the label “011” in the DDC as an example, it refers to a class of documents that has the common feature that they are bibliographies and not about bibliographies (Gunnarsson, 2011, p. 16).

What Gunnarsson illustrates here is that there is an important difference in whether a thing or a document (A) is *a kind of* X or is *about* X. His quote shows that the criteria

for assigning A to X depend on whether we are speaking of form classes or subject classes, although both kinds of classes are defined by a concept (e.g. “bibliography”). He does not, however, discuss the criteria used to decide whether or not A should be assigned to X (which corresponds to define the meaning of a concept, for example “bibliography”).

In the context of librarianship, “classification” is often used for systems such as the DDC, the UDC, the LCC, Colon, or Bliss, and its process involves assigning individual books or other documents to one or more classes within such a system, as opposed to verbal indexing systems (see Figure 1).

The fundamental distinction in Figure 1 is between classification systems and verbal indexing systems, which I consider to be “traditional” in KO, but which has been criticized by Lancaster (2003, pp. 20-2), who argued that one should not speak of assigning classification codes as “classification” in opposition to the assignment of indexing terms as “indexing”. “These terminological distinctions”, he writes, “are quite meaningless and only serve to cause confusion” (Lancaster, 2003, p. 21). The view that this distinction is purely superficial is also supported by the fact that a classification system may be transformed into a thesaurus and vice versa (cf. Aitchison, 1986, 2004; Broughton, 2008; Riesthuis and Bliedung, 1991).

Mikael Gunnarsson further writes:

Indexing in particular, but classification as well, stresses *the assignment of labels to documents rather than the assignment of documents to classes of documents*” Gunnarsson (2011, p. 85; emphasis in original).

However, I would dispute this distinction because the act of labeling a document (say by assigning a term from a controlled vocabulary to a document) is at the same time to assign that document to the class of documents indexed by that term (all documents indexed or classified as X belong to the same class of documents).

It is therefore important to realize that all of the different kinds of systems in Figure 1 (with the possible exception of free text systems) are different kinds of “classification systems”. The most important difference between the different indexing languages shown in Figure 1 is between “free text systems” and the others (Figure 2).

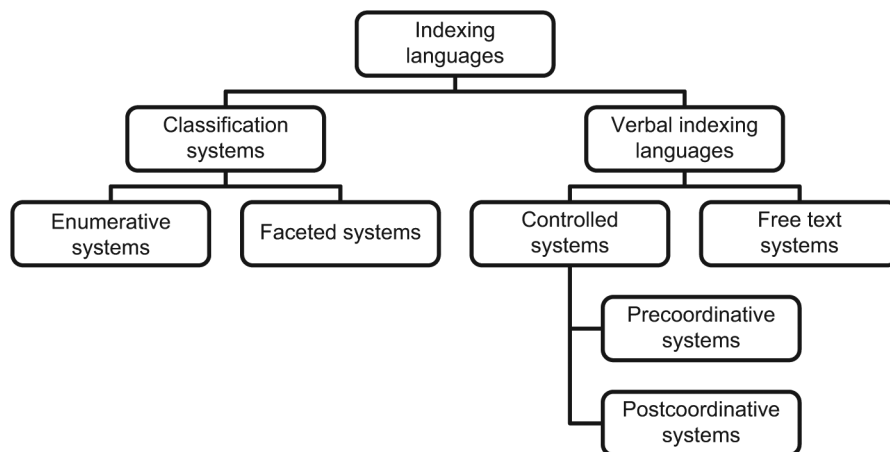


Figure 1.
The traditional view of the different kinds of indexing languages

In free text systems, none of the classification is carried out by personnel associated with the bibliographical system, as there are only possible “classifications” made by the authors of the documents. These are represented in the free text systems in the form of relations between meanings in a given text (for example, when it is said that A is a kind of B), which are developed in the discourse communities to which a given text belongs.

In all other kinds of systems, there are classifications made by LIS personnel (or others responsible for developing, maintaining, and using the systems as “metadata” in bibliographical systems). All such systems may be termed “controlled vocabularies,” and they are “normative” and “prescriptive”. They may also be termed “bibliographic classifications” (classification *of* documents about, for example, plants, animals, chemicals, religions, languages, genres, and sports). As such, they are opposed to “scientific and scholarly classifications” (e.g. classifications of plants, animals, chemicals, religions, languages, genres, and sports) developed in different disciplines and reflected in the documents being indexed. Bibliographical classification is not independent of scientific classification; rather it must, to a large extent, be based on and reflect scientific/scholarly classifications. Therefore, LIS cannot develop adequate theories of classification if problems of scientific classification are ignored.

A more recent overview of knowledge organization systems have been presented by Hodge (2000), who grouped them into three general categories:

- (1) term lists, which emphasize lists of terms, often with definitions;
- (2) classifications and categories, which emphasize the creation of subject sets; and
- (3) relationship lists, which emphasize the connections between terms and concepts.

What Hodge has actually provided, though recognizing that this list is not comprehensive, is a suggestion for a taxonomy of KOS:

Term lists

- Authority files
- Glossaries
- Dictionaries
- Gazetteers

Classifications and categories

- Subject headings
- Classification schemes
- Taxonomies

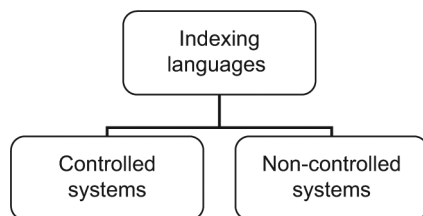


Figure 2.
The two basic kinds of
indexing
languages/information
languages

- Categorization schemes
- Relationship lists
 - Thesauri
 - Semantic networks
 - Ontologies

In this paper, we shall not consider each kind of KOS or analyze Hodge's taxonomy. The most important difference between the different kinds of KOS seems to be the different kinds of semantic relations being displayed. In traditional classification systems, hierarchical relations and the relationship between synonyms and homonyms are the most important. In ontologies, a large range of semantic relations are possible; it is a question of whether other kinds of KOS are needed for KO theory or whether other kinds of KOS should be considered special cases of ontologies with more limited ranges of semantic relations. Dagobert Soergel wrote:

Unification: Ontologies

The relationships between document components in a document model, the tags in a document template or a metadata schema, the table structure in a relational database (or the object structures in an object-oriented database), and the relationships between concepts can all be traced back to (or defined in terms of) an entity-relationship model [. . .] Such a model is an ontology, so all structures in a digital library can (and should) be conceived as subsets of an overarching ontology (Soergel, 2009, p. 38).

Or, in the words of Lars Merius Garshol:

With ontologies the creator of the subject description language is allowed to define the language at will [. . .] [A s]ummary of the relationship between topic maps and traditional classification schemes might be that topic maps [i.e. ontologies[13]] are not so much an extension of the traditional schemes as on a higher level. That is, thesauri extend taxonomies, by adding more built-in relationships and properties. Topic maps do not add to a fixed vocabulary, but provide a more flexible model with an open vocabulary.

A consequence of this is that topic maps can actually represent taxonomies, thesauri, faceted classification, synonym rings, and authority files, simply by using the fixed vocabularies of these classifications as a topic map vocabulary (Garshol, 2004).

Both Hodge's taxonomy of KOS and the taxonomy of indexing languages in Figure 1 emphasize form rather than content. It has already been said that, for example, classifications may be converted into thesauri. This is an indication that the form or formal aspects of KOS and indexing languages are less important. The central issue in classification is, as previously stated, the way in which the semantic relations between concepts are determined (e.g. how we decide whether or not A is a kind of B).

It should also be noted that a given KOS is much more than its systemic features. When, for example, a classification system has been applied for years in a given library many qualified decisions may have been made about what to classify as X. In this way the system may then represent an important accumulation of knowledge, which is seldom reflected in the scientific literature concerning KO (but may exist in the form of minutes from meetings and as "tacit knowledge" among the indexers).

The Cranfield experiments, which were founded in the 1950s, introduced the famous measures of "recall" and "precision" as evaluation criteria for system efficiency. They found that classification systems such as the UDC and facet-analytic systems were less

efficient compared to free text searches or low-level indexing systems (“UNITERM”). According to Ellis (1996, pp. 3-6), the Cranfield I test revealed the following results:

- UNITERM: 82.0 percent recall;
- alphabetical subject headings: 81.5 percent recall;
- UDC: 75.6 percent recall; and
- facet classification scheme: 73.8 percent recall.

Is classification
necessary after
Google?

307

Although these results have been criticized and questioned, the IR tradition became much more influential whilst library classification research relatively lost its influence. Since then, a split has appeared between IR researchers (who tend to assume that controlled vocabularies are not important) and KO researchers (who continue to assume they are still important).

With regard to the construction of controlled vocabularies, Maniez (1997) found that “compatibility is the paradise lost of information scientists, the dream of a universal communication between information languages [indexing languages]. Paradoxically, the information languages increase the difficulties of cooperation between the different information databases”. Maniez’s insight is an argument against the need for classification as it is traditionally performed; it seems to suggest that each controlled system tends to make its own arbitrary decisions (rather than reflecting an underlying order), and thereby tends to make a federated search or integrated search less efficient.

To sum up, to classify is to define the “kind” which a given “thing” is, and how that kind is related to other kinds. This is a fundamental process that all human beings carry out many times each day (e.g. when sorting which things are “food” and which are not, which fruit is fresh and which is rotten, which desserts a child likes and which s/he dislikes, etc.). All the different kinds of “indexing languages” and KOS are based on the same sort of fundamental decisions.

Criteria for classifying

The underlying theoretical questions are: “How do we decide the criteria for assigning document A to class X?” and the related question: “what are the criteria by which decisions such as assigning the semantic relation X between the concepts A and B can be made?”.

Concerning the first problem (how do we decide whether we should classify document A as belonging to class X) there are different principles in play. In the mainstream library tradition, it has been a rule that at least 20 percent of a book should be about X in order for it to be assigned to that class; in automatic indexing, on the other hand, it is the occurrences of specific terms in A that determines whether or not it is assigned to X. In request-oriented indexing it is the anticipated request from users that determines when to assign A to X (the indexer asks himself: “Under which descriptors should this entity be found?” and “Think of all the possible queries and decide for which ones the entity at hand is relevant” (Soergel, 1985, p. 230). Finally, two views in accordance with request-oriented indexing have been formulated by Rowley and Farrow and by Hjørland, respectively. The core of indexing as stated by Rowley and Farrow is to evaluate a paper’s contribution to knowledge and index it accordingly (Rowley and Farrow, 2000, p. 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, as in the words of Hjørland (1992, 1997): “the subjects of a document are its informative potentials”. While mainstream classification research is still based on the objectivist understanding (a document has a subject), the minority view (that document A is assigned subject X by somebody in order to support some specific activities) is gaining a footing. I believe this last view is decisive for making a future for classification in both theory and practice.

Concerning the second question (how to assign semantic relations between concepts, for example how to decide that A is a kind of B) LIS personnel classifying or assign terms from different kinds of controlled vocabularies may often assume that such decisions are based on logic or on fixed semantic relationships between concepts. Svenonius (2000, p. 131), for example, found hierarchical relations to be “wholly paradigmatic or a priori [*sic*]”. However, in my opinion, this is not a fruitful perspective. As the following quotation expresses, a controlled vocabulary should be understood as an interpretation:

A controlled vocabulary is a way to insert an interpretive layer of semantics between the term entered by the user and the underlying database to better represent the original intention of the terms of the user (Fast *et al.*, 2002).

Such an interpretation is often an empirical question, and often different perspectives and interest have different criteria for deciding semantic relations. In chemistry, for example, helium is considered a noble gas; however, in Stowe’s Periodic Table (based on quantum mechanics) helium appears with the alkaline earth metals (cf. Channon, 2011). In contemporary science the place of helium has not been solved, and it is an open question whether or not there is one correct answer (cf. Scerri, 2007). Such considerations make it important not just to consider classification to be about logical decisions, but to be based on interpretations of discourses and on the negotiation of different interests.

Although researchers such as Andersen (2004), Cornelius (1996), Feinberg (2008), Frohman (1983, 1990), Mai (2000, 2011), Olson (2001, 2002) and Ørom (2003) have promoted interpretative views of classification and have formed part of an important “social turn” in classification research, this view is still not very influential and has not resulted in the formation of a coherent theoretical approach. There are, for example, no textbooks on indexing and classification written from such a social and interpretative perspective in which it is explained how the relation X between concepts A and B should be determined and assigned. Theoretically speaking, the quotation from Fast *et al.* (2002) raises an important question regarding the basis on which LIS personnel construe and apply this “interpretative layer”.

Olson (2001) examined the way in which LCSH were used to index some books in the field of gender studies, and found the indexing to be problematic. For example, the concept of “voice” (in relation to the views of a minority) is not represented in LCSH. According to Olson’s interpretation, this may be due to poor indexing of marginalized topics (as opposed to mainstream topics). My interpretation is unfortunately more pessimistic: I believe that this poor indexing is due to a lack of subject knowledge. The terms and relations in LCSH in relation to this field (gender studies) seem to be speculative and far away from the points of view that the literature in the field tries to put forward. This probably does not only apply to the views of minorities, which seem to be poorly represented by LC, but constitutes a general tendency that will also affect majority views. I agree with Olson that it is problematic that LC fails to see that the

books examined by Olson are about “voice” (this is particularly so because, in my view, one of the most important functions of libraries – and of KO – is to help different “voices” be heard). However, I have no statistical basis for claiming that LC indexing is of a poor quality; I simply wish to state that this is an important issue for investigation and that we need more qualitative, interpretative studies like Olson’s in order to better understand and improve indexing and retrieval.

More generally, what principles form the basis for the concepts classified in a given KOS? I consider subject headings, classification codes, etc. to be, in principle, the equivalent of chapter titles and section headings in a book about the history of the domain to be classified. Consider, for example, histories of Danish fiction. During the twentieth century, several important histories of Danish fiction were written, first from more traditional viewpoints, and then some reflecting particular views such as feminism, Marxism, postmodernism, etc. Each work reflects the subjectivity of both its author and its time (*Zeitgeist*). In spite of this subjectivity, such histories also reflect different levels of quality (they are reviewed and evaluated, and some are considered to be brilliant while others are considered to be bad). There is no way to avoid such subjectivism, and it would probably be a bad idea to try to do so. A given history may attach new labels to a given set of books, for example “magic realism”. Such labels are not necessarily derived from the literature, but are often assigned to it on the basis of a new conceptualization or interpretation. In the same way, libraries and the LC should assign labels such as “voice” to certain books in order to support important goals associated with the discourses of which the books are a part. By implication, it is important for LIS to study how concepts of genre and other methods of conceptualizing documents are used, and for researchers to incorporate them into their own conceptualizations[14]. This may seem extremely costly; however it should be considered that this kind of knowledge is not only necessary in relation to indexing, but in relation to any kind of qualified work concerning information. Furthermore, this kind of knowledge is assumed in high standard libraries and bibliographical databases such as the National Library of Medicine and the MEDLINE database. Lowering the cost of employing qualified staff may ultimately mean that KO and classification will disappear because there will be no need for classifications other than the best[15] ones.

In the past, it has often been assumed that to say of what kind a given “thing” is and how that kind is related to other kinds is a task which has one true solution, i.e. that things have an “essence” and that classification reflects these “essential” properties. Plato and Aristotle, for example, are generally considered to be essentialists. The idea of essentialism is related to the concept of natural kinds and may also be expressed by Plato’s metaphor of “carving nature at its joints”. Essentialism has come under fierce criticism, however, as stated by Rachel Cooper:

In recent years, traditional essentialist accounts of natural kinds have come in for fierce criticism. A major difficulty is that for biological species, which are traditionally considered amongst the best examples of natural kinds, no plausible candidates for the essences can be found. Several different criteria may be employed by biologists seeking to determine species: morphological features, evolutionary lineages, the criteria of reproductive isolation, or genetic features. On examination none of these appears suitable candidates for being the essential properties of biological species (Cooper, 2005, p. 47).

This critique claims that classifications are made in order to facilitate certain human tasks, and that the properties chosen for classification depend on the purpose of the

classification, as different purposes require different classifications, and that classifications are theory-dependent (or indeed that a classification is a theory of sorts). In a paper published in *Knowledge Organization* (Hjørland, 2011b), I suggested that the ideal form of the periodic system of chemistry and physics depends on whether quantum mechanics (QM) or more traditional chemical views are emphasized[16]. The question of essentialism may, however, still be considered to be an open question.

To conclude this section, classification involves considering and negotiating different theories and interests within the domain that is being classified[17].

How is library classification related to other forms of classification?

In Hjørland (2008), I provided the following model for “the traditional view of classification” in KO (see Figure 3).

This view may be expressed by stating that there is only one way in which nature has joints, or as Stamos wrote: “Nature itself has supplied the causal monistic essentialism. Scientists in their turn have simply discovered and followed (where ‘simply’ ≠ ‘easily’)” (Stamos, 2004, pp. 138-9). Library and information scientists, in turn, must study scientific classifications and “simply discover and follow” scientific classifications (where again “simply” ≠ “easily”). When LIS professionals classify a given book, the concepts which they use are derived from the literature, and are not primarily constructed by LIS professionals. As Hulme (1911, pp. 46-7) has stated: “The real classifier of literature is the book-wright, the so-called book classifier is merely the recorder”. This view, however, almost disappeared in KO in the second half of the twentieth century (to be ousted by, for example, facet-analytic and user-oriented perspectives)[18].

In opposition to this traditional view it may be suggested that classifications reflect the purposes for which they are designed and that different sciences, theories, and human activities classify the world (more or less) differently[19]. Both the practice of science and the practice of information science can therefore be seen as more constructive. The periodic system of physics and chemistry seems to be the ultimate challenge to this view.

For LIS and KO, the implication is that, in both cases, classification is not constructed within our field but is dependent on subject knowledge produced outside LIS. The pragmatic view emphasizes that LIS/KO should consider the purpose of its classification and the activities that classification is made to support.

Why classification is needed: the case of evidence-based practice

If classifications are to be relevant, they must consider and negotiate between different views and interests. In order to support these given views, classifications must enable

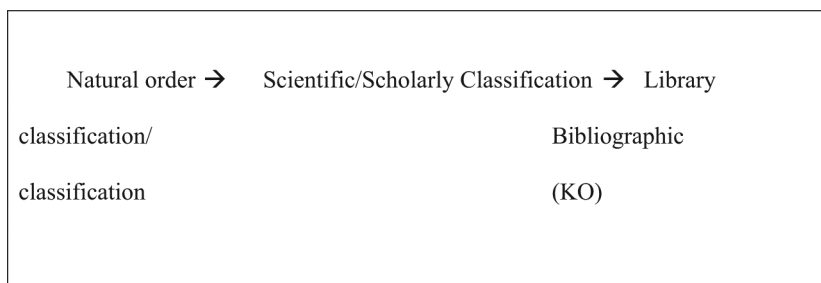


Figure 3.
Model for the traditional
view of classification in KO

IR according to the relevance criteria associated with them. To classify should be to make *relevant* distinctions in relation to the goals of the system, and therefore implies a consideration and negotiation of different views and interests. Google and other similar IR systems are certainly impressive, but how do they classify and prioritize the relevant information? We tend to think of such systems as neutral and objective tools, but they cannot be. Any system is always biased in some way or another (see, for example, Fortunato *et al.*, 2005; Gerhart, 2004; Introna and Nissenbaum, 2000). Search engines may be calibrated in order to provide different findings or rankings. In order to make such a calibration (or simply to evaluate the systems), we need to have some kind of classification of what should be found. Thus far, in the field of LIS, we have mainly used assessments of relevance based on “user relevance” (see Hjørland, 2010). However, if we are to trust, for instance, retrieved medical documents, it would be better to base our relevance judgments on scientific criteria (such as research methods), rather than on the opinions of users. This is done most explicitly in the interdisciplinary movement known as evidence-based practice (EBP), which developed from evidence-based medicine. In this movement, documents are classified according to very explicit criteria (based on a hierarchy of research methods[20], i.e. criteria for what counts as evidence). EBP is presented and discussed in Hjørland (2011a), in which it is concluded that the focus on scientific argumentation in EBP constitutes an important and long overdue contribution from EBP to LIS. However, parts of the underlying epistemological assumptions should be replaced: EBP is too narrow, too formalist, and too mechanical an approach on which to base scientific and scholarly documentation. Whether or not we accept the philosophy of EBP wholeheartedly, or whether we accept the epistemological criticism raised against it, classification appears to be a necessary activity. If a given hierarchy of research methods is accepted as generally the best, papers need to be classified according to the research methods listed in this hierarchy. If such a hierarchy is formed, it might be possible to construe algorithms that – to some degree of certainty – can classify the documents automatically (although the classification itself has to be constructed beforehand, leading to so-called “supervised machine learning” or “supervised classification”). If the critique put forward by me (Hjørland, 2011a) is accepted, this means that the classification of research methods is less formalist and mechanical and that classification therefore becomes more interpretative and presupposes more contextual knowledge. The point to note is that medical researchers, for example, cannot rely on IR systems that are not based on metadata and do not reflect the scientific criteria in a given domain. I am not arguing that EBP is the final answer, but I believe that it is a much healthier approach on which to base systems of IR and classification than the paradigms which have dominated LIS and KO for the last 30 years. For the moment, suffice it to say that EBP offers a rationale from which we can reach the important conclusion: we cannot manage without classification.

Is classification
necessary after
Google?

Conclusion: why classification is necessary after Google

We have now considered how classification is challenged in library practice as well as in information retrieval theory. We have considered different forms of classification including those done by search engines or algorithms. We have pointed out that at the core is classification theory regarding whether A should be considered a kind of X and how concepts are related. All kinds of knowledge organization systems may be seen as

kinds of ontologies, which consist of selections of concepts and their selected semantic relations from a specific conceptualization. Ontologies are not just neutral reflections of an objective reality, but are constructed from a world-view that is fruitful for some purposes and values, though at the expense of others. In developing this issue we saw that such decisions involve consideration of subject-specific theories as well as metatheories and “paradigms”. All this is a major challenge to traditional theories of classification in our field. I believe my stated principles are necessary, but will encourage information specialists to challenge them so that some consensus may develop and we can work together for the future of KO.

Given this new theoretical understanding of classification and KO we can understand why classification is necessary in any kind of library, documentation, or information work: the criteria of classification are simply identical to criteria of relevant information provision. It is my claim that an information specialist dealing with questions in Arts would be better situated to do so if he or she has read and understood Ørom (2003) about major paradigms in that domain and how they have affected library classification systems. We should work in this direction for all domains.

Notes

1. There is a tendency to change the term from knowledge organization to information organization (IO) or information architecture (IA). The term IO is now used, for example, at the School of Information Studies at the University of Wisconsin-Milwaukee. In this article, the more traditional term “KO” is used.
2. The DDC is published by OCLC Online Computer Library Center, Inc. It is developed and maintained in the Library of Congress, but most of the editorial staff are employed by OCLC. “The Dewey editorial office is located in the Decimal Classification Division of the Library of Congress, where classification specialists annually assign over 110,000 DDC numbers to records for works cataloged by the Library. Having the editorial office within the Decimal Classification Division enables the editors to detect trends in the literature that must be incorporated into the Classification” (cited from the Dewey Editorial Office; see <http://staff.oclc.org/~dewey/dewey.htm>).
3. According to the website of the Royal Library: “At present most of the books in English will be classified in the DDC code, whereas books in Danish are classified in the DK5” (see www.kb.dk/en/kub/fag/hum/tors/kulturstudier/boeger.html).
4. For my own part, for example, I often find the books that I need for my research in reference lists, at Amazon.com or the Web of Science, or in other places (and some of the books that I find in these ways I subsequently borrow from my local library). I also find more relevant books displayed as “new books” at the library of RSLIS, compared to what I find in the classified catalogue.
5. One example of a combined function is the role of classification systems as browsing systems on open shelves with physical books and materials. This function, however, will in future be dependent on the display of physical documents rather than electronic documents and is therefore also threatened by digital media. It is more likely that the future will bring an increase in special exhibitions and that browsing whole libraries will be possible electronically in the form of “virtual libraries”, which can be displayed by multidimensional organizations.
6. It is important to realize that the kind of knowledge which is needed in order to classify knowledge is equally relevant when providing guidance to users, searching for information, and other related tasks. Shachaf (2009) indicates that the Wikipedia reference desk, which

relies on volunteers, is as good as a library staffed by professionals. This is a provocative suggestion that needs to be considered carefully. My suggestion is that the kind of knowledge that is needed for answering questions overlaps a great deal with the kind of knowledge that is needed in order to classify documents.

Is classification
necessary after
Google?

313

7. There may be a need for services to help users without access to the internet. Such services cannot, in my opinion, justify the maintenance of a research field such as KO.
8. The “best” KO means the best for a given task. Although it is argued in the article that no system is the best for all purposes, a system may be the best one in a specific search situation.
9. This question may be reformulated as: what are the relative contributions that different kinds of subject-access contribute to successful retrieval? (cf. Hjørland and Kyllèsbech Nielsen, 2001).
10. Large-scale international development of broad ontologies might take place in some subject fields (e.g. GeneOntology), but in that case this seems not to be something that the community of KO is engaged in and which contributes to make our field visible.
11. The UDC was originally designed for a universal bibliography at the International Office of Bibliography, created in 1895 in Belgium and based on a card catalog system. It was therefore not designed for shelving purposes, but it has mostly been used in large research libraries in which it has also served the needs of shelving.
12. Most advanced in the sense of allowing many kinds of semantic relations to be expressed.
13. A topic map is a standard for representing knowledge based on an ontology.
14. See Abrahamsen (2003) for an example of how a genre label originated in popular music.
15. “Best ones” means best for needed tasks. Although no KO is optimal for all purposes it cannot be inferred that any system is optimal for some purposes: Many systems may in reality be considered superfluous.
16. Timothy Stowe’s periodic table for physicists from 1988 may best represent QM, while other versions may better reflect other points of view. Scerri (2007, p. 286) suggests “the left-step table” as the best solution, but writes: “I am well aware of the resistance that this proposal will meet, especially from the chemical community”. This suggests that different theories and interests are at play even in this, the hardest of the natural sciences. However, of course, one theory at any given point in time may be able to combine different views and thus appear to be the only theory, and by implication, provide one true classification of the elements – at least for a period of time.
17. The anthropologist Jan Ovesen, during his term of office at the Royal School of Library and Information Science in Copenhagen, criticized the way in which anthropology was classified according to the Danish Decimal Classification: “In the academic world, both in Denmark and internationally, cultural and social anthropology (in Denmark still termed by its previous name, i.e. ethnography) is an independent science with its own institutes at universities, its own terminology and scientific historical development, and a relatively well delimited subject literature. [Note omitted] This is, however, not the case in DK5! [The Danish Decimal Classification System, 5th ed., which is a Danish modification of the Dewey system]. In this system it is more important, apparently, to draw meaningless distinctions between, for example, European and non-European cultures, between ‘developed cultures’ and ‘primitive people’, between the life of ordinary people and cultural processes, between social anthropology and ethnography etc. These distinctions obviously come from the strange understanding that ethnography should only deal with the pure *description* of *primitive* people, which are *not* in a *process of cultural change* caused by the meeting with the Western world. From the point of view of the discipline such delimitation is totally absurd.

Firstly, there is today general consensus that 'pure descriptions' do not exist in ethnography or in other humanistic disciplines. Descriptions are not independent of the cultural background, the education, the personality etc. of the ethnographer. All description involves at the same time one or another kind of interpretation or even analysis. Secondly, decades of discussion about the concept of primitive have resulted in the view that it seems no longer meaningful to use the phrase 'primitive people'. Thirdly, today few if any groups of people exist who are not culturally changed in one way or another caused by contact with Western industrialized societies. One thus faces the paradoxical situation that the discipline of anthropology, which over the decades has experienced an important growth and development, in DK5 is represented by a class (59.5) in which the increase of literature must by necessity be very limited. The result has been that the central works of anthropology are not kept together, but are scattered in a number of rather different classes. Besides class 59.5 the classes 29, 30.12, 39 and 98 are the most important ones" (Ovesen, 1989, pp. 120-2; translated by BH, emphasis in original).

18. My own position is therefore closer to this traditional view than, for example, the facet-analytic, user-oriented or cognitive views.
19. A main spokesman of this view is Dupré (e.g. 1993).
20. "I a Evidence from a meta-analysis of RCTs [randomized controlled trials]; I b Evidence from at least one RCT; II a Evidence from at least one controlled study without randomisation; II b Evidence from at least one other type of quasi-experimental study; III Evidence from non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies; IV Evidence from expert committee reports, or opinions and/or clinical experience of respected authorities" (Geddes and Harrison, 1997).

References

- Abrahamsen, K.T. (2003), "Indexing of musical genres. An epistemological perspective", *Knowledge Organization*, Vol. 30 Nos 3/4, pp. 144-69.
- Aitchison, J. (1986), "A classification as a source for thesaurus: the bibliographic classification of H.E. Bliss as a source of thesaurus terms and structure", *Journal of Documentation*, Vol. 42 No. 3, pp. 160-81.
- Aitchison, J. (2004), "Thesauri from BC2: problems and possibilities revealed in an experimental thesaurus derived from the Bliss music schedule", *Bliss Classification Bulletin*, Vol. 46, pp. 20-6.
- Andersen, J. (2004), "Analyzing the role of knowledge organization in scholarly communication: an inquiry into the intellectual foundation of knowledge organization", PhD dissertation, Department of Information Studies, Royal School of Library and Information Science, Copenhagen, available at: <http://biblis.db.dk/Archimages/78.PDF> (accessed July 20, 2010).
- Bawden, D. (2007), "Editorial: The doomsday of documentation?", *Journal of Documentation*, Vol. 63 No. 2, pp. 173-4.
- Broughton, V. (2008), "A faceted classification as the basis of a faceted terminology: conversion of a classified structure to thesaurus format in the Bliss bibliographic classification, 2nd edition", *Axiomathes*, Vol. 18 No. 2, pp. 193-210.
- Channon, M. (2011), "The Stowe table as the definitive periodic system", *Knowledge Organization*, Vol. 38 No. 4, pp. 321-7.
- Cooper, R. (2005), *Classifying Madness: A Philosophical Examination of the Diagnostic and Statistical Manual of Mental Disorders*, Springer, Berlin.
- Cornelius, I.V. (1996), *Meaning and Method in Information Studies*, Ablex, Norwood, NJ.

- Dahlberg, I. (1978), *Ontical Structures and Universal Classification*, Sarada Ranganathan Endowment for Library Science, Bangalore.
- De Rosa, C., Cantrell, J., Hawk, J. and Wilson, A. (2006), "College students' perceptions of libraries and information resources. A report to the OCLC membership. A companion piece to 'Perceptions of libraries and information resources'", OCLC Online Computer Library Center, Dublin, OH, available at: www.oclc.org/reports/pdfs/studentperceptions.pdf (accessed October 10, 2009).
- De Rosa, C., Cantrell, J., Cellentani, D., Hawk, J., Jenkins, L. and Wilson, A. (2005), "Perceptions of libraries and information resources. A report to the OCLC Membership", OCLC Online Computer Library Center, Dublin, OH, available at: www.oclc.org/reports/pdfs/Percept_all.pdf (accessed October 10, 2009).
- Dupré, J. (1993), *The Disorder of Things: Metaphysical Foundations for the Disunity of Science*, Harvard University Press, Cambridge, MA.
- Ellis, D. (1996), *Progress and Problems in Information Retrieval*, Library Association Publishing, London.
- Fast, K., Leise, F. and Steckel, M. (2002), "What is a controlled vocabulary?", available at: http://web.archive.org/web/20030811115443/http://www.bboxesandarrows.com/archives/what_is_a_controlled_vocabulary.php (accessed December 7, 2010).
- Feinberg, M. (2008), "Hidden bias to responsible bias: an approach to information systems based on Haraway's situated knowledges", *Information Research: An International Electronic Journal*, Vol. 13 No. 2, article 07, available at: <http://informationr.net/ir/12-4/colis/colis07.html> (accessed March 24, 2011).
- Fortunato, S., Flammini, A., Menczer, F. and Vespignani, A. (2005), "The egalitarian effect of search engines", arXiv:cs.CY/0511005 v1, available at: <http://arxiv.org/abs/cs/0511005> (accessed July 20, 2010).
- Frohmann, B. (1983), "An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG", *Cataloging & Classification Quarterly*, Vol. 4 No. 1, pp. 11-27.
- Frohmann, B. (1990), "Rules of indexing: a critique of mentalism in information retrieval theory", *Journal of Documentation*, Vol. 46 No. 2, pp. 81-101.
- Garshol, L.M. (2004), "Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all", *Journal of Information Science*, Vol. 30 No. 4, pp. 378-91, available at: www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html (accessed July 20, 2010).
- Geddes, J.R. and Harrison, P.J. (1997), "Evidence-based psychiatry: closing the gap between research and practice", *British Journal of Psychiatry*, Vol. 171 No. 9, pp. 220-5.
- Gerhart, S. (2004), "Do web search engines suppress controversy?", *First Monday*, Vol. 9 No. 1.
- Gross, T. and Taylor, A.G. (2005), "What have we got to lose? The effect of controlled vocabulary on keyword searching results", *College & Research Libraries*, Vol. 66 No. 3, pp. 212-30.
- Gunnarsson, M. (2011), "Classification along genre dimensions: exploring a multidisciplinary problem", dissertation, Valfrid, Borås, available at: <http://bada.hb.se/handle/2320/7920> (accessed May 11, 2011).
- Hjørland, B. (1992), "The concept of 'subject' in information science", *Journal of Documentation*, Vol. 48 No. 2, pp. 172-200.
- Hjørland, B. (1997), *Information Seeking and Subject Representation. An Activity-Theoretical Approach to Information Science*, Greenwood Press, Westport, CT.

- Hjørland, B. (2007a), "Arguments for 'the bibliographical paradigm'. Some thoughts inspired by the new English edition of the UDC", *Information Research*, Vol. 12 No. 4, paper colis06, available at: <http://informationr.net/ir/12-4/colis/colis06.html> (accessed November 13, 2010).
- Hjørland, B. (2007b), "Semantics and knowledge organization", *Annual Review of Information Science and Technology*, Vol. 41, pp. 367-405.
- Hjørland, B. (2008), "What is knowledge organization (KO)?", *Knowledge Organization*, Vol. 35 Nos 2/3, pp. 86-101.
- Hjørland, B. (2009), "Concept theory", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 8, pp. 1519-36.
- Hjørland, B. (2010), "The foundation of the concept of 'relevance'", *Journal of the American Society for Information Science and Technology*, Vol. 61 No. 2, pp. 217-37.
- Hjørland, B. (2011a), "Evidence based practice: an analysis based on the philosophy of science", *Journal of the American Society for Information Science and Technology*, Vol. 62 No. 7, pp. 1301-10.
- Hjørland, B. (2011b), "The periodic table and the philosophy of classification", *Knowledge Organization*, Vol. 38 No. 1, pp. 9-21.
- Hjørland, B. and Kylesbech Nielsen, L. (2001), "Subject access points in electronic retrieval", *Annual Review of Information Science and Technology*, Vol. 35, pp. 3-51.
- Hodge, G. (2000), *Systems of Knowledge Organization for Digital Libraries. Beyond Traditional Authority Files*, Council on Library and Information Resources, Washington, DC, available at: www.clir.org/pubs/reports/pub91/contents.html (accessed May 16, 2011).
- Hulme, E.W. (1911), "Principles of book classification", *Library Association Record*, Vol. 13, pp. 354-358, 389-94, 444-9.
- Introna, L. and Nissenbaum, H. (2000), "Shaping the web: why the politics of search engines matters", *The Information Society*, Vol. 16 No. 3, pp. 1-17, available at: www.nyu.edu/projects/nissenbaum/papers/searchengines.pdf (accessed June 7, 2007).
- Lancaster, F.W. (2003), *Indexing and Abstracting in Theory and Practice*, Library Association, London.
- Larson, R.R. (1991), "The decline of subject searching: long-term trends and patterns of index use in an online catalog", *Journal of the American Society for Information Science*, Vol. 42 No. 3, pp. 197-215.
- Leydesdorff, L. (2006), "Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports?", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 5, pp. 601-13.
- McIlwaine, I.C. (2010), "Universal decimal classification (UDC)", in Bates, M.J. and Maack, M.N. (Eds), *Encyclopedia of Library and Information Sciences*, 3rd ed., Vol. VII, Taylor & Francis, London, pp. 5432-9.
- Mai, J.-E. (2000), "The subject indexing process: an investigation of problems in knowledge representation", unpublished doctoral dissertation, The University of Texas, Austin, TX, available at: http://individual.utoronto.ca/jemai/Papers/2000_PhDdiss.pdf (accessed July 20, 2010).
- Mai, J.-E. (2011), "The modernity of classification", *Journal of Documentation*, Vol. 67 No. 4, pp. 710-30.
- Maniez, J. (1997), "Database merging and the compatibility of indexing languages", *Knowledge Organization*, Vol. 24 No. 4, pp. 213-24.
- Olson, H.A. (2001), "The power to name: representation in library catalogs", *Signs*, Vol. 26 No. 3, pp. 639-68.

-
- Olson, H.A. (2002), *The Power to Name: Locating the Limits of Subject Representation in Libraries*, Kluwer Academic, Dordrecht.
- Ørom, A. (2003), "Knowledge organization in the domain of art studies – history, transition and conceptual changes", *Knowledge Organization*, Vol. 30 Nos 3/4, pp. 128-43.
- Ovesen, J. (1989), "Classificare necesse est. Lidt om klassifikation i antropologi-en – og også ellers (Classification is necessary. A little about the classification of anthropology – and also in general)", in Harbo, O. and Kajberg, L. (Eds), *Orden I Papirerne – En Hilsen Til J.B. Friis-Hansen*, Danmarks Biblioteksskole, Copenhagen, pp. 117-22.
- Pattuelli, M.C. (2010), "Knowledge organization landscape: a content analysis of introductory courses", *Journal of Information Science*, Vol. 36 No. 6, pp. 812-22.
- Pors, N.O. (2005), "Studerende, Google og Biblioteker: En Undersøgelse af 1694 Studerendes Brug af Biblioteker og Informationsressourcer", Biblioteksstyrelsen og Danmarks Biblioteksskole, København, available at: www.statensnet.dk/pligtarkiv/fremvis.pl?vaerkid=45550&repid=1&iarkiv=1 (accessed May 16, 2011).
- Riesthuis, G.J.A. and Bliedung, St. (1991), "Thesaurification of the UDC", *Tools For Knowledge Organization and the Human Interface*, Vol. 2, Index Verlag, Frankfurt, pp. 109-17.
- Rowley, J.E. and Farrow, J. (2000), *Organizing Knowledge: An Introduction to Managing Access to Information*, 3rd ed., Gower, Aldershot.
- Sparck Jones, K. (2005), "Revisiting classification for retrieval", *Journal of Documentation*, Vol. 61 No. 5, pp. 598-601.
- Salton, G. (1996), "Letter to the editor. A new horizon for information science", *Journal of the American Journal for Information Science*, Vol. 47 No. 4, p. 333.
- Scerri, E. (2007), *The Periodic Table, Its Story and Its Significance*, Oxford University Press, Oxford.
- Shachaf, P. (2009), "The paradox of expertise: is the Wikipedia reference desk as good as your library?", *Journal of Documentation*, Vol. 65 No. 6, pp. 977-96.
- Soergel, D. (1985), *Organizing Information: Principles of Data Base and Retrieval Systems*, Academic Press, Orlando, FL.
- Soergel, D. (2009) "Digital libraries and knowledge organization", in Kruk, S.-R. and McDaniel, B. (Eds), *Semantic Digital Libraries*, Springer, Berlin, pp. 9-39, available at: www.dsoergel.com/NewPublications/SoergelDigitalLibrariesandKnowledgeOrganization.pdf (accessed May 16, 2011).
- Stamos, D.N. (2004), "Book review of: *Discovery and Decision: Exploring the Metaphysics and Epistemology of Scientific Classification*", *Philosophical Psychology*, Vol. 17 No. 1, pp. 135-9.
- Svenonius, E. (2000), *The Intellectual Foundation of Information Organization*, The MIT Press, Cambridge, MA.
- Villen-Rueda, L., Senso, J.A. and Moya-Anegón, F. (2007), "The use of OPAC in a large academic library: a transactional log analysis study of subject searching", *The Journal of Academic Librarianship*, Vol. 33 No. 3, pp. 327-37.

Corresponding author

Birger Hjørland can be contacted at: bh@iva.dk