Theories of knowledge organization — theories of knowledge

Keynote speech given March 19, 2013, at the 13th Meeting of the German ISKO (International Society for Knowledge Organization), Potsdam, 19th to 20th March 2013

Birger Hjørland
Outline

1. Ontological commitment
2. Scientific versus bibliographic classifications
3. The epistemological basis of classifications
4. Approaches to knowledge organization
   4.1 Automatic versus human classification
   4.2 User-based and cognitive approaches
   4.3 Facet classifications
   4.4 Numeric taxonomic approaches
   4.5 Bibliometric approaches
   4.6 Domain-analytic classification
5. Conclusion
1. Ontological commitment
1. Ontological commitment

The core issues in knowledge organization are to determine semantic relations between concepts and to ascribe subjects to documents.

For example:

- To say that a cat is a mammal
- To say that a given document is about cats

These issues also involve determining the meaning of the words “cat”, “mammal”, “aboutness” (and further concepts such as “species”, “concepts”, “generic relation” etc.)
1. Ontological commitment

Normally non-experts would just say that we know what a cat is and that we know that it is a mammal. If challenged we might look it up in an authoritative source, either a general encyclopedia or an authoritative biological handbook or ask some experts.

In the case of cats and mammals, this might be safe, but in many other cases there are no consensus among experts. In main-stream biological systematics major groups of animals (such as fishes and reptiles) are no longer regarded as valid systematic units.
1. Ontological commitment

Cladistics is a novel classificatory method and philosophy adopted by zoologists in the last few decades, which has provided a rather turbulent state of zoological classification (see Blake, 2011).

In brief the new cladistic approach prefers to group animals based exclusively on *shared ancestry* rather than on *shared features* such as in the traditional Linnaean system.
1. Ontological commitment

How to classify animals (and even to say what a species is both in an abstract and concrete sense as in cats) is not something that is simply “given” but it is something which depends on background theories (“paradigms”), and in biological systematics we have today different competing paradigms.

This may be hard to understand and to accept, because in the traditional epistemology we start seeing and from there construe our theories based on the given. Here things are turned up-side down: Our theories determine what we see something as!
1. Ontological commitment

“The notion of ontological commitment has come to prominence in the second half of the twentieth century, mainly through the work of Willard Quine [1908-2000] […]. On Quine’s view the right guide to what exists is science, so that our best guide to what exists is our best current scientific theory: what exists is what acceptance of that theory commits us to.” (Craig, 1998)
1. Ontological commitment

Of course things may be classified in many ways. Animals may, for example, be classified according to weight and size, color, “sweetness”, and usefulness for human beings.

This is often done, for example, in books for children, or by authorities: pests etc. There is no one correct way of classifying things, and classifications should always be considered in relation to their purposes, and everything is always classified in many ways.

Where does this leave KO as a scholarly field?
“At present, many, perhaps most, current bibliographic classifications for mammals reflect quite outdated science. The latest edition of DDC, for example, arranges mammals in essentially the same way as the second edition of 1885” (Blake, 2011, p. 469)

“Such “outdated” classifications may still do their job well. The library of the Zoological Society of London uses its own scheme, devised in the 1960s and largely based on the Bliss Bibliographic Classification, to classify the monographs it holds. The librarian reports that, in most cases, her patrons are able to retrieve items and browse the collection effectively” (Blake, 2011, p. 469-470)
1. Ontological commitment

Can it really be true that such “outdated’ classifications may still do their job well?

Might the reason be that library classifications do not serve advanced retrieval purposes (e.g. within front-end research)? Or that libraries and databases do not support the dissemination of new knowledge to the general public?

If we have such low level of ambition concerning classification systems, is there then a need for KO as a scholarly research discipline? (Can the kind of work done in revising DDC be said to be scholarly?)
1. Ontological commitment

Even if there is no “one true classification”, it does not mean that any classification is as good as anyone else.

If KO is to be taken seriously as a scholarly field, it must be based on knowledge about the implications of alternative ways of classifying.

We should not say “we know that X is a kind of Y”. We should say that according to a given view X is considered a kind of Y (but according to other views and interests X is a kind of Z).

The scholarly basis for classification is to consider the underlying “paradigms” and make an informed choice.
2. Scientific / scholarly versus bibliographic classifications
2. Scientific versus bibliographic classifications

Mai (2004, p. 41) argued that “scientific classification of natural objects, and the bibliographic classification of the content of a document, are distinct …”.

I find this understanding harmful because it undermines the important relation between subject knowledge and bibliographical classification (e.g. between knowledge about zoological taxonomy and the design of classification systems on animals for bibliographic databases).
2. Scientific versus bibliographic classifications

The way biologists classify living organisms is, for example, reflected in bibliographical classification schemes such as the UDC (with some delay).

Blake (2011) writes that cladistics is a novel classificatory method and philosophy adopted by zoologists in the last few decades, which has provided a rather turbulent state of zoological classification.

Both scientific classification and bibliographic classification are subject to the same kinds of theory-dependence, interpretation and difficulties (i.e. the paradigm shift from Linnaean to cladistic classification).
2. Scientific versus bibliographic classifications

The same was shown by Ørom in the field of art: The way works of art are presented in museums, the way histories of art are organized and the way bibliographical classifications are organized are all dependent on the view of art (theory or philosophy of art) which has dominated the people making the organization.
2. Scientific versus bibliographic classifications

Figure 1
Social values – worldviews – scholarly paradigms
(After Ørom, 2003, p. 132)
3. The epistemological basis of classifications
3. The epistemological basis of classifications

- Some classifications are based on logic (e.g., that even numbers are numbers). The philosophical school of "conceptual analysis" is an attempt to generalize the use of a priori analysis for classification (see Hanna, 1998).

- Some classifications are based on empirical studies. A drug is classified as, e.g. tranquilizer, based on medical experiments.

- Some classifications are based on human conventions (e.g. the borders of a country, who is a royal person).
3. The epistemological basis of classifications

• Some classifications are based on heritage (e.g., who belongs to a certain family). The cladistics school in biological systematics which today is the dominant school is based on this principle.

• Some classifications are based on purpose (e.g. tools for cooking).

• Some classifications are based on a mixture of criteria (e.g., combined logical, empirical, historicist and pragmatic criteria)
Logical, empirical, historicist and pragmatic methods may each have applications for which they are especially relevant but each may also be generalized and used more widely because of traditions and ideologies.

How do we determine whether one or another classification is best? To evaluate a classification is to consider the methods by which it has been produced and to evaluate the logic, empirical studies, knowledge of human conventions, the genealogy, and the goals the classification is meant to serve.
3. The epistemological basis of classifications

In other words: To evaluate a classification is to engage in the research which lies behind the classification in order to check its validity and relevance.

All research is influenced by epistemological norms or commitments. There is no simply “correct way of doing research” or one correct and all-encompassing “scientific method” and also in the theory of knowledge consensus is rare. In my view versions of pragmatism/activity theory are the best candidate for fruitful philosophy of enquiry, but this issue is still open and is in a somewhat confused condition today.
3. The epistemological basis of classifications

The classical theories of empiricism and rationalism are still very much alive and influential in contemporary research (although mostly unrecognized). These theories have been characterized as a trap, and the point here is that if we understand their shortcomings, we may avoid the trap and do better research leading to better classifications. Empiricism and rationalism used to be considered the fundamental epistemological positions (and their combination was tried by the logical positivists in the beginning of the 20th century without success).
3. The epistemological basis of classifications

Because of their shortcomings, we need to include some alternatives. I consider four theories the basic epistemological theories: Empiricism, rationalism, historicism and pragmatism/activity theory.

But there are many labels in use today, including actor-network theory, cognitivism, critical rationalism, critical realism, critical theory, dialectical materialism/Marxism, feminist epistemology, hermeneutics, paradigm theory, phenomenology, postmodernism (late modernism), semiotics, social epistemology and social constructivism.
I do not think that all these epistemologies have important different implications for KO. As pragmatic philosophers say: If a theory is of no practical consequence it is indifferent and trivial.

The most important thing in the criticism of empiricism, rationalism and “positivism” is:

Knowledge is a social and historical product made to serve certain purposes and interests. It is important to reconsider knowledge claims in the light of new purposes, conditions and interests.
Example: “Textbooks like Harter (1986), Lancaster (2003), Large, Tedd & Hartley (2001), and Svenonius (2000) can be characterized as texts that solidify the use of technical and managerial language in LIS in the sense that they are basically how-to books, constantly referring to techniques, standards, principles, methods and rules. If one’s professional knowledge base has such texts at its foundation, no critical attitude is developed nor demanded because these textbooks do not question at all the role of information seeking or of knowledge organization systems in culture and society. …
3. The epistemological basis of classifications

They do not provide students with a language, an understanding, a knowledge that make them capable of participating in public discourse debating the functionality and legitimacy of these systems” (Andersen, 2005)

I believe Jack Andersen’s quote can be interpreted as a critical epistemological view of KO. (And, by the way, JA is inspired by activity theory).

However, a critical view cannot be separated from knowledge about technical aspects.
4. Approaches to knowledge organization

4.1 Automatic versus human classification
4.2 User-based and cognitive approaches
4.3 Facet classifications
4.4 Numeric taxonomic approaches
4.5 Bibliometric approaches
4.6 Domain-analytic classification
4.1 Automatic versus human classification
4.1 Automatic versus human classification

In overviews of KO a fundamental difference between computer based versus human based classification and indexing is often made. In Hjørland (2011) I argue, however, that this distinction is theoretically unfruitful.

Both human indexers and programmers are guided by their knowledge (theories) which – at the deepest – is connected to their (implicit) theories of knowledge (of which the most important are (empiricism, rationalism, positivism) on the one hand and (historicism, hermeneutics, pragmatism, AT) on the other (to be demonstrated below))
4.2 User-based and cognitive classifications
4.2 User-based and cognitive classifications

User-based and cognitive views have been influential in KO since the 1970s. Hjørland (2013b) is a critical analysis of this approach.

If KO addresses questions such as *Should document A be classified in class X? Is term A synonymous with term B?* It is difficult to understand how the study of users (qua users) can provide the knowledge needed to answer such questions.

It is therefore difficult to understand why user-studies have been so popular an approach in KO.
One hypothesis is that this is caused by the influence of empiricist/positivist ideals of science:

It seems better science to make empirical studies of users than to engage in say literary genre studies, theories of art or the philosophy of biological taxonomy. (It is also much easier to avoid difficult theoretical problems by basing KO on the study of user-preferences).

User-based approaches is thus seen as an ideology with a problematic basis with roots in empiricism.
4.3 Facet classifications
4.3 Facet classifications

In Hjørland (2013a) I found that the facet analytic approach is based on the epistemology of rationalism.

The strength of this approach is its logical principles and the way it provides structures in knowledge organization systems (KOS).
4.3 Facet classifications

The main weaknesses are

1) its lack of empirical basis and
2) its speculative ordering of knowledge without basis in the development or influence of theories and socio-historical studies.

It seems to be based on the problematic assumption that relations between concepts are a priori and not established by the development of models, theories and laws.

This tradition thus demonstrates how rationalism as a theory of knowledge has influenced KO.
4.4 Numeric taxonomic approaches
4.4 Numeric taxonomic approaches

Statistical methods such as cluster analysis, factor analysis etc. are used in many different sciences and on many different kinds of data (e.g. for classification of diseases or biological organisms).

They are also used for classifying documents (vector space models, latent semantic indexing, etc.) and may therefore also be considered an approach to KO.

(Normally Information Retrieval (IR) and KO are considered different fields, but I argue that they should be considered different approaches).
4.4 Numeric taxonomic approaches

IR is an extremely wide and complex field, and it may seem hasty and problematic to go into this field in such an overall way as is attempted here. However, these techniques are competing with other approaches to KO (and seemingly have much more success and authority in academia today). I therefore feel that we in KO have to take numeric taxonomic/IR-approaches very seriously, and if we want to make room for other approaches, we have to provide convincing argumentation about the limits of mainstream IR.
4.4 Numeric taxonomic approaches

In this presentation I shall limit myself to approaches based on statistical measures of similarity between documents (including between queries and documents).

Such statistical measures look like “objective science”. However, things may be similar in many different ways, and the seemingly objective nature of these approaches may be illusionary:
“Even in the field of numerical taxonomy, where the use of similarity coefficients has been even more widespread than in information retrieval, Jackson, Somers and Harvey (1989) were moved to conclude that “the choice of a similarity coefficient is largely subjective and often based on tradition or on a posteriori criteria such as the “interpretability” of the results”, and went on to quote Gordon (1987): “Human ingenuity is quite capable of providing a post hoc justification of dubious classifications”. ” (Ellis, Furner-Hines & Willett, 1993, p. 144)
4.4 Numeric taxonomic approaches

“Presented in these terms, the history of research into the use of similarity coefficients in text retrieval appears to betray a lack of progress” (Ellis, Furner-Hines & Willett, 1993, p. 141).

My claims are 1) that all numeric taxonomic approaches need to have criteria on how to choose among alternative algorithms 2) that such criteria must be based on subject theories (e.g. cladistic research) on what should be considered “similar”.
4.4 Numeric taxonomic approaches

It should also be considered that IR is applied to collections of descriptions of objects (e.g. animals). We therefore have different levels of epistemology:

1) The way the objects are described in the literature influenced by various paradigms.

2) The way a given collection is made influences which paradigms are dominating.

3) The way a given similarity measure is chosen supports given paradigms at the expense of others.
Without information from substantial theories, IR-approaches are based on problematic empiricist ideals and must fail.

(If they are successful after all, it might be that subject knowledge has been used indirectly, e.g. by relevance-assessments).
4.5 Bibliometric classifications
Methods based on *citation analysis* (e.g. co-citation analysis and bibliometric coupling) represent a unique set of approaches, which can be used to construe atlases of sciences, provide candidate terms for thesauri and many other kinds of knowledge organizing systems (KOS).

What are the *principal* strengths and limits of such methods?
4.5 Bibliometric classifications

In general, it cannot be expected that methods based on citation analysis are able to produce intellectual maps such as geographical maps, biological taxonomies or periodical systems. “A geographical structure, for example, places different regions in a structure that is autonomous in relation to the documents that are written about those regions. You cannot produce a geographical map of Spain by making, for example, bibliometric maps of the literature about Spain [Yet such autonomous structures as maps of Spain are often very useful for information retrieval about Spain]” (Hjørland, 2002, p. 452).
4.5 Bibliometric classifications

In Hjørland (submitted) I make the distinction between KOS reflecting intellectual KO versus KOS reflecting social KO:

• The intellectual aspect of KO is knowledge organized in concepts, propositions, models, theories, and laws. Such intellectual organizations are primarily structured via relations of explanatory coherence (Thagard, 1992, p. 9), which are again primarily related to questions concerning truth.
4.5 Bibliometric classifications

• The social aspect of KO is knowledge organized into academic departments, disciplines, cooperative networks, administrative bodies etc. Such social organizations are primarily structured by the social division of labor in societies which are again primarily related to questions concerning social relevance, authority and power.

I argue that citation-based approaches are by nature social organizations, and by implication the principal limits of these approaches are tied to the relation between intellectual and social KO.
4.5 Bibliometric classifications

Bibliographic methods therefore cannot render subject knowledge superfluous (but are themselves – like numeric taxonomy – dependent on subject knowledge).

Although bibliometrics is often associated with domain-analysis, I here argue for considering these approaches separately.
4.6 Domain-analytic classification
4.6 Domain-analytic classification

The domain-analytic (DA) view first of all recognizes the need for subject knowledge in classification and indexing.

A fine domain-analytic study is Blake (2011) who demonstrates solid knowledge about zoological taxonomy and the competing approaches in the field (cladistics, evolutionary taxonomy and the Linnaean system).

He also carefully discusses the relations between scientific theory, quasi-taxonomic groupings, the specific demands that information retrieval puts on classifications (including the principles of literary warrant).
Finally Blake describes the classifications used by biologists in their writings (monographs) and reveals the tendency to use conflicting or inconsistent classifications (corresponding to Ørom’s (2003) concept “bricolage”).

The domain-analytic view is the only view which is fully committed to exploring knowledge organization in the light of subject knowledge and substantial scholarly theories (and their epistemological basis).
4.6 Domain-analytic classification

The criticism raised against other approaches to KO is made from the theoretical position of domain-analysis. In all cases it was found that the subject-knowledge perspective of DA cannot be dismissed.

That is not to say that other approaches (e.g. IR-approaches) are not the most efficient approaches we have today. I am just claiming that their theoretical bases are problematic and that such approaches might be improved by considering domain-analytic perspectives.
4.6 Domain-analytic classification

Have arguments been put forward against domain-analysis?

Not really. Even leading scholars of other approaches seem to avoid direct discussions. The resistance towards DA seems mostly to be unvoiced or unspoken.

Although DA is gaining influence, there still seems to be a silent resistance. Why?
4.6 Domain-analytic classification

Some possible reasons may be:

1) the need to design concrete systems rather than to develop theoretical principles

2) the problematic tendency to use techniques, not theory to direct scientific practice (the so-called “law of the instrument”)

3) The displeasure to be involved in subject knowledge (and the fear of loosing one’s own disciplinary identity)
5. Conclusion
Conclusion

The necessity of subject knowledge in KO (as in the broader field of information science/ library and information science) is certainly not a new idea. This kind of knowledge has always been assumed in high standard libraries and bibliographical databases such as the National Library of Medicine and the MEDLINE database.

(In parallel to teaching qualifications: the higher the level of teaching, the bigger are the demands on subject knowledge, at the university level research qualifications are demanded).
Conclusion

In spite of this subject knowledge has been and still is extremely neglected in KO. My claim is that the neglect of the importance of subject knowledge has brought forward a crisis in KO, and that no real progress can be observed in the field.

There is real progress in digital technologies, and these are used to improve KO and IR, but this is mainly progress caused by developments outside our field.
Conclusion

Where does this place the theory of knowledge in KO? The first thing to say is that you cannot classify domains on the basis of theories of knowledge (or other metadisciplines, including genre studies, the sociology of knowledge etc.). Our studies have to be based on concrete domains.

Epistemology is, however, the best *general* background it is possible to educate people within in information science. It is the best general preparation we can provide people with in order to study any domain.
Conclusion

The general lesson I draw from epistemology is that knowledge is created by humans for some specific purposes and serve some interests better than others. Concepts and semantic relations are not a priori or neutral, but should be examined in relation to their implications for the users they are meant to serve.

I hope that my discussion of existent approaches to KO has been able to make this claim probable.
Thank you for your attention!


References

Ellis, David; Furner-Hines, Jonathan & Willett, Peter (1993). Measuring the degree of similarity between objects in text retrieval systems. Perspectives in Information Management, 3(2), 128-149


References


References


Hjørland, B. (submitted). Bibliometrics: a dynamic approach to knowledge organization
References


References


References


References


