IFLA Satellite Pre-Conference of the Classification and Indexing Section
"Looking at the Past and Preparing for the Future"
Florence, Italy, 20-21 August 2009

Animals belonging to the emperor: enabling viewpoint warrant in classification

Claudio Gnoli
University of Pavia. Library Service
via Ferrata 1, 27100 Pavia, Italy
gnoli@aib.it

Author biography: Dr. Claudio Gnoli has been working as an academic librarian since 1994. His main interest is classification theory. He has published papers on this subject in several international journals and conference proceedings. He is currently vice-president of the International Society for Knowledge Organization (ISKO), and member of the scientific advisory boards of the Universal Decimal Classification Consortium (UDCC) and of the journal "Knowledge organization".

Abstract: Recent research in knowledge organization has emphasized the need for representing different local perspectives, synthesized in Beghtol's principle of *viewpoint warrant*. A typical case is the taxonomy of animals: folk or non-Western taxonomies differ from those of academic biology, an extreme example being Borges's paradoxical "Chinese" classification. On the other hand, global services require interoperability between different viewpoints. The Integrative Level Classification (ILC) project is working at the basic structure of a general, interdisciplinary, freely-faceted system. Among its features is a set of special classes (*deictics*) that acquire different meanings according to the local context, thus allowing for interoperability between different local extensions of the scheme. Examples of their application to the classification of animals are shown.

Introduction

The theory and practice of intellectual tools like classifications, thesauri, subject heading lists, taxonomies, and ontologies is collectively known today as *knowledge organization*. Research on knowledge organization addresses the principles and techniques by which knowledge items can be ordered.

In recent years, much knowledge organization research has focused on a critical examination of these principles and techniques. While having developed from practical needs, like the arrangement of books on library shelves, or of bibliographic records in directories and catalogues, knowledge organization systems (KOS) have brought various theoretical biases with themselves. Indeed, a KOS is an expression not only of the structure of the real world (*ontological dimension*), and of our means of perceiving it (*epistemological dimension*), but also of the cultural milieu and pragmatic purposes providing the context for its development (*sociological dimension*) [Hjørland & Hartel 2003].

Classical KOSs, like the Dewey Decimal Classification, the Universal Decimal Classification or the Bliss Bibliographic Classification, have adopted a universal perspective, basically expressing the first two dimensions, ontological and epistemological, which can be assumed as common to any individual human user. However, the third dimension, sociological, seems to be another unavoidable component of any KOS. Some even think that this component be the main one, making any attempt of universality problematic [Maniez 1997; Hjørland 2004]. This view can be too pessimistic [Szostak 2008], leading to a relativistic way of thinking, in which the only possible task of knowledge organization research would be a sociological analysis of how communities working in given domains produce their own KOS. In principle this would even imply the impossibility of building such interoperability tools as multilingual thesauri or top-level ontologies, although in practice these are produced and used in some form.

Viewpoint problems

In any case, contemporary authors generally agree that the perspective of a KOS, including its philosophical assumptions, its cultural origins, and its pragmatic purposes, should be made explicit, rather than remain implicit hence potentially misleading for its users. This idea has been received in formulating the theme for the next conference of the International Society for

Knowledge Organization: “Paradigms and conceptual systems in knowledge organization” [Gnoli & Mazzocchi in prep.].

A wide discussion of bias in classification, both scientific and bibliographical, has been opened by Bowker and Star [1999], who showed the effects of classifying e.g. diseases in a way or another. This theme has been received by Ridi [in press] to warn information users about the importance and the implications of using classifications and taxonomies, both in bibliographic searches and in everyday life. The terminology used to label subjects itself can be biased towards culturally dominant groups, like middle-class white males: various social prejudices can thus be hidden in such a widely spread KOS as the Library of Congress Subject Headings [Olson 2002]. This can make its use problematic for different perspectives, like women's studies and feminism [Kublik et al. 2004].

Another possible kind of bias is political. In the thesauri of international organizations, the term `development` is defined only in its economic meaning, suggesting that developing countries should “develop” in a capitalistic sense, but not in social, educational, artistic, or spiritual senses [Severino 2005]. The Library of Congress Classification is remarkable for treating military sciences and naval sciences as two main classes, where another KOS could represent them as subclasses or facets of political sciences. Naval sciences are particularly irrelevant for countries without any coast in their territory. Also evident are biases in the Soviet and Chinese library classifications, adopting Marxism, Leninism, Maoism etc. as their first main classes, like a presupposition for any other form of knowledge.

Cultural differences in classification are also reported on small scale. While studying knowledge of potato varieties in the traditional agriculture of Liguria (Italy), Angelini [2005] observes that “different local names can be referred to the same variety, but also, on the contrary, different varieties are called by the same name. To Giacumin from Vobbia, the same *quarantina* potatoes that are cultivated in Croce, Pentema, and Montoggio, villages just few kilometers away,

are completely different varieties that he calls by different names, and I don't know how but he is able to tell them apart!”

Probably the most basic kind of bias in knowledge organization is that produced by profound differences between cultures that developed separately, like the Western vs. Eastern ones. Relevant cases have been studied by Kwaśnik & Chun [2004], reporting how the Korean version of the Dewey Decimal Classification required a new subclass of 700 “arts” for calligraphy, as in Far-Eastern culture this is listed among the major arts. Kinship structures also need to be represented in different ways according to the cultural context [Kwaśnik & Rubin 2004]. In the second edition of the Bliss Classification, a radical choice has been made to organize Eastern philosophy by different facets than Western philosophy [Biagetti 2009]; still, the separation of philosophy from religion, built in the main classes of all Western classifications, seems itself to be an unnatural representation of Eastern wisdom, and the same could be said for Medieval Western wisdom.

The ultimate example of unexpected categories in an exotic classification is that described by Borges [1964], claiming that it comes from a Chinese encyclopaedia. Actually this is presented to serve as a summa of all kinds of inconsistency that can be found in real classifications:

a animals that belong to the Emperor
b embalmed animals
c animals that are trained
d suckling pigs
e mermaids
f fabulous animals
g stray dogs
h animals included in the present classification
i animals that tremble as if they were mad
j innumerable animals
k animals drawn with a very fine camelhair brush
l other animals
m animals that have just broken a flower vase
n animals that from a long way off look like flies

This classification is often quoted to discuss problems of inconsistency. Indeed, each class appears to be the result of applying a different characteristic of division, a different facet, a different perspective. The whole scheme is thus extremely idiosyncratic, suggesting that any other

classificationist, or even the same classificationist in another moment, could produce a different scheme. Which can seem quite discouraging when we are looking for an optimal KOS to be shared on the global scale.

Matching local and global KOSs

Despite these problems, scholars dealing seriously with a given phenomenon usually agree on many parts of its classification, leaving aside those aspects that are not yet clear at the current stage of research. No zoologist starts her classification by main classes like Borges's ones: she will rather mention such general groupings as molluscs, annelids, arthropods, chordates. Is this arbitary? Is it only the expression of a particular academic community, having imposed its KOS over those of the minorities for some accidental or tendentious reason? Probably most zoologists would agree that the standard taxonomy is rooted in real relationships between animals, that can be understood in some reliable way by our means of knowledge, though conceding that many details could be corrected and developed in future. In other words, the ontological dimension seems to play a major role in determining scientific taxonomy, as compared with the epistemological and the sociological.

A method to check this assumption is to compare the standard scientific taxonomies with those used in cultures that had little or no contact with the modern West, so that they have presumably not been influenced by it. Diamond [1966] reports that the Fore people in New Guinea used 110 specific names to identify birds: these largely corresponded to the 120 species identified by zoologists, with 93 one-to-one correspondences and most of the remaining names referring to strictly-related species or to male and female forms of species with a great sexual dimorphism. Even stronger correspondence was found by Mayr with bird names used in the Arfak mountains, also in New Guinea, although the same people made no distinction between the many species of ants identified by biologists in their region [Wilson 1992]. Berlin et al. [1966] compared 200 plant names in the Tzeltal language spoken by a community in Chiapas (Mexico) with the respective

species in the standard botanical nomenclature: 82 names resulted to be underdifferentiated as compared to the botanical species, 68 (including 40 introduced after the Spanish conquest) exactly matched them, and 50 were overdifferentiated. Such results generally suggest that both the taxonomical units identified by biologists and those identified by native people do have a natural foundation; on the other hand, natives are less specific or precise for organisms that lack any practical interest for them, like in the case of ants.

Taxonomies of organisms thus have both ontological and pragmatical bases, which should be reflected in knowledge organization in some way. There is a need both for a general way to refer to concepts as objectively as possible, and for ways to represent local uses and perspectives. The latter requirement is considered by Beghtol [1998] as a kind of *warrant*: in the same sense as the traditional classification principle of literary warrant recommends that classes reflect the occurrence of topics in actual documents, *viewpoint warrant* should ensure that they reflect the occurrence of concepts, and the relationships between them, in actual cultures. This is also an ethical principle [Beghtol 2002], as in the global information context no culture should be privileged by knowledge organization, rather every one should find its own perspective represented.

How can viewpoint warrant be enabled in practice? One simple way is to develop systems explicitly reflecting the particular perspective of a community of knowledge users, as is recommended in the domain analytic approach. This, however, conflicts with the other need that information can be shared on a global basis, that is, with the requirements of interoperability. In order to establish connections between different classifications, indeed, one needs some way to refer each couple of classes coming from different schemes to a common frame, independently from their specific viewpoints, domains, contexts.

In other words, for mapping two KOSs that adopt special viewpoints, a third “neutral” KOS should exist, at least in the minimal form of concept identifiers to which concepts of each KOS can be referred [Coates 1970]. Although complete neutrality can be viewed as utopian, some intendedly neutral scheme is needed for technical purposes. In order to minimize its biases, such switching

scheme should adopt a maximally general and objective viewpoint. This will then be distinguished from explicitly biased KOSs, aiming at reflecting knowledge from particular viewpoints. Indeed, Beghtol [1998] suggests that a system enabling viewpoint warrant should "be able to support multiple perspectives in a looser structure"; it thus "would presumably have the advantage of providing infinite hospitality for adding any viewpoint – cultural, multidisciplinary, disciplinary, or sub-disciplinary – that might arise in future".

A similar structure was attempted already by Wåhlin [1974], who worked at an "AR-Complex", that is "a coherent complex of classification systems" composed of a Reference unit (R) to which many different Adapted systems (A) could be attached: he created two adapted systems, one for products and another for building trade documentation. Parsons [1996; 2002] looked in the same direction for his MIMIC system, to manage "multiple views" in data modeling. Discussion about an "international comprehensive KOS" is still current in the context of digital information sharing [Boteram 2009].

The ILC project

Integrative Level Classification (ILC) is an international research project aimed at developing the basic structure of a general, interdisciplinary, freely-faceted KOS, able to serve as a reference scheme for organizing any kind of information collection. ILC follows the principle, recently expressed in the León Manifesto [ISKO Italia 2007], of representing the objects treated (phenomenon), the perspective under which they are treated (aspect), and the information medium (carrier) as three separate dimensions. Viewpoint refers to the aspect dimension, which includes communicative function, modality, application, discipline, theory, method, place and epoch of the recorded knowledge.

The basic structure of ILC is a tree of phenomenon classes, expressed by lower cases. These generally follow the standard knowledge of phenomena held in contemporary sciences. Thus,

animals are a subclass of organisms, and in turn have various subclasses (only the main ones shown here):

m	organisms
mq	animals
mqb	sponges
mqc	cnidaria
mqm	molluscs
mqn	annelids
mqr	arthropods
mqrd	arachnids
mqrh	crustaceans
mqri	insects
mqt	echinoderms
mqv	chordates
mqvg	cartilaginous fish
mqvh	ray-finned fish
mqvj	amphibians
mqvl	reptiles
mqvo	birds
mqvt	mammals

Viewpoint in ILC

Among the features of ILC is a set of special classes that acquire different meanings according to the local context. In linguistical terms they are *deictics*, that is expressions that change their meaning according to the present situation, like words such as “you”, “here”, or “tomorrow” do.

Deictics are represented in ILC by capital letters. Therefore, while mq always means “animals”, F can mean anything, depending on how it has been defined. In other words, any scheme adopting a particular viewpoint can potentially be represented by ILC classes A, B, C, D..., their subclasses AA, AB..., etc. Deictics can also occur as subclasses of standard classes: mqA, mqB etc. will mean animals of some type according to a local context.

If a KOS is maintained as a database, as is the case with ILC, separate tables can be used for the general reference schedule and for any local schedule expressing particular viewpoints. For any class containing deictics, like A, a special field will be filled with the equivalent class in terms of the standard scheme (written in square brackets in schedule display). Such equivalent classes can be just a simple class, e.g. mqvo "birds", that is stated to be equivalent to A for practical convenience, like having a more manageable notation, and for local relevance: as in the ASCII character set capital letters are ordered before small ones, computers will list them before the standard classes, as the "favoured host classes" [Ranganathan 1967, Section DG34-35] of the present information system.

The equivalent class can also be a compound, defined as the syntactical combination of various facets, that under the local viewpoint takes the status of a single whole. In the extreme case of Borges's taxonomy, the first subclass of "animals" is defined by the combination of relationships "belonging to the emperor [of China]". The relationships of these concepts would be represented in the standard scheme by quite complex combinations of facets:

tp55a2mo	empires, with emperor, in China
u8(mq)6i	economies, of animals, by private owner

from which the combination mq98u(6i(955(a)tp(2mo))) "animals, being a good, by private owner, being emperor, of empire, in China" can be constructed.

Clearly, this compound notation is not very practical to be managed by users adopting the viewpoint of the Chinese encyclopedia. This can then be defined as equivalent to the first subclass of animals in this viewpoint, mqA. In the same way it would be possible to define the other subclasses, so to produce a representation of this classification as mqA, mqB, mqC, etc.:

mqA	[mq98u(6i(955(a)tp(2mo)))]	animals belonging to the emperor
mqB	[...]	embalmed animals
mqC	[...]	animals that are trained
...		

In the current version of ILC, letter A is used to mean the favoured host class (or subclass) according to the present viewpoint; letters B to T for other favoured classes; and letters U to Z for other special meanings also depending on the local context:

mqU	the typical animals
mqX	some animals
mqXA	the actual animals (i.e. there are animals)
mqXX	what animals?
mqY	the individual animal (e.g. Laika)
mqZ	the mentioned animal (anaphoric or cataphoric)

The deictic U is of special interest in the present discussion, as it allows to express the general viewpoint of mankind, rather than that of a specific community. It can be described as the *anthropocentric favoured class*. For example, the phenomenon "stars" in ILC is hk. A general, neutral taxonomy of stars would list them according to some general astronomical principles. In this general perspective, our Sun is just one star among an immense number of others, therefore it would get a very specific notation, say hkxxxxxx. But human documents deal with the Sun much more often than with any other star, and consider it as by far the most relevant star. Shortened notation hkU then allows to represent it in a shorter way, and to list it before all other stars, except any one of specific interest to a local context (a research centre focused on the Dog Star could represent it as hkA, preceding hkU in ordered displays).

This is still a biased viewpoint, as the Sun in itself has no special character as compared to all the other stars; still, as the bias is the same for all human users, it does not need to be changed according to specific information contexts. Similar cases have already been identified in ILC for expressing concepts like "air" as a special gas mixture, "water" as a special chemical compound, "the Earth", "the continents of contemporary Earth", "currently prominent languages" such as English and Spanish, "contemporary countries" (the viewpoint of the last examples is actually dependent on time, but only on a big scale).

Testing and development

The ILC system is currently being tested in indexing on-line bibliographies and other Web resources in different domains, including human geography, bioacoustics, chemistry, and facet analysis. In particular, the system of deictics has been used until now in a bibliography on traditional culture and geography of a mountainous area in North-Western Italy, where deictics A to T stand for valleys in the Apennine range, which would have a much more complicate notation if had to be represented by their standard notation for landforms in the whole Earth [Gnoli 2008].

The citation order of compound classes including deictics follows the inversion principle of analytico-synthetic schemes, prescribing that classes listed before in the schedules be cited last within a single compound class. A document concerning the Curone valley H and its animals mq will thus be notated mq H, rather than H mq. This appears to be an effective solution, as deictic classes expressing the local valleys are less discriminating, in the context of a bibliography on that very region, than classes expressing other phenomena. Documents concerning animals will thus be primarily listed all together, and only subsequently differentiated according to any specific valley (although digital search will also allow to extract all records containing H, that is, concerning the Curone valley in any way).

This experience of indexing is providing some idea of how viewpoint can actually be represented and managed in a classification system. Still the case described is quite simple. While the technique assumes that any concept in a special KOS can be translated into some combination of facets of the general KOS, cases of problematic translation could be encountered in the actual work, and should then be analyzed in detail. Also, the complete mapping of a local scheme with a greater number of classes, including complex compounds like those of Borges's taxonomy, requires that the general reference scheme be developed at a further stage. For example, class mqB "embalmed animals" cannot be really defined until the general scheme lacks a class for the rather specific meaning "embalming". This is a good example of how many components of the system,

like the general scheme of phenomena, the way to express local meanings, and the syntax of faceted compounds, are all connected, hence need to develop slowly all together. Although basic principles for representing viewpoints have been described in this paper, more work is required in order to develop the system in a complete way.

Acknowledgments

I am grateful to Clare Beghtol, Thomas M. Dousa, and Riccardo Ridi for providing useful suggestions and references.

References

Angelini, Massimo, 2005, Varietà tradizionali, prodotti locali ed esperienze, L'Ecologist italiano, 1, n. 3, p. 230-275.

Beghtol, Clare, 1998, General classification systems: structural principles for multidisciplinary specification, in Structures and relations in knowledge organization: proceedings 5th International ISKO Conference, Lille, 25-29 August 1998, eds. Widad Mustafa el Hadi, Jacques Maniez, A. Steven Pollitt, Advances in knowledge organization, 6, Ergon, Würzburg, p. 89-96.

Beghtol, Clare, 2002, A proposed ethical warrant for global knowledge representation and organization systems, Journal of documentation, 58, 5, p. 507-532.

Berlin, Brent, Dennis E. Breedlove & Peter H. Raven, 1966, Folk taxonomies and biological classification, Science, 154, p. 273-275.

Biagetti, Maria Teresa, 2009, Philosophy in bibliographic classification systems, in The philosophy of classifying philosophy, eds. Cristiana Bettella & Massimiliano Carrara, Knowledge organization, 36, n. 2-3, p. 92-102.

Borges, Jorge Luis, 1964, Other inquisitions, 1937-1952, transl. Ruth L. C. Simms, University of Texas press, p. 101-106.

Boteram, Felix, 2009, Semantic interoperability in an international comprehensive knowledge organisation system, in Content architecture: proceedings ISKO UK conference, London, 22-23 June 2009, Aslib proceedings, in prep., also ISKO UK, <http://www.iskouk.org/conf2009/proceedings.htm>.

Bowker, Geoffrey C. & Susan Leigh Star, 1999, Sorting things out: classification and its consequences, MIT Press, Cambridge and London.

Coates, Eric J., 1970, Switching languages for indexing, Journal of documentation, 26, n. 2, p. 102-110.

Diamond, Jared M., 1966, Zoological classification system of a primitive people, Science, 151, p. 1102-1104.

Gnoli, Claudio, 2008, Potential of freely faceted classification for knowledge retrieval and browsing, presentation at 7th NKOS workshop, Aarhus, 19 September 2008, <http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2008/>.

Gnoli, Claudio & Fulvio Mazzocchi eds., in prep., Paradigms and conceptual systems in knowledge organization: proceedings 11th ISKO Conference, Rome, 23-26 February 2010, Advances in knowledge organization, 12, Ergon, Würzburg.

Hjørland, Birger, 2004, Theory of knowledge organization and the feasibility of universal solutions, paper presented at 8th ISKO Conference, London, 13-16 July 2004, DLIST, <http://dlist.sir.arizona.edu/389/>.

Hjørland, Birger & Jenna Hartel, 2003, Afterward: ontological, epistemological and sociological dimensions of domains, Knowledge organization, 30, n. 3-4, p. 239-245.

ISKO Italia, 2007, The León manifesto, <http://www.iskoi.org/ilc/leon.htm>, republ. in Knowledge organization, 34, 2007, n. 1, p. 6-8.

Kublik, Angela, Virginia Clevette, Dennis Ward & Hope A. Olson, 2004, Adapting dominant classifications to particular contexts, in Knowledge organization and classification in international information retrieval, eds. Nancy J. Williamson & Clare Beghtol, Haworth, Binghamton, p. 13-32.

Kwaśnik, Barbara Hanna & You-Lee Chun, 2004, Translation of classifications: issues and solutions as exemplified in the Korean Decimal Classification, in Knowledge organization and the global information society: proceedings of 8th ISKO Conference, London, 13-16 July 2004, ed. Ia C. Mc Ilwaine, Advances in knowledge organization, 9, Ergon, Würzburg, p. 193-198.

Kwaśnik, Barbara Hanna & Victoria Rubin, 2004, Stretching conceptual structures in classifications across languages and cultures, in Knowledge organization and classification in international information retrieval, eds. Nancy J. Williamson & Clare Beghtol, Haworth, Binghamton, p. 33-48.

Maniez, Jacques, 1997, Database merging and the compatibility of indexing languages, Knowledge organization, 24, n. 4, p. 213-224.

Olson, Hope A., 2002, The power to name: locating the limits of subject representation in libraries, Kluwer: Dordrecht.

Parsons, Jeffrey, 1996, On the relevance of classification theory to database design, in Advances in classification research: proceedings 5th ASIS SIG/CR Classification research workshop, ed. Raya Fidel, Barbara Hanna Kwaśnik, Clare Beghtol & P. J. Smith, Information today, Medford, p. 131-140.

Parsons, Jeffrey, 2002, Effects of local versus global schema diagrams on verification and communication in conceptual data modeling, Journal of management information systems, 19, n. 3, p. 155-183.

Ranganathan, S.R., 1967, Prolegomena to library classification, 3rd ed., with assistance of M.A. Gopinath, SRELS, Bangalore, also in DLIST, <http://dlist.sir.arizona.edu/1151/05/ProlegomenaD.pdf>.

Ridi, Riccardo, in press. Il mondo dei documenti. Laterza, Roma.

Severino, Francesca, 2007, The term `development` in the thesauri of international organizations, The European journal of development research, 19, n. 2, p. 327-351.

Szostak, Rick, 2008, Classification, interdisciplinarity, and the study of science, Journal of documentation, 64, n. 3, p. 319-332.

Wåhlin, Ejnar, 1974, The AR-Complex: adapted systems used in combination with a common reference system, in Conceptual basis of the classification of knowledge, ed. Jerzy A. Wojciechowski, Verlag Dokumentation, Pullach bei München, p. 416-449.

Wilson, Edward O., 1992, The diversity of life, Penguin, London and Belknap Press, Cambridge (Mass.).