Articles

Rick Szostak.
Classifying the Humanities ......................................................... 263

Rongying Zhao and Shengnan Wu.
The Network Pattern of Journal Knowledge Transfer
in Library and Information Science in China .................. 276

Vinod Kumar B.L. and Khaiser Nikam.
Development of an Information Support System
for Yogic Science using Knowledge Organization
Systems .......................................................................................... 288

Papers from I Congress of ISKO Spain and Portugal /
XI Congress ISKO Spain, 7-9 November 2013,
University of Porto

Gustavo Silva Saldanha.
The Philosophy of Language and Knowledge
Organization in the 1930s: Pragmatics of Wittgenstein
and Ranganathan ................................................................. 296

Maria Aparecida Moura.
Emerging Discursive Formations, Folksonomy and
Social Semantic Information Spaces (SSIS):
The Contributions of the Theory of Integrative
Levels in the Studies carried out by the Classification
Research Group (CRG) .......................................................... 304

Virginia Bentes Pinto, Camila Regina de Oliveira Rabelo,
and Igor Peixoto Torres Girão.
SNOMED-CT as Standard Language for Organization
and Representation of the Information in Patient
Records .......................................................................................... 311

Fernanda Ribeiro.
The Use of Classification in Archives as a Means
of Organization, Representation and Retrieval
of Information ................................................................................ 319

Feature

Rebecca Green.
ISKO and Knowledge Organization’s 25th Anniversary:
The Future of Knowledge Organization and ISKO
Panel Discussion ....................................................................... 327

Book Review

Handbook of Metadata, Semantics and Ontologies.
Edited by Miguel-Angel Sicilia. Singapore: World

Books Recently Published ................................................. 335
KNOWLEDGE ORGANIZATION

KNOWLEDGE ORGANIZATION

This journal is the organ of the INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (General Secretariat: Vivien PETRAS, Humboldt-Universität zu Berlin, Institut für Bibliotheks- und Informationswissenschaft, Unter den Linden 6, 10099 Berlin, Germany. E-mail: secert@isko.org.

Editors

Richard P. SMIRAGLIA (Editor-in-Chief), School of Information Studies, University of Wisconsin, Milwaukee, NorthWest Quad Building B, 2025 E. Newport St., Milwaukee, WI 53211 USA. E-mail: smiraglia@uwm.edu

Hanne ALBRECHTSSEN (Book Review Editor), Institute of Knowledge Sharing, Frisersvej 1, 2, DK-2920 Charlottenlund, Denmark. E-mail: hal@iva.dk

Nancy WILLIAMSON (Classification Research News Editor), Faculty of Information Studies, University of Toronto, 140 St. George Street, Toronto, Ontario M5S 3G6 Canada. E-mail: william@its.utoronto.ca

Melodie Joy FOX (Editorial Assistant), School of Information Studies, University of Wisconsin, Milwaukee, NorthWest Quad Building B, 2025 E. Newport St., Milwaukee, WI 53211 USA.

Daniel Martinez AVILA (Editorial Assistant), School of Information Studies, University of Wisconsin, Milwaukee, NorthWest Quad Building B, 2025 E. Newport St., Milwaukee, WI 53211 USA. dmartinezavila@gmail.com

Hailey E. STRICKON (Editorial Assistant), School of Information Studies, University of Wisconsin, Milwaukee, NorthWest Quad Building B, 2025 E Newport St., Milwaukee, WI 53211 USA.

Editors Emerita

Hope A. OLSON, School of Information Studies, University of Wisconsin-Milwaukee, Milwaukee, NorthWest Quad Building B, 2025 E. Newport St., Milwaukee, WI 53211 USA. E-mail: holson@uwm.edu

Clare BEGHTOL, Faculty of Information Studies, University of Toronto, 140 St. George Street, Toronto, Ontario M5S 3G6, Canada. E-mail: clare.beghtol@utoronto.ca

Ingertraut DAHLBERG, Am Hirtenberg 13, 64732 Bad König, Germany. E-mail: IDalhberg@t-online.de

Editorial Board

Jonathan FURNER, Graduate School of Education & Information Studies, University of California, Los Angeles, Box 352840, Mary Gates Hall Ste 370, Seattle WA 98195-22250-900, Brazil. E-mail: renato.souza@fgv.br

Birger HJORLUND, Royal School of Library and Information Science, Copenhagen Denmark. E-mail: pnil167@iva.ku.dk

Barbara H. KWASNIK, School of Information Studies, Syracuse University, Syracuse, NY 13244 USA. E-mail: bkwasnik@syr.edu

Maria J. LÓPEZ-HUERTAS, Universidad de Granada, Facultad de Biblioteconomía y Documentación, Campus Universitario de Cartuja, Biblioteca del Colegio Máximo de Cartuja, 18071 Granada, Spain. E-mail: mllopez@ugr.es

Kathryn LA BARRE, The Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, MC-493, Champaign, IL 61820-6211 USA. E-mail: klbarre@illinois.edu

Widad MUSTAFA el HADI, Université Charles de Gaulle Lille 3, URFIST, Domaine du Pont de Bois, Villeneuve d’Ascq 59653, France. E-mail: widad.mustafoa@free.fr

H. Peter OHLY, Prinzennstr. 179, D-53175 Bonn, Germany. E-mail: president@isko.org

K. S. RAGHAVAN, K-LiteOE (Centre for Knowledge Analytics & Ontological Engineering), PES Institute of Technology, 100 Feet Ring Road, BSK, 3rd Stage, Bangalore 560085, India. E-mail: ksravagav@hotmail.com.

M. P. SATIJA, Guru Nanak Dev University, School of Library and Information Science, Amritsar-145 005, India. E-mail: satija_mp@yahoo.com

Aida SLAVIC, UDC Consortium, PO Box 90407, 2509 LK The Hague, The Netherlands. E-mail: aida.slavic@udc.org

Dagobert SOERGEL, Department of Library and Information Studies, Graduate School of Education, University at Buffalo, 534 Baldy Hall, Buffalo, NY 14260-1020. E-mail: dsmeergel@buffalo.edu

Renato R. SOUZA, Applied Mathematics School, Getulio Vargas Foundation, Praça de Botafogo, 190, 3o andar, Rio de Janeiro, RJ, 22250-900, Brazil. E-mail: renato.souza@fgv.br

Joseph T. TENNIS, The Information School of the University of Washington, Box 352840, Mary Gates Hall Ste 370, Seattle WA 98195-2840 USA. E-mail: jtennis@u.washington.edu

Maja ŽUMER, Faculty of Arts, University of Ljubljana, Askerceva 2, Ljubljana 1000 Slovenia. E-mail: maja.zumer@ff.uni-lj.si

Official Bi-Monthly Journal of the International Society for Knowledge Organization

ISSN 0943 – 7444

International Journal devoted to Concept Theory, Classification, Indexing and Knowledge Representation

Knowl. Org. 41(2014)No.4

José Augusto Chaves GUIMARÃES, Departamento de Ciência da Informação, Universidade Estadual Paulista–UNESP, Av. Hyginio Muzzi Filho 737, 17525-900 Marília SP Brazil. E-mail: guima@marilia.unesp.br

Birger HJORLUND, Royal School of Library and Information Science, Copenhagen Denmark. E-mail: pnil167@iva.ku.dk

Barbara H. KWASNIK, School of Information Studies, Syracuse University, Syracuse, NY 13244 USA. E-mail: bkwasnik@syr.edu

Maria J. LÓPEZ-HUERTAS, Universidad de Granada, Facultad de Biblioteconomía y Documentación, Campus Universitario de Cartuja, Biblioteca del Colegio Máximo de Cartuja, 18071 Granada, Spain. E-mail: mllopez@ugr.es

Kathryn LA BARRE, The Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel Street, MC-493, Champaign, IL 61820-6211 USA. E-mail: klbarre@illinois.edu

Widad MUSTAFA el HADI, Université Charles de Gaulle Lille 3, URFIST, Domaine du Pont de Bois, Villeneuve d’Ascq 59653, France. E-mail: widad.mustafoa@free.fr

H. Peter OHLY, Prinzennstr. 179, D-53175 Bonn, Germany. E-mail: president@isko.org

K. S. RAGHAVAN, K-LiteOE (Centre for Knowledge Analytics & Ontological Engineering), PES Institute of Technology, 100 Feet Ring Road, BSK, 3rd Stage, Bangalore 560085, India. E-mail: ksravagav@hotmail.com.

M. P. SATIJA, Guru Nanak Dev University, School of Library and Information Science, Amritsar-145 005, India. E-mail: satija_mp@yahoo.com

Aida SLAVIC, UDC Consortium, PO Box 90407, 2509 LK The Hague, The Netherlands. E-mail: aida.slavic@udc.org

Dagobert SOERGEL, Department of Library and Information Studies, Graduate School of Education, University at Buffalo, 534 Baldy Hall, Buffalo, NY 14260-1020. E-mail: dsmeergel@buffalo.edu

Renato R. SOUZA, Applied Mathematics School, Getulio Vargas Foundation, Praça de Botafogo, 190, 3o andar, Rio de Janeiro, RJ, 22250-900, Brazil. E-mail: renato.souza@fgv.br

Joseph T. TENNIS, The Information School of the University of Washington, Box 352840, Mary Gates Hall Ste 370, Seattle WA 98195-2840 USA. E-mail: jtennis@u.washington.edu

Maja ŽUMER, Faculty of Arts, University of Ljubljana, Askerceva 2, Ljubljana 1000 Slovenia. E-mail: maja.zumer@ff.uni-lj.si

Knowl. Org. 41(2014)No.4
Classifying the Humanities

Rick Szostak

Department of Economics, University of Alberta, Edmonton, Alberta, T6G 2H4, Canada
<rszostak@ualberta.ca>

Rick Szostak is professor of economics at the University of Alberta. His research is focused on facilitating interdisciplinary research and teaching; knowledge organization has become the dominant element in that research agenda. He authored two books and several articles that develop classifications of things (phenomena), theories and methods applied, types of data, ethical perspectives, research practices, and relationships. He contributed to the Integrative Levels Classification and is developing the Basic Concepts Classification, and is collaborating on a book about interdisciplinary knowledge organization. He is exploring how these classificatory principles can be applied to galleries, museums, archives, and the Semantic Web, as well as to libraries.


Abstract: A synthetic and universal approach to classification which allows the free combination of basic concepts would better address a variety of challenges in classifying both humanities scholarship and the works of art (including literature) that humanists study. Four key characteristics of this classificatory approach are stressed: a universal non-discipline-based approach, a synthetic approach that allows free combination of any concepts but stresses a sentence-like structure, emphasis on basic concepts (for which there are broadly shared understandings across groups and individuals), and finally classification of works also in terms of the theories, methods, and perspectives applied. The implications of these four characteristics, alone or (often) in concert, for many aspects of classification in the humanities are discussed. Several advantages are found both for classifying humanities scholarship and works of art. These four characteristics are each found in the Basic Concepts Classification (which is briefly compared to other faceted classifications), but each could potentially be adopted elsewhere as well.

Received: 12 November 2013 Revised: 3 June 2013 Accepted: 6 June 2014

Keywords: art, classification, classifying, works, scholarship, humanities, concepts, subjects

1.0 Introduction

There are multiple ways in which disciplines or groups of disciplines might be classified: subject matter, method, theory, epistemological perspective, or historical evolution among them. The Basic Concepts Classification (BCC) (Szostak 2013a) classifies in terms of subject matter. It recognizes that philosophy, and to a lesser extent history and geography, are different from other disciplines: they each treat a wide range of subjects. History and geography treat the temporal and spatial characteristics respectively of diverse subjects. Philosophy asks a variety of questions—what should be?; how should we study?; what is the real nature?—of diverse subjects. Efforts to fit philosophy and history into the humanities—for classificatory rather than administrative purposes—have thus generally proven problematic; recourse is generally made to a common tendency toward close reading of texts, but of course this is to some degree a characteristic of all scholarship (and humanists often do other things such as interviews). It is in general much harder to distinguish disciplines in terms of the methods they employ (of which there are only about a dozen, broadly speaking) than subjects, especially in the contemporary academy where humanistic approaches infuse the social sciences. Culture (including language) is one subject that is often treated in the humanities, but it has long also been studied by social scientists. Art (including literature) receives far more attention from humanists than from social scientists. This paper thus, somewhat unusually, defines the humanities as the study of art, including literature (see Szostak 2003).

Gnoli (2007) has urged the development of classifications that can simultaneously address both documents and the objects in the world that those documents address. One key purpose of this paper is to show that a synthetic approach grounded in basic concepts can simultaneously classify works of humanities scholarship and
works of art themselves. More generally, this paper will strive to show the value of the following four characteristics for classifying the humanities:

- A classification that is not grounded in disciplines.
- A classification that allows all concepts to be freely combined, but stresses a sentence-like combinatory structure linking things (phenomena), relationships, and properties.
- A classification that breaks complex concepts into basic concepts that at least potentially have broadly shared understandings across groups.
- A classification that treats the theories and methods and perspectives applied in a work.

The paper will suggest that a classification with each of these four characteristics is best suited to the needs of the humanities. It is nevertheless true that each of these characteristics is valuable in its own right. The paper can thus also inform and support efforts to add or enhance one or more of these characteristics in other classifications.

Still, it is useful in the next section to briefly distinguish the Basic Concepts Classification—which possesses each of the four characteristics above—from other classifications. The third section of the paper briefly discusses certain challenges that arise in classifying the humanities. The fourth section of the paper then addresses in turn various elements of a classification of the humanities, and shows how one or more of the characteristics outlined above can be applied. A brief concluding section summarizes the main advantages that result from each of the four characteristics above for classifying both humanities scholarship and works of art.

2.0 The Basic Concepts Classification

The advantages and disadvantages of a faceted versus enumerative approach have been long discussed in the KO literature (e.g. Broughton 2006). It is worthwhile here to briefly distinguish the BCC from other faceted approaches. The Bliss and Colon Classifications are organized around disciplines, and facets are described for each disciplinary class; the Universal Decimal Classification has some general facets but also subject-specific facets (Broughton and Slavic 2007, 731-2). The faceted approach to classifying the humanities described in Broughton and Slavic (2007) likewise has some general and some subject-specific facets. It is notable that in the penultimate paragraph of their conclusion, Broughton and Slavic (2007, 750) urge further research on interdisciplinarity and on the possibility of a classificatory structure not dependent on disciplines (they had earlier worried that it might not be possible to apply facets universally; 131-2). All of these faceted approaches necessarily treat synthetic combinations within disciplines differently from synthetic combinations across disciplines. The BCC shares with the Integrative Levels Classification (ILC) (2014) the distinction of being entirely grounded in things, relationships, and properties rather than disciplines.

The BCC can be distinguished from the ILC (and all other faceted classifications) by its approach to facets. In particular the BCC is able to eschew the use of facet indicators. Though the BCC allows the free combination of any concepts, it encourages a sentence-like structure. Most scholarly works, and likely most general works, investigate how one thing or set of things influence in a particular way a different thing or set of things (Szostak 2012a). The best way to classify such works is thus to synthetically link things and relators: (chemical)(reduces) (blood pressure) or (dogs)(bite)(mail carriers). A majority of works describe the properties of a thing: (steel)(is) (strong). Such works are also best captured synthetically, this time by linking a thing and a property. This sentence-like structure will prove particularly useful in classifying the humanities (see below). In particular it can be applied to both humanities scholarship and the works they study: (woman)(riding)(horse) or (gods)(celebrating). But eschewing facet indicators has a more general advantage: it greatly simplifies both user queries and machine programming for neither user nor machine need be acquainted with the logic of facet analysis (Broughton and Slavic 2007, 749 speculate that users might be able to just input concepts they wish to see combined in queries). Nor is it necessary to provide a strict ordering for facets as all of the above faceted systems do. Szostak (2013a) describes in detail how each of the dozen facets identified in the Bliss Classification, plus additional facets employed in the ILC, are captured in BCC.

The BCC employs the classification of relationships developed in Szostak (2012b). A class of some 200 adjectival/adverbial “properties” has been developed and organized into a couple of dozen flat hierarchies. These relators and properties can be combined with any of the thousands of things (phenomena) in the classification. The BCC thus encourages classification in terms of concept strings that capture the key arguments of a work. Such an approach allows us to simultaneously capture works and ideas, a desiderata noted by Gnoli (2007). We shall see below that such an approach is also well-suited to capturing the subject of a work of art. That is, it can cope simultaneously with works, ideas, and objects.

A further shared characteristic of the ILC and BCC is that they allow and encourage works to be classified in terms of the theory and method and perspective applied. This is information that users often seek: it is critical in
allowing scholarly users in particular to evaluate the likely relevance of a work (Szostak forthcoming). This facility might be achieved within any classification but is rarely pursued. This paper will suggest a novel application of this approach to works of art.

Szostak (2011) argued that ambiguity could be substantially reduced by breaking complex concepts—which are understood differently across disciplines or groups—into their constituent basic concepts, which are understood in a broadly similar fashion across groups and disciplines. Conceptual atomism suggests that shared understanding will be most likely for things and relationships that we regularly perceive. The BCC (and ILC) pursues the classification of works in terms of combinations of such things and relationships and properties.

3.0 Challenges in Classifying the Humanities

There is no one widely accepted approach to the subject classification of works of art. Dozens of classifications are used in music libraries, suggesting that classification is difficult and existing schemes do it poorly (Lee 2011). The Categories for Description of Works of Art (Getty Research Institute 2014a) and Cataloguing Cultural Objects (Baca et al. 2006; Getty Research Institute 2014b), by far the most widely used approaches to classifying art, recommend the use of controlled vocabulary in subject classification of art, but leave it to individual galleries to choose among several possible vocabularies. These vocabularies often take the form of thesauri (notably the Art and Architecture Thesaurus (AAT) (Getty Research Institute 2000)) rather than formal classification schemes; this opens the door to further differences in classificatory practice across institutions. This paper will recommend a simple but powerful approach that allows the sort of terms in the AAT—and, importantly, beyond—to be synthetically combined in a formal classification.

Subject is one key element that we would wish to capture in classifying works of art. But there are many others, including style, purpose, technique, provenance, and form. Each of these can also benefit, as we shall see, from one or more of the four characteristics listed above. Indeed the purpose of this paper is to suggest the breadth of uses of these classificatory characteristics in classifying the humanities.

Classifications should be evaluated in terms of their impact on users. Medaille (2010) notes that theatre artists find it troublesome to navigate large research libraries. Yet they list “seeking inspiration” to be one of their six main goals. If artists would turn to the literature for inspiration, but find it daunting to do so, then society suffers. And if they find libraries daunting, it must seem that it is not easy for them to find what they need. Inspiration will often come in art in the same way that innovation occurs in science: by juxtaposing previously separate ideas. In art as in science, then, we need a classification system that alerts users to items that might be related to their core area of interest. Users should not have to be familiar with and master a device such as the relative index in the Dewey Decimal Classification in order to achieve this kind of inspiration.

If there are challenges in classifying works of art, we might reasonably anticipate related challenges in classifying the scholarship of art. Casual empiricism suggests that humanities scholarship is also characterized by diverse and contested terminology, and by a multiplicity of theories. These challenges can also be addressed through application of a classification that has one or more of the characteristics listed above.

This paper, then, is broad in its approach. Its purpose is to show how a small set of classificatory characteristics can potentially address a wide range of actual or potential challenges in classifying the humanities. Future research (Szostak 2014b) will take a narrower focus on subject classification, and more carefully compare the approach recommended here to extant approaches, and provide many examples of the classification of actual works of art.

4.0 Phenomena

We must first classify art itself before worrying about subject, style, and other characteristics. A first challenge here is that many artifacts—notably pottery and textiles—are intended to be both useful and aesthetically pleasing. Indeed most of the artifacts created by humans embody some aesthetic elements. Our buildings, bridges, shoes, cutlery, furniture (the list goes on) could all be made much more cheaply if we focused only on their function and not their appearance. Works that address the practical side of such artifacts are usually given quite different subject headings, and shelved in quite different parts of the library, from works that treat their aesthetic side.

Should “china” or “fashion clothing” themselves be classified as works of art or as items of utility? General practice tends toward the latter, except when a work stresses their aesthetic elements. This is perhaps unfair: a set of china that is put on display in a china cabinet year-round, and only pulled out once a year for a special family dinner, is likely misclassified as an item that is primarily useful. But we hardly want either classificationists or classifiers to have to decide where particular artifacts (or works about these) fall on a continuum of art to practicality. If we will follow common practice and privilege utility over art by classifying most artifacts as useful arti-
facts, then it becomes particularly important to be able both to readily identify and find aesthetic treatments of such artifacts, and connect these to works regarding their practical attributes.

In other words, we have here a diverse set of artifacts (and many works about these) that do not fit neatly into just one hierarchy. And the hierarchies of works of art tend to be quite distant in most classification systems from the hierarchies of items of utility. Such artifacts are inevitably treated poorly in “universal” classifications that are organized around disciplinary silos, and treat engineering and economic outcomes in a quite different fashion from works of art. These artifacts reflect an intersection of utility and aesthetics that is hard to signify in an enumerative classification. It is much easier to address in a classification that stresses relationships. A user interested in china often wants to move seamlessly from reading works on pottery technology to works on pottery aesthetics.

Szostak (2014a) argues that only a synthetic classification without disciplinary boundaries could instantiate a “web-of-relations” approach to classification. One of the values of a web-of-relations approach is that it should be much easier for users to move between studies of aesthetics and of practicality. The user need not master disciplinary classes (or even know of the Relative Index) in order to move from (technology of) (Ancient Greek) (pottery) to (trade in) (Ancient Greek) (pottery) to (art of) (Ancient Greek) (pottery). And this facility in turn could encourage ever-better integration of beauty and practicality in our lives. The web-of-relations will likewise facilitate the search for works that are similar in terms of any characteristic included in the synthetic classification of a particular work.

What sort of subclasses might we attribute to art? Szostak (2014a) argued also that much of the concern that there are diverse ways of subdividing classes into subclasses simply disappears within a synthetic approach. For example it is often suggested that pharmacologists would want a quite different classification of chemicals from that which chemists want. But in fact pharmacologists are interested in causal relationships such as (chemical) (reduce) (blood pressure) and their concerns are thus not just adequately but best addressed through a classification that stresses free combinations of things and relationships.

We might wish to classify art by purpose. But “art as propaganda” is a causal relationship, as indeed is “art (that) enhances understanding” or “art (that) evokes a better world.” We might seek to classify art by audience, distinguishing fine from popular art. Yet this boundary is unclear and shifting. It is much better to classify works which emphasize the audience or purpose of a work or works utilizing causal relationships. In particular, masterpieces are works of art thought to be particularly aesthetic. We might characterize differences in materials, but again these are best seen as relationships: (art)(made from)(wood). We might classify art primarily in terms of time and place it was developed. Such distinctions are readily captured by allowing free combination with classifications of time and place. In order to facilitate a web-of-relations approach, it is important that the same classification of time and place be used throughout the universal classification (which can be achieved with respect to time by focusing on decades and centuries rather than “The Golden Age” or “Middle ages” or “Renaissance”). This will facilitate connections being drawn between the art of a particular time and place and the culture, politics, economy, and technology (and so on) of the same time and place. Finally, we might distinguish art in terms of artistic styles. We will make the novel argument below that styles can be treated in a very similar manner to scientific theories.

We are left, then, with the classification of art in terms of “type of,” the standard avenue of developing subclasses (occasionally replaced by “parts of”) in the Basic Concepts Classification (Szostak 2013a). In the case of art, “type of” primarily captures what is often termed “medium”:

Non-reproducible art
1. Painting
2. Sculpture
3. Collage

Reproducible art
1. Prose
2. Poetry (rhyming or not)
3. Theater
4. Cartoon
5. Graphic art
6. Film
7. Music
8. Dance

The reproducible versus non-reproducible distinction is stressed here because it has a host of implications for how art is produced and appreciated. Non-reproducible art is by its nature unique whereas reproducible art can be copied or repeated (as with all classes the boundary can be fuzzy here, as with limited-edition prints). Though it is a logical distinction, it is one that reflects the importance of particular relationships: it was thus discovered inductively rather than deductively.

Several of these subclasses merit further subdivision. At this lower level we also seek subclasses in terms of medium. In the case of music, this would involve distinguishing singing from a variety of musical instruments. Lee (2011) notes that musical performers are most likely to search by medium. This may well be true for perform-
ers more generally. We would want a notation that allowed easy recognition of when two or three different types of music were combined (and also some notational shortcuts for indicating common combinations of many instruments such as orchestras). And we would want our subclasses to be free of cultural bias: instruments common in non-western music should not be disadvantaged. Poetry can be distinguished by rhyming scheme and rhythm (and again we want cultural fairness such that haiku is treated similarly to sonnet). Painting can be distinguished by: oil versus water, canvas versus fresco. Film likewise can be subdivided in several ways: silent, black and white, 3D, large screen, made-for-television, 360 degrees. Folklore can be distinguished as tales, legends, fables, and so on (La Barre and Tilley 2012). In all cases we want flexibility to allow new mediums to be recognized.

This classification does serve to distinguish art in terms of human senses: auditory, verbal, visual, mixed. It is not explicitly organized in terms of these. In part this is to not ground the classification unnecessarily in an approach where there is scope for dispute: some have claimed that sculpture is really an art of touch. Arguments about how various senses are applied in both production and appreciation of art are best captured through causal relationships.

The classification is flexible such that new subclasses can be added. It is fairly standard, except for incorporating literature within art (where it logically belongs). Architecture was included in previous versions of the BCC but is, admittedly, a misfit. The argument above would suggest that we treat buildings primarily as items of utility, and capture their aesthetic elements through causal linkages: Architecture is thus (aesthetics)(applied to)(buildings). And if we were to include architecture then we risk the slippery slope of demands to include pottery and other artifacts that mix utility and art.

It is also necessary to classify works of art by artist. Indeed much (or most) art scholarship focuses on individual artists (Ørom 2003). This is probably best done separately from the classification above, since artists often produce more than one type of art, and it is then annoying to have to search for them multiple times (Ørom 2003). But each artist should then be connected to the type(s) of art that they produce.

4.1 Causal Links in Humanities Scholarship

As noted above, most scholarly works, and likely most general works, investigate some causal link(s): how one or more of the things in the world affects one or more others. This is true of humanities scholarship as well. Existing classification systems unnecessarily privilege some causal links over others. There is thus tremendous value in moving toward a system that allows works to be freely classified in terms of any causal relationship.

It is fairly straightforward to imagine causal influences between art and every other category of human science scholarship (see Szostak 2000):

- The non-human environment provides both raw materials and inspiration. Art in turn affects both how we perceive nature and how we construct the built environment.
- Our genetic predisposition generates aesthetic universals (though literary theory at times suggests that these do not exist), our senses; and our ability to structure say music; many have posited that humans are evolutionarily selected for art because it teaches, raises the spirit, and creates social bonding (note that these are then links from art to yet other phenomena).
- Cultural elements and aesthetic sense are combined in most (or all) artworks. Thus works of art are generally seen to express certain cultural values or beliefs. Art is emphasized in all religions. Art may help us cope with cultural change.
- As for individual differences, artists likely display unusual personality traits and behaviors, and perhaps so do their audiences. And if art does impart meaning, then art changes what people believe and do. Art may be cathartic.
- Economic circumstances influence both the demand for art and supply of art. Art may also reflect and communicate economic ideology. As noted above, artistic sensibility increases the cost of all goods.
- Politics exerts various influences: funding, censorship, ideology, nationalism. The fact that many regimes have funded artistic propaganda suggests a belief that art can have important political influences.
- Social structure affects who becomes an artist and who supports the arts and thus likely the content of art. Art can support group solidarity. It can thus either encourage or level social distinctions.
- Technology influences the cost and quality of art. As noted above, art influences the design of almost everything we use.

The central importance of art to human existence is lost in a classification that does not allow the myriad influences on and of art to be readily captured. Yet in practice art scholarship has emphasized some links more than others. And classification systems grounded in literary warrant thus make it easier for the user to investigate some links than others. Notably Ørom (2003) speaks of the “bricolages” of today’s classification schemes. The classifications most in use today were developed over a century ago, and have developed slowly over the succeed-
ing decades. The result is that these classifications thus include elements of many schools of thought regarding the arts over the last centuries, but do not represent any one coherent view. Ørom (2003) argues that the new art history since the 1970s has been interdisciplinary in orientation and thus poorly served by pre-existing classifications. The challenge, then, is to take a new approach that encompasses these many ways of looking at art. A universal synthetic classification automatically meets this challenge, for art can then be linked to any other subject. A discipline-based classification can only achieve this equivalence in treatment of all possible causal linkages by reproducing all other classes within the classification of art.

4.2 Causal Links in Works of Art

In the preceding section we made a fairly modest recommendation for a classification that can encompass all of the myriad causal links that humanities scholarship might engage. In this section we make a more radical recommendation that the exact same principle should be applied to the works of art that humanists study.

It is useful to return to Ørom here. Notably, Ørom (2003, 134) references an increased importance of thematic study in art scholarship. To this end he looks at recent Danish art exhibitions: “In ‘Symbolism in Danish and European painting 1870-1910’ there are five themes: Beauty and Death, The Greatness of Man and Nature, Silence till Death, Eros and Melancholy, and The Prophets of Beauty. The painter’s nationality, the art form, and the date of the exhibited works are subordinated to the themes.” If scholars of art want to study such themes, and especially if art galleries wish to gather together works that exemplify such themes, then surely we should attempt to classify works in terms of the themes they express. Imagine how much easier it would be for scholars—or indeed anyone interested in art—to explore how artworks across different times and places expressed any particular theme if we took this simple step.

Rossett (2013) reports widespread interest in subject classification of works of art, but very limited progress. Financial constraints are one barrier, but even more important is the lack of a clear vision of how to proceed. Some galleries have experimented with social tagging as a result, but there are naturally worries both about the lack of controlled vocabulary and expertise. Rossett wonders whether experts would be willing to tag for free, and finds in a survey that there is some limited willingness to do so. Such an approach might address the financial challenges of classification, but as Shatford Layne (2002) suggests will most likely achieve a consensus classification if a shared and simple controlled vocabulary is employed.

Scatturo (2013) summarizes impressive efforts to classify the European Collected Library of Artistic Performance. Yet the approach to subjects is quite limited (p. 28): “The ‘Subject’ facet can be used to explore themes which are common to different collections. Its foci may include: artistic movements (e.g., Expressionism, Futurism), performing arts disciplines (e.g., history of theatre, physical training, voice, directing), specific issues (e.g., catharsis, jealousy, feminism, racism), and well-known characters (e.g., Medea, Hamlet, Oedipus, Nora). This will help to create monographs and virtual exhibitions, as well as helping teachers to gather the content needed for their courses. The user community may propose to the Content Board how this facet can be enriched.” This paper would suggest that these goals would be better achieved by allowing any subject to be captured synthetically.

Panofsky has identified three levels in the description of works of art. One level, description, simply describes the main elements (woman on horse). Another level, identification, gives specifics (name of woman). Users, of course, often search for particular people, places, or times. The third level, interpretation, records cultural significance (e.g., Christian parable). Baca et al. (2006) are far from alone in wondering if a classification can possibly capture all three levels. Shatford Layne (2002) surveyed art historians and found that 20 percent of art history research addresses what works are “about” (Panofsky’s third level) and 35 percent addresses what works are “of” (Panofsky’s first two levels). She concludes that art historians will benefit significantly from subject access of both types. She notes that scholars from a variety of other disciplines also would benefit from subject access to works of art, and wonders whether works should be classified in diverse ways to suit these diverse audiences. The sort of universal approach to classification recommended in this paper would spare us from such a complicated procedure.

Shatford Layne also recognizes that description and identification lie on a continuum; it is thus desirable to have a hierarchal classification such that a specific (Lady Godiva, Westminster Abbey) denotes also the general (woman, church); one need not then separately treat identification and description. It is also often the case that a work is of and about the same thing. But often also this is not the case: Shatford Layne provides several examples of paintings expressing Biblical themes but employing medieval settings. It is thus sometimes but not always valuable to distinguish “of” from “about.” A classificatory strategy that insisted on separate treatment of Panofsky’s three levels would thus be inefficient. A synthetic approach to classification allows us to specify a difference between “of” and “about” only as necessary. We can also employ identification to imply description as long as we place all specifics employed within general classes.
Harpring (2002, 21) notes that subject matter may be narrative as in “slaying the Nemean lion” or “capturing the wild boar of Mr. Erymanthus.” Her first example of non-narrative art is “young woman bathing.” The importance of verbs in each of these three examples is noteworthy. A concept chain that allows things and actions to be linked seems the best way to address narrative works (where the subject is a sentence fragment) and at least many non-narrative works.

Harpring urges extensive subject classification. For a painting described as “Nike crowning the victor, with the judge on the right and defeated opponent on left” she suggests a long list of subject terms: Nike, game, judge, competition, victory, games, prize, festival, and athlete. This list, notably, does not capture the combination at the heart of the painting; there is no mention of “crowning” at all, nor of defeat. A synthetic approach utilizing both (basic) things and relators would be: (Nike)(crowning)(victor)(of) (particular game)(beside)(judge)(and)(loser), where the particular game would be linked hierarchically to games in general. This synthetic entry is more compact than Harpring’s list but communicates much more of the nature of the work.

Shatford Layne had worried that there is often controversy regarding what a painting is “about.” Harpring urges us to use a word like “probably” in situations where there is controversy regarding the subject of a work. We could easily insert (probably) into a synthetic concept chain. But we should be cognizant of the fact that the theme attributed to a work may evolve through time: a religious work may be appreciated long after the religion has been eclipsed. Though no classificatory strategy can address this possibility perfectly it could be that a synthetic approach has the added advantage of allowing multiple interpretations to be recognized.

Of course, some works of art may be about a single thing (or perhaps a single relator) rather than some combination. If a work seems to be about “small girl” or “vineyard” or “training” so be it. But most works are better described in terms of combinations of basic concepts: (girl)(smiling) or (vineyard)(at)(sunrise). And many/most works of art will express some causal relationship: (girl)(smiling)(because)(gift).

If a synthetic approach is valuable, we should then ask what sort of concepts we wish to combine synthetically. The danger in classifying art is the same as the danger encountered above with respect to the scholarship of art: that we become captured by the themes thought to be important at some place and time. La Barre and Tilley (2012) discuss the (admittedly valuable) efforts to classify folktales in terms of a hundred or so themes. The leading classification of this type would classify “Beauty and the Beast” as falling under “Tales of Magic,” then “Super-natural or Enchanted Wife (Husband) or Other Relative,” and more precisely “The Girl as the Beast’s Wife.” Despite their widespread use, such “motif and tale type indices” are subject to frequent criticism. Critics often note that the classes are arbitrary. And it is felt that these indices are not updated regularly but need to be. There are also complaints that these particular indices are not themselves part of document classification schemes, and thus users must then track down cited works themselves. It would be much better to allow works of art (including fiction) to be freely classified in terms of dominant (and even subsidiary) themes within our documentation classification schemes. That is, we should be able to employ combinations of any concepts employed in the entire classification in order to indicate the theme of a work of art. It is noteworthy in this respect that the tale types and motifs employed in the classifications referenced by La Barre and Tilley are combinations of basic concepts.

We do try to some extent to classify the subject of art works within our classifications. For example, the Library of Congress Classification attempts a few precise classes for the subject of paintings: animals, birds, hunting, and fishing (which Orom 2003 argues represent a Renaissance sense of art subjects; for our purposes it is worth noting that these are each basic concepts or very simple combinations rather than very complex concepts). If we will attempt to classify art in terms of some arbitrary subjects, why not instead classify works of art in terms of any subjects that they seem to address?

It should be noted that controlled vocabularies such as the Art and Architecture Thesaurus naturally focus on the terminology of art itself. The AAT provides controlled vocabulary for artistic styles, materials, objects such as furniture, appearance (colors, for example), and artistic processes, but limited treatment of subject. These controlled vocabularies are of limited use in describing the subjects of works of art, unless these are works about art itself. Capturing the subject of a work requires access to a universal controlled vocabulary such as can (best) be provided by a universal classification. If this universal vocabulary takes the form of a universal classification then we dramatically increase the likelihood that different classifiers will apply the same classification to a particular work.

We noted above that humanities scholarship evolves, and can thus outgrow any classification grounded entirely in literary warrant. The same is true for art itself. Art is an inherently evolutionary endeavor, where artists build on what has gone before but try to create something new. This will mean among other things that artists will seek out new subjects (say, soup cans) that have not been treated before. We thus need to be able to classify works in terms of any subject.
One disadvantage of employing an ad hoc classification is that a user needs to identify precise classes of interest under which to search. If terminology is instead freely borrowed from a logically organized universal classification, then the user need not worry. If they search for (woman) (smiling) they will find instances of (girl) (smiling) if girl is in some way a logical subset of woman. (In the BCC, girl is achieved by combining woman and an age indicator.)

This universal classification is best grounded in basic concepts; only then can individuals and groups be expected to attach similar meanings to the terminology employed. Lee (2011) stresses that scholars, performers, and the public have different search needs and styles. This provides a further justification for allowing search in terms of combinations of basic concepts.

There is one classification used in classifying works of art that does strive to capture a broad range of subjects: ICONCLASS (2012). ICONCLASS is organized as a classification system: ten broad categories are each subdivided ten times, and then a further 25 times, and further as necessary. ICONCLASS attempts a broad coverage: the ten main classes address religion, literature, humanity, nature, history, society, and history. Very detailed identification is possible (the main webpage gives detailed notation for the biblical story of David and Bathsheba). Like many classifications, ICONCLASS provides some limited scope for synthesis: for example, the names of flowers can be inserted in brackets after the notation for flower, notational tricks allow nude men to be distinguished from nude women; and there are a handful of “keys” that allow symbolic animals to be distinguished from real animals or male from female. But as with most enumerated schemes there is no easy way to capture synthetic subjects. Many ICONCLASS classes are thus necessarily compounds themselves: 25L cities represented allegorically or symbolically; 25F8 extinct animals; 33A11 lifting one’s hat, baring one’s head; 71H713 Bathsheba (alone) with David’s letter. Many classes capture elements of a more general classification: 25F animals; 25F6 fishes; 25C geological phenomena; 23 Time. By pursuing a synthetic approach we can allow greater precision with shorter schedules. The developers of ICONCLASS have clearly striven to identify compounds found in many works, but can hardly capture compounds of importance to every work. By pursuing a universal classification, we achieve even greater precision in treatment of natural and social phenomena, while facilitating searches that span works of art and other objects.

It was straightforward to translate each class in ICONCLASS into the terminology of the BCC (see Szostak 2013b for the full translation). Some of the more specific subclasses in ICONCLASS, such as particular Biblical stories, would be handled in BCC by the use of Cutter numbers to indicate characters of a particular type from a particular source. So the BCC allows us to capture every subclass identified by ICONCLASS. Yet the BCC does so with a couple of key advantages. First, the classifier is not limited to the subclasses identified in ICONCLASS (which, notably, exhibit a strongly Western cultural bias, and seem better suited to historic than contemporary art). Second, the user need not master ICONCLASS but can input synthetic queries employing basic concepts. The classifier in turn is not limited by enumerated ICONCLASS classes but can classify very complex work by combining multiple basic concepts. The BCC, that is, is more universal than ICONCLASS, and more flexible because of its synthetic approach. At the same time, its reliance on basic concepts facilitates both classifier and user.

4.3 Relationships

Most of what needs to be said about relationships was captured in our discussion of causal relationships. But it deserves to be stressed that what is important about a work of art is often some relationship. If a painting is of a (woman) (riding) (horse), we will not be able to describe it very well if limited only to noun-like phenomena. The concept “riding” is essential to accurate classification and retrieval.

Humanities scholarship is likewise characterized by relationships. These are sometimes external relationships, as when art influences politics or is influenced by cultural values. They may also be internal relationships. As we have seen, art is an evolutionary process. Artists want to innovate, but start from what is. Mutations are selected culturally and thus we can usually point to a dominant style of any time and place, but one that necessarily allows change. Rising incomes and a mass market allow different styles to cohabit in the contemporary world. Much of art scholarship focuses on how a particular artist (or group of these) was influenced by other artists and/or the wider world, and how they in turn influenced other artists and/or the wider world. And this sort of scholarship will be hard to classify and hard to search if we do not classify different types of influence. Among the basic relators identified in the BCC (using a combination of deduction and induction, and drawing among other sources on the AIT) are transforming, energizing, combining, creating, facilitating, experiencing, performing, believing, evaluating, feeling, intending, rehearsing, perceiving, selecting from, thinking, cooperating, imitating, paying, and talking. These can be combined to generate hundreds of further relators.
4.4 Qualifiers

It was noted above that elements of the form and content of works of art might usefully be classified. It should be noted here that such an approach is of particular importance for works of abstract art. Some of these elements are best captured synthetically through relationships, such as links to classifications of shapes and sizes and colors, or to materials or instruments. But the Basic Concepts Classification (Szostak 2013a) possesses a class of adverbial/adjectival qualifiers that can be freely combined with any concept. Many of these may be particularly useful in classifying both art and humanities scholarship. They capture not only elements of form and content but also of subject matter and intent: beautiful, ethereal, polished, bright/dull, intense, sleek, sublime, thankful, superior/ inferior, successful, good, interesting, enjoyable, suitable, safe, simple, popular, necessary, effective, mechanical, strategic, informative, secretive, true, illusory, romantic, familiar, artificial, realistic, authentic, hard/soft, thick, clear, clean, complete, balanced, united, orderly, modern, radical, tidy, holy, and legitimate. This classification of qualities (of which the preceding is just a selection) has been developed inductively (but then organized into two dozen classes) and can readily be added to if warranted by humanities scholarship. It may be useful at times for scholars or other users to explore what range of items is classified using a particular qualifier. This is easiest and perhaps only possible if a universal synthetic approach is taken.

4.5 Perspective

Gnoli (2012) notes that information scientists have been talking about classifying works by authorial perspective for over a century. And he worries that, without a clear understanding of perspective, elements of perspective may be either ignored or conflated with subject (or type of work, an issue addressed below). Langridge (1989, 45-7) also notes that the failure to classify by what he calls “viewpoint” often interferes with subject classification: a book on the Christian approach to education may be miscategorized as on education about Christianity. Gnoli recognizes that classifying by perspective may be especially important in art. A poem, he appreciates, can communicate the same message—say, sadness—through many subjects.

Much but not all of what Gnoli surveys in terms of authorial perspective—theory, method, time, and place—are captured elsewhere in this paper. What is left? As I have argued elsewhere (Szostak 2014a), we wish with perspective to capture key motives and beliefs of the author or artist. Kleinberg (2013) also urges us to capture “why” in our classification, along with “who” and “what.” I suggested that a variety of dimensions (beyond those addressed else-where) might be useful in this respect: rhetorical, epistemological, ideological, aesthetic, ethical. These dimensions seem quite useful for classifying works of art:

- Some artists may be focused on communicating meaning of various sorts and in various ways.
- Art scholarship has at times suggested that art serves a revolutionary function, and at other times argued that art always supports the status quo. Both need to have a place in our classifications.
- Some artists may be focused on creating a certain type of aesthetic pleasure.
- Religion is the most obvious but hardly the only avenue through which an artist may aspire to encouraging particular values.

What about humanities scholarship? Clavier and Paganelli (2012) argue that we should classify all works by authorial stance: criticism, agreement, consensus, and so on. It would seem that it would be useful to distinguish different texts about art:

- Art criticism (which evaluates one or more works in terms of aesthetic standards).
- Connoisseurship (similar, but with a goal of identifying particularly valuable works).
- Contextual analysis, which analyses the influences on or of a work(s) without necessarily passing an aesthetic judgment.

This list is certainly not exhaustive but gives a flavor of what a classification by perspective might look like and accomplish. For works of scholarship it can also be valuable to identify the discipline of the author.

4.6 Theory

In classifying works of scholarship in terms of theory applied, it is necessary both to classify by the name of the theory and by theory type (a classification of theory types was developed in Szostak 2004 and applied there and in Gnoli and Szostak 2008). This is because diverse types of theory operate under the same name, and the same type of theory goes under many names. This approach will be particularly valuable in humanities scholarship which has tended to be characterized by an abundance of theories.

It is suggested here that the “theory” dimension be employed to capture “artistic style” when classifying works of art. The style pursued by an artist is at least somewhat analogous to the theory pursued by a scientist. Here again we should classify both in terms of style name and style type, and for the same reason: style names are not well defined and similar works of art may have...
quite different styles ascribed to them. It may be useful to employ Cutter numbers in designating style names, given the profusion of styles. In any case, we want a system that is hospitable, for new styles emerge with some frequency. We must not privilege western styles as existing classifications tend to do (Lee 2011), but allow any style found anywhere in the world to be readily represented.

Scholarly theory types were classified in terms of the who, what, where, when, and why questions (Szostak 2004). A similar approach would be useful in capturing key elements of artistic style:

- The “what?” question could capture degree of realism.
- The “why?” question would address issues of purpose—is the intent to inform, shock, energize, and so on.
- The “who?” question might capture intended audience.
- The “where?” question captured degree of generalizability when classifying theories and could capture here the degree to which a work expresses universal aesthetic value relative to particular cultural values.
- The “when?” question might capture the historical relations between one style and another.

As with perspective just above, work remains to be done to flesh out the details of the classification. The purpose here is to suggest both the desirability and feasibility of doing so.

4.7 Method

Paintings are often classified by technique (Ørom 2003). The practice is less common for some other art forms. But theatre depends on techniques of vocalization, sound and lighting systems, and techniques for building and moving sets. In the world of film the set of techniques is even greater. Music employs not only techniques for amplification and recording but increasingly techniques for artificially creating musical sounds. Scholars, performers (perhaps especially directors and producers), and members of the public may wish to search by technique. And for works of art we may need to differentiate the time and place that a work is set from the time and place it was produced. For literature especially the time and place in which a work is set is of great interest to users; again it is useful for setting to be captured with respect to a universal categorization of time and place. It will, though, be necessary to develop some classification of imaginary times and places for works set in these.

One advantage of treating time and place systematically is that it facilitates the use of the same classification system in archives and museums as in libraries. Archives especially tend to prioritize classification by time and place (and of course source). This need not prevent them from classifying also in terms of different classifications of time and place produced. This can easily be done through recourse to classifications of time and place. There is no good reason for different types of art to be classified in terms of different classifications of time and place. Indeed, it will be easier to draw connections between art, politics, economy, and culture if the same categorizations of time and place are used throughout a universal classification.

For reproducible works of art, we may need to classify with respect to multiple times and places: a work may be performed in a quite different time and place from where it was produced. Even for non-reproducible works we may wish to note where and when it was owned and exhibited. And for all works of art we may need to differentiate the time and place that a work was set from the time and place it was produced. For literature especially the time and place in which a work is set is of great interest to users; again it is useful for setting to be captured with respect to a universal categorization of time and place. It will, though, be necessary to develop some classification of imaginary times and places for works set in these.

4.9 The Nature of a Work

Smiraglia (2001) has carefully examined the nature of a work, and in particular when a work has changed enough to deserve designation as a new work. He stresses throughout that a work is primarily defined by the ideas that it conveys. I have suggested that “ideas” comprise some set of: descriptions of phenomena or relationships, causal arguments, theories applied, methods applied, and perspectives applied (Szostak 2015). Both Smiraglia and I appreciate that the subject matter of any work cannot be particular. It might be important to ensure that works on these would be found by all interested users.

4.8 Time and Place

It goes without saying that works of art should be classified also in terms of time and place produced. This can easily be done through recourse to classifications of time and place. There is no good reason for different types of art to be classified in terms of different classifications of time and place. Indeed, it will be easier to draw connections between art, politics, economy, and culture if the same categorizations of time and place are used throughout a universal classification.
separated from its semantic content. Smiraglia was focused on written works for the most part but his analysis applies even more forcefully to works of art: these must be defined (and classified) in terms of both subject matter and what might be termed style. We have striven to capture both elements above.

When does a work change enough to be classed as a new work? When there is a significant change in either style or substance. Smiraglia appreciates that changes in media always generate a new work: a film based on a novel or play is a new work no matter how strictly it adheres to the original. Likewise a print made from a painting is a new work. And a musical score is a different work from a recording or a textual description of a piece of music. For performances, the engagement of a new performer (at least in a key role) generates a new work. This is especially important for genres such as jazz music in which a performer has considerable scope to reinterpret a piece of music. But the argument holds more broadly. Yet existing classification systems that stress the composer or playwright often leave little scope for appreciating changes in performer.

Even written texts can be changed through time. In their study of a random sample of folktales, La Barre and Tilley (2012, 697) found that “illustrators were identified more than 80% of the time, although the original author was identified just over 40% of the time. Records of stories that were retold indicated an agent in only one third of the instances, and editors or compilers were acknowledged in only one quarter of the records. Translators were identified only 10% of the time, and adapters less than 5% of the time.” Yet La Barre and Tilley found that both scholars and storytellers were very interested in this sort of information.

To summarize, it is critically important to distinguish different “editions” of a work of art. Since these can be identified in terms of causal links, artistic theory (style), method, and perspective, we will be better able to identify new works if each work is classified along these myriad dimensions.

4.10 A Brief Note on Terminology

Ørom (2003) notes that humanities scholarship is characterized by the development of new terms that need to be reflected in a classification. Though Ørom is not specific, new terminology can be used in most of the areas discussed above: phenomena, relationships, theory, and method most obviously. The approach recommended here treats all complex concepts as combinations of basic concepts. It thus alleviates the need to expand the classification when a new term is coined. The challenge for the classifier is that the precise meaning of a new term may be unclear. Scholars should, of course, try to clarify their terminology, but often do not. The classificationist should respect literary warrant but not engage in unnecessary contortions in order to reflect unnecessarily ambiguous scholarship. The classificationist should thus provide an exhaustive set of basic concepts (which can be achieved in a very manageable set of schedules, at least in the social sciences and humanities), and the classifier (or indeed the author) should render complex concepts in terms of these. This strategy has been pursued throughout this paper. It will allow multiple types of user to better find both works of humanities scholarship and works of art that are relevant.

5.0 Conclusion

Only a universal classification free of disciplinary boundaries can allow for the inherently interdisciplinary nature of both works of art and scholarly examinations of these. Discipline-based classifications have privileged some relationships over others. A universal approach also facilitates the access to art (and art scholarship) of scholars from outside the humanities. Utilizing the same classification of time and space throughout a universal classification facilitates connections between the art of a particular time and place and all other aspects of life in that time and place.

A synthetic approach allows diverse elements to be combined in describing both art and scholarship of art. This again is critical in allowing any combination pursued by artist or scholar to be identified. A synthetic approach allows us to focus our classification of art on medium (“type of” subdivision), while capturing audience, purpose, material, and time/place synthetically. A structured synthetic approach that mimics (usually causal) sentence structure allows works—of both art and scholarship—to be precisely characterized. The combination of a synthetic approach with a universal classification instantiates a web-of-relations in which users can easily follow their curiosity from works on the aesthetic nature of an artifact to works on its technical or economic aspects—or indeed to similar aesthetic characteristics found in different works. The particular synthetic approach urged in this paper—which links things, relators, and properties in a sentence-like structure—is particularly valuable for it is the combination of these that usually signals the essence of a work of art or indeed a work of scholarship.

Synthesizing basic concepts—those for which there is broadly shared understanding across individuals and groups—facilitates both classification and search. Classifiers can be precise by combining multiple basic concepts. Users can then search also by combining basic concepts.
This may be especially important in humanities scholarship where new complex terminology is generated with some frequency.

Classifying theory types is important for all scholarship, but arguably especially so in the humanities. The approach taken to classifying theory types in scholarship can, with some adjustment, be applied to the classification of styles in works of art. Classifying both scholarly and artistic methods is also important. For scholarship these are a particularly important signal of relevance. We can aspire to capture in the “perspective” dimension rhetorical, aesthetic, ideological, ethical, and epistemological characteristics of both art and scholarship. This will aid users in assessing relevance.

Classification in each of the four ways discussed above has a final advantage in aiding us in identifying when a work has changed enough to be considered a new work. These four characteristics of a classification can each be found in the Basic Concepts Classification (Szostak 2013a). Most can also be found in the Integrative Levels Classification (2014). It would be possible to adapt at least some of these to other classifications.

The next step in this research will focus more narrowly on subject analysis. A more detailed comparison with extant approaches to classification can then be provided. It is also possible to provide detailed synthetic classifications of a sample of works from various leading galleries and museums (Szostak 2014a, b).

References


Broughton, Vanda. 2006. The need for a faceted classification as the basis of all methods of information retrieval. Aslib proceedings 58: 49-72.


Shatford Layne, Sara. 2002. Subject access to art images. In Baca, Murtha, ed., Introduction to art image access: issues,

Szostak, Rick. 2014b. The Basic Concepts Classification and the archipelago of cultural memory. Unpublished paper
The Network Pattern of Journal Knowledge Transfer in Library and Information Science in China

Rongying Zhao* and Shengnan Wu**

School of Information Management, Wuhan University, Luojiashan, Wuchang, Wuhan, Hubei Province, P.R.China, 430072

*<zhaorongying@126.com>, **<vivian19870220@163.com>

Rongying Zhao has a Ph.D. in information science (2006) and is a professor at the School of Information Management, Wuhan University, Hubei, P.R.China. She teaches knowledge management at the School of Information Management of Wuhan. Her research interests are knowledge management, informetrics and scientific evaluation.

Shengnan Wu is a Ph.D. student at the School of Information Management, Wuhan University, Hubei, P.R.China. She has a bachelor's degree in information management, and has been taking successive postgraduate and doctoral programs of study for doctoral degree since September, 2012. Her research interests are knowledge organization, domain analysis, knowledge management, informetrics and altmetrics.


Abstract: Using the library and information science journals 2003-2012 in Nanjing University’s Chinese Social Sciences Citation Index as data sources, the paper reveals the citation structure implied in these journals by applying social network analysis. Results show that, first, journal knowledge transfer activity in library and information science is frequent, and both the level of knowledge and discipline integration as well as the knowledge gap influenced knowledge transfer activity. According to the out-degree and in-degree, journals can be divided into three kinds. Second, based on professional bias and citation frequency, the knowledge transfer network can be divided into four blocks. With the change of discipline capacity and knowledge gap among journals, the “core-periphery” structure of the knowledge transfer network is getting weaker. Finally, regions of the knowledge transfer network evolved from a “weak-weak” subgroup to a “strong-weak” subgroup or a “weak-strong” subgroup, and then move to a “strong-strong” subgroup.

Received: 26 March 2014; Revised: 21 June 2014; Accepted 28 June 2014

Keywords: knowledge transfer network, journal citation, library and information science, informetrics, China

1.0 Introduction

In recent years, studies in library and information science (LIS) in China have received much attention in theoretical research and practical applications, and a lot of achievements have been scored. It is necessary for us to grasp the current status of LIS in China and its development trends. Domain analysis (DA) as a mature quantitative research method has been applied to many disciplines at home and abroad. As a functionalist approach, it attempted to understand the implicit and explicit functions of information and communication, and to trace the mechanisms underlying informational behavior from this insight (Hjörland and Albrechtsen 1995). At present, research advances in domain analysis come from many researchers (Smiraglia 2012; Tennis 2003). In the perception of these researchers, the concept of domain refers to a field of knowledge, activity, or interest in which certain knowledge is marked with defined limits and whose professionals or groups are entangled both in thought and in language, generating differentiated thought and interpretations (Freitas et al. 2012). The objects of domain analysis not only involve the entire science but also a specific discipline. For example, López-Huertas and Jiménez
Contreras (2004) conducted research to give an initial appraisal of research activity in Spain surrounding “knowledge organization” from 1992-2001 by the method of DA. And, more remarkable, De la Moneda Corrochano, López-Huertas, and Jiménez-Contreras (2013) conducted a similar study to analyze Spanish research in knowledge organization from 2002 to 2010 with the method of DA.

When assessing scientific output in a special field, the most usual approach is to collect the literature. Literature is an essential source of information in scientific research. The reason why human knowledge can be passed on for generations mainly lies in the citation practices in scientific literature, where old literature transfers intelligence into new literature. The references or citations between scientific literatures not only reveal the accumulation, continuity and succession of scientific knowledge but also indicate the presence of knowledge transfer. Thus it is feasible and meaningful to make use of the citation relationship to study knowledge transfer. In scientific research, scholars tend to take academic journals as the first choice to publish their newfound knowledge, but also indicate the presence of knowledge transfer. Therefore we believe the activity of knowledge transfer is clearly defined; rather, it stands as a broad concept. Moreover, the purpose of knowledge organization is to accelerate knowledge communication as well as knowledge innovation, and knowledge transfer is a good way to achieve knowledge communication and knowledge innovation. This kind of study can help to accelerate knowledge communication and knowledge organization as well as the development of library and information science.

The study of the history of the citation relationship can be traced back to the occurrence of concepts like citation identity and citation image. After White (1998, 2001) proposed citation image and citation identity, many scholars became committed to similar research. Bonnevie-Nebelong and Frandsen (2006) introduced the concepts of citation image and citation identity to the process of journal assessment and analysis and then put forward the notions of journal citation image and journal citation identity. Since then, related research studying journals has sprung up, but it has been found that most studies still concentrate on citation itself, such as citation indicators, citation characteristics, and citation styles (Bonnevie-Nebelong 2006; Fangli 2013) but pay little attention to the knowledge transfer and knowledge communication behind the journal-citing activity. In fact, citation image and citation identity can be regarded as two aspects of the citation relationship, only when the two are combined can we see the panorama of citation and further understand the knowledge transfer process and the growing and maturing of a subject revealed in a citation network. By building journal citation networks at different times and applying social network analysis (SNA), we investigate the changes in the centrality and the roles of citation networks of journals in library and information science in China at different times to dig deeper into the knowledge transfer pattern and traits of this field. We also utilize the associations generated by journal citation networks to calculate the knowledge transfer ratio of journals in library and information science in China to professionally similar journals and draw corresponding knowledge maps. We then summarize the pattern of journal knowledge transfer in library and information science in China in the hope of providing a basis for the formulation of corresponding knowledge transfer policy in this field.

In this article we present a systematic and visual domain analysis of the LIS field in China in terms of citation networks to explore the knowledge transfer evolution of this knowledge domain over time. The objectives of this study are:

- to examine the character of journals involved in the field of LIS in China in terms of knowledge input and output;
- to study the block traits of journal knowledge transfer networks for identifying the knowledge transfer evolution of the field of LIS in China; and,
- to detect the journal knowledge transfer model of the LIS field in China.

2.0 Data and methods

2.1 Data

At present there are approximately 70 journals in library and information science in China, and there is no consensus...
sus on which journals are the core periodical group in this field. The relatively authoritative journals are what have been included in Nanjing University's Chinese Social Sciences Citation Index (CSSCI), and these are also the criteria used by most colleges and scientific research institutions to evaluate talent. So based on CSSCI, we chose 16 journals as the objects of study (see Table 1, we use the abbreviated journal titles in the tables and figures that follow). Meanwhile, we take the reference database in CSSCI as a data source to retrieve the citation data of these 16 journals. To obtain a better grasp on the knowledge transfer condition in library and information science, we set a relatively long study period (2003-2012) out of the consideration that a long study period would reveal more accurately the variation of citation patterns in library and information science, from which we can find the traits of knowledge transfer. At the same time, to eliminate the impact of negative factors, our research divides the data into two phases: 2003-2007 and 2008-2012, so the five-year data can more accurately show the citation relationships of journals.

The initial data set from the database is more than 25000, and the research employs our self-made VBA program to calculate how many times the journals are cross-cited and construct a journal cross-citation matrix. More details are shown in Figures and 2. The rows in the matrix are the citing journals, and the columns are the cited journals. The establishment of the journals' citation matrix is also an embodiment of mutual knowledge transfer to some extent. When journal cross-citation is drawn in a network chart, a directed weight map forms. In this map, network nodes represent journals; sides represent the citation and reference relationship between the journals; the weight of each side represents the citation and reference frequency; the direction of the sides points from citing journals to the cited journals. In the figure, citation relationship is basically from citing journals to the cited journals, while the knowledge flow is just the opposite. However, the differences of citation frequency determine the differences of the knowledge transfer level.

The more times papers are cited, the more knowledge flows out, and vice versa.

2.2 Research methodology

Social network analysis (SNA) is a set of theories and methods which can analyze the structure and properties of various relations in the social network. It mainly analyzes the relation schema among actors and this method has been proved to be valid when applied to research on citation relationships (Yuanyuan and Qinghua 2008). We use SNA to systematically analyze the citation network of journals in library and information science in China; this mainly involves two levels, the properties of the whole network and the roles and locations of network nodes. This process would be conducted using UCINET’s network analysis tools. Specifically, the process consists of the aspects described below.

2.2.1 Centrality analysis

Centrality measures power in the network. We can measure with the indexes of centrality and central potential. Centrality measures the extent to which members are in the network's center; the degree of centrality is the most commonly used and in the topology network it represents the number of members who are directly connected to the chosen member while in the weighted network it represents the sum of edge weights directly connected to this member. Central potential measures the centralized degree of the whole network; the higher the central potential is the more concentrated the internet connections are on fewer nodes (Fengchao and Rongking 2013). Because we are investigating a directed network, both centrality and central potential can be divided into out-degree and in-degree.
2.2.2 Block model analysis

The notion of location is an essential notion which discusses the analysis of network structure. It mainly describes the location of nodes in the network and the interaction they have with other nodes. Block model analysis places similar nodes in a block according to equivalence in structure in order to simplify the complicated relationship between network nodes into the relationship between blocks (Liu 2009). Block model analysis usually consists of two-sided information: a) the nodes in the network which are divided into specific subsets are the locations; b) for each pair of locations, the existence or vacancy of links within or between locations reveals the relationship. Block models can simplify the complicated networks and then explain the network structure (Lin 2009). We can use the CONCOR program in the software UNINET to analyze the block model and the analysis result is the density of block models. We can further simplify the results of the block model into a matrix so as to briefly describe the block features of network structure.

3.0 Results and Analysis

3.1 The characteristics of journals

We use the UCINET software to calculate the centrality of the journal citation network in library and information science in China during the periods of 2003-2007 and 2008-2012. The exact results are shown in Figures 1 and 2.

In Figures 1 and 2, in-degree represents knowledge input, and out-degree represents knowledge output. From the perspective of knowledge output, Journal of Library Science in China has always been in the leading position and reveals the professional strength of Chinese academic journals of library science. As a significant knowledge source in the field of library and information science, Journal of Library Science in China has provided crucial technical support for the evolution of library and information science and development of other journals. Comparing Figure 1 with Figure 2, we can find that Library and Information Service gradually dominates knowledge output because of its huge number of published papers but also its strong ability to absorb knowledge. Library and Information Service is both an essential knowledge output and vital for knowledge absorption in the field of library and information science, thus it is the core and the bond in the whole knowledge transfer process of library and information science. We can also conclude that each journal’s different numerical values of out-degree and in-degree give rise to the knowledge potential difference. By the dimension of knowledge potential difference, we can divide the journals in library and information science into three categories: knowledge output-based journals, knowledge-sucking journals and knowledge balanced journals. Journal of Library Science, Journal of the China Society for Scientific and Technical Information, Journal of Academic Libraries, and New Technology of Library and Information Service are all knowledge output-based journals; they output more knowledge than they input. This demonstrates their authority in the field of library and information science. As knowledge output-based journals, they usually publish the most cutting-edge or up-to-date articles in aspects such as research contents and research methods. In general, high-level research achievements will have more opportunities to be cited. However, Information Science, Journal of Information, Information and Documentation Service, Library Work and Study, and Library Tribune have maintained their positions as knowledge absorbers during the period 2003-2012. As knowledge-sucking journals, they publish research results, which take the research achievements in knowledge output-based journals much further and have relatively narrower intellectual vanity and better relative independence (closure); so it also means less impact on other journals. Document, Information & Knowledge, Library Journal, Library and Information Service, and Library & Information are viewed as equation-type journals for the positive and negative value of their potential difference changes little. Apart from Library Journal, Document, Information & Knowledge, Library and Information Service, and Library & Information involve both the disciplines of library science and informatics judging from their names. Knowledge in the field of library and information science realizes their intersection and mutual complementation and referencing in the subgroups of these journals. It is their positive interactions that lead other journals to advocate this field’s development. There also exist some journals whose types vary a lot, such as Library Theory and Practice, Library and Library and Information Studies: Theory & Application. Library Theory and Practice changes from a knowledge-sucking journal into a knowledge output-based journal; Library turns from a knowledge balanced journal into a knowledge-sucking journal; Library and Information Studies: Theory & Application varies from a knowledge output-based journal into a knowledge-sucking journal. After analysis, we find that it is perhaps related to the forwardness, creativity and disciplinary expansibility of the published articles. These journals should not only keep their advantages of academic styles and traits, but also publish more articles that are forward-looking and that have discipline integration in order to expand their influence.

To sum up, the gap between knowledge strength and discipline integration causes the relatively big difference in each journal’s status in the knowledge transfer network. Journals with strong knowledge strength and discipline integration output knowledge constantly and absorb
external knowledge to maintain their own strength. But journals with weak knowledge strength and discipline integration mainly gain resources by absorbing knowledge; because the field of articles they publish is narrow and not forward-looking, these journals do not have much progress in their knowledge output ability.

3.2 The blocks’ traits in the knowledge transfer network

When establishing the cross-citation matrix in journals, we will find the relationship between knowledge output and knowledge absorption in journals of library and information science. We use the CONCOR procedure in
UCINET to produce the partitioning results of the journal cross-citation knowledge transfer network during the periods of 2003-2007 and 2008-2012. Considering that we are studying the structural equivalence of citing and cited activity between journals, we take no account of journals' self-citation, which means that there is no value of diagonal lines in the calculating process. The results are shown in Figures 3 and 4; Tables 2 and 3 are respectively the patch density matrix in the periods of 2003-2007 and 2008-2012 (because we are studying weighted networks, the density is more than 1.)
According to Figures 3 and 4, we find that the knowledge transfer networks of cross-citation in journals during 2003-2006 and 2008-2012 are both divided into 4 blocks. Based on the relationship between blocks, we can divide the blocks into a core block, a strong edge block, a weak edge block and a periphery block (Gao X et al. 2011).

According to Figures 3 and 4, the networks of journal cross-citation knowledge transfer during periods of 2003-2007 and 2008-2012 are clustered based on both professional bias and citation frequency. From the aspect of professional bias, we divide the journals into library science journals and informatics journals, and then further divide them based on citation condition, although there also exist exceptional cases. For instance, judging from their names, Information and Documentation Services and Journal of Library Science in China in Figure 3 should belong to informatics and library science respectively, the partitioning results are just the opposite; this phenomenon is principally due to the professional bias of articles in the two journals.

Combining the four tables, our research points to further analysis. From 2003 to 2007, the first block including Journal of Library Science in China and Journal of Academic Libraries belongs to the core block of the network. It outputs knowledge to the second and third block of informatics and the fourth block of library science, while it sucks knowledge merely through its interior communication. From this we can see its core status. From 2003 to 2007, informatics had just sprung up and was not mature, so it was eager to refer to its brother disciplines and suck knowledge to develop itself. It well explains why journals in block 1 transfer large sums of knowledge to the journals in blocks 2 and 3. Block 2 consists of Journal of The China Society for Scientific and Technical Information, New Technology of Library and Information Service and other journals of informatics. Block 2 is not only closely tied to journals in block 1, but also transfers knowledge to informatics journals in the third block. So the second block is a strong-edge block. Journal of Information and Information and Documentation Services are in block 3, and these journals mainly suck knowledge from others and have tight connections with the core block and the strong edge block; these journals belong to a weak-edge block. The fourth block contains many library science journals such as Library Tribune and Library Journal, and this block only absorbs knowledge of its own discipline and never communicates with informatics journals. So block 4 belongs to a periphery block.

From 2008 to 2012, the value of $R^2$ of the block model for the journal citation knowledge transfer network increased to 0.462 from 0.418; the number of nodes which have similar structure in the network decreased slightly and the blocks’ traits became more and more obvious. During that period, Journal of Library Science in China, Journal of The China Society for Scientific and Technical Information, and New Technology of Library and Information Service were assigned to the same block. It outputs knowledge to the second block, to which the journals of informatics belong, and the third block where the journals of library science are, respectively. Thus, it becomes the core block of the knowledge transfer network at this time. However, the second block includes a lot of journals of informatics. It not only maintains close relationship with the core block, but also outputs knowledge to the third block where the journals of library science are. Thus, it is called a strong-edge block. Going further, the journals in the two blocks all belong to informatics, excluding Journal of Library Science in China. This indi-

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>153,000</td>
<td>83,600</td>
<td>47,000</td>
<td>62,071</td>
</tr>
<tr>
<td>2</td>
<td>226,200</td>
<td>224,600</td>
<td>135,900</td>
<td>67,886</td>
</tr>
<tr>
<td>3</td>
<td>345,000</td>
<td>346,600</td>
<td>183,500</td>
<td>129,357</td>
</tr>
<tr>
<td>4</td>
<td>286,429</td>
<td>119,343</td>
<td>95,857</td>
<td>131,762</td>
</tr>
<tr>
<td>$R^2$=0.418</td>
<td>density = 144.8958</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The patch density matrix from 2003 to 2007

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88.833</td>
<td>77.733</td>
<td>27.833</td>
<td>36.583</td>
</tr>
<tr>
<td>2</td>
<td>273.133</td>
<td>334.050</td>
<td>108.300</td>
<td>129.400</td>
</tr>
<tr>
<td>3</td>
<td>148.333</td>
<td>186.900</td>
<td>169.833</td>
<td>186.750</td>
</tr>
<tr>
<td>4</td>
<td>82.167</td>
<td>79.600</td>
<td>71.250</td>
<td>84.000</td>
</tr>
<tr>
<td>$R^2$=0.462</td>
<td>density = 138.6417</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. The patch density matrix from 2008 to 2012

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4. The matrix of the cross-citation journal knowledge transfer network from 2003 to 2007

<table>
<thead>
<tr>
<th>Block</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5. The matrix of the cross-citation journal knowledge transfer network from 2008 to 2012
cates that the journals of informatics act as the knowledge source of library science and informatics. From 2008 to 2012, the development of science and the progress in computer technology provide an opportunity for the study of informatics; informatics becomes mature as a science gradually, getting rid of the restraints of the library science knowledge and developing into a subject which needs strong theory and practice. But after long term development of library science, its research approaches a bottleneck. So it needs to refer to the related theories, methods, and technology in informatics urgently to make a breakthrough. Thus, it heavily cites related studies of informatics. But as the leading journal in the field of library science, Journal of Library Science in China also follows the needs of the subject’s development and carries a lot of articles which have a strong fusion of disciplines; its professional bias inclines to informatics science gradually. That’s why it is assigned to the same block with the other two journals of informatics. Journals in the third block not only have a close relationship with core block journals and strong-edge block journals, but also absorb substantial knowledge in the fourth block. Knowledge exchanging activities are frequent. Thus, it is called a weak-edge block. However, the fourth block where a lot of journals of library science are, just transfers knowledge from the weak-edge block. So it is isolated gradually and belongs to the periphery block.

According to the matrix of the journal cross-citation knowledge transfer network during the periods of 2003-2007 and 2008-2012, we can draw the related simplified diagram which is shown in Figure 5.

From Figure 5, we find that from 2003 to 2007, the blocks’ traits in the journal cross-citation knowledge transfer network are obvious and the core-periphery structure whose core is the first block is formed. Further analysis reveals that journals in block 1 and block 4 are library science journals, while those in block 2 and block 3 are informatics journals. During this period, library science has more professional strength and develops quickly. The professional strength of informatics is comparatively relatively weaker. In order to develop and grow, knowledge transfer and knowledge communication within informatics can no longer satisfy the need. So they seek large quantities of cutting-edge and essential knowledge in library science journals. For specific performance, informatics journals in block 2 and block 3 absorb knowledge in block 1, where most authoritative journals in library science lie. At this time, block 1 can be viewed as the knowledge source of the whole knowledge transfer network and effectively annexes the knowledge authority status in the network. According to the simplified block figure of the knowledge transfer network during the period of 2008 to 2012, the status and roles of blocks 2 and 1 gradually converge, and the traits of the core-periphery structure of the journals’ network transfer network wear off. During this time, the development of technology, especially the growth of computer technology, offers further opportunities for informatics to advance. The professional power of the subject strengthens. On the contrary, library science comes across adversity and traditional research finds it hard to meet the requirements of a new epoch. At this moment, it is imperative for it to re-

![Figure 5. The simplified block figure of library and information science journals’ knowledge transfer network during the periods of 2003-2007 and 2008-2012](image-url)
ference the advanced technologies and research methods in informatics in order to make a breakthrough. Thus blocks 1 and 2 where informatics journals lie can be regarded as the knowledge source of the entire knowledge transfer network. In sum, with the alternations of the professional strengths of the disciplines, the knowledge gap and the interactive model among journals changes with it. Thus the hierarchical structure of the knowledge transfer network of cross-citation in library and information science journals evolves towards a fuzzy core structure and a distinct periphery structure.

4.0 Discussion: the analysis of the pattern of journal knowledge transfer in library and information science

According to the structure of the network of journal knowledge transfer in library and information science, the knowledge power and discipline integration force of a journal determine their intensity of knowledge supply and demand and further determine their roles in the network; the gap between the professional powers of the journals’ subjects decides their choice of the objects of knowledge supply and demand, and further decides the interactive model of knowledge transfer. According to the block traits of the network of journal knowledge transfer in library and information science, we find that the knowledge transfer network of the two periods of 2003-2007 and 2008-2012 can both be divided into four blocks, and these four blocks are all divided by the two dimensions of professional bias and citation frequency. To be specific, journals in two blocks belong to library science and journals in the other two blocks belong to informatics. At the same time, the journals in the same discipline are divided into two blocks by reference conditions. Thus the relationships between blocks are of two kinds, which are respectively the professionally close block and professionally distant block. We set blocks as the X-axis and the power of professionally close blocks as the Y-axis; knowledge power increases with the direction of the arrow. By recognizing the traits of knowledge input and output among the blocks, we construct the pattern of journal knowledge transfer in library and information science. The detailed process is shown in Figure 6.

Figure 6. The pattern of journal cross-citation knowledge transfer in library and information science
According to Figure 4, when the power of journals in a certain block and its professionally close block are comparatively weak, that is to say, the power of the discipline where the two blocks’ journals lie is weak, their knowledge gap is narrow; thus they can be viewed as weak-weak subgroups. In this situation, the blocks have to take advantage of the convenience of their close disciplines and output knowledge to the professionally close blocks with a platform of citing activity. In order to maintain their own power, this process will merely be realized by knowledge transfer within the blocks. At the same time, for their own development as well as that of the whole discipline, journals in this block have to overcome professional restrictions and cite articles from professionally distant journals to absorb knowledge needed from professionally distant blocks whose professional power is stronger in order to promote their knowledge power.

During the period of 2003 to 2007, informatics journals reveal this condition and block 2 is a typical example. According to Figure 3, it can be seen clearly that block 2 is outputting knowledge to block 3, which is professionally close. While the self-communication in block 2 is frequent and it sucks knowledge from the professionally distant block, block 1, whose disciplinary power is stronger in order to promote its own power.

As time goes on, if the knowledge transfer activity and methods above promote the development of journals in the block, while the power of professionally close blocks has little promotion and remains relatively weak, the gap between the block and its professionally close block further widens. Thus this can be viewed as a strong-weak subgroup. At this time, the block improves itself by interior knowledge transfer activity and outputs large quantities of knowledge to professionally close blocks. This shows that journals in the block develop well and have already formed superior academic strengths and academic styles; they have relatively independent knowledge transfer patterns. In this condition, it has already overcome its professional restrictions and the professionally distant blocks make full use of citations to make knowledge connections with it. The power of the discipline the journals lie in is the main factor affecting the pattern of knowledge transfer. From 2003 to 2007, journals in block 1 belong to this condition. The technical power of block 1 far surpasses that of its professionally close block, block 4; block 1 primarily communicates and transfers knowledge with professionally distant blocks. This is specifically expressed as the phenomenon that journals in block 2 and block 3 cite mass information in block 1. So we can deduce that block 1 is in the knowledge authority status in the journal cross-citation knowledge transfer network.

If the knowledge power of a block professionally close to a certain other block rises fast while its own knowledge strength is still in a bad situation, that is to say, the knowledge potential difference is big, these blocks make up the weak-strong subgroup. In this condition, the block will make full use of the advantage of close professions and suck knowledge from professionally close blocks by citation. However, the knowledge supplied by professionally close blocks is limited in the weak-strong subgroup and its needs cannot be fully met. So the block would absorb mass knowledge from professionally distant blocks in order to satisfy its needs and gain improvements at the same time. In this case, knowledge demand and the discipline’s development are the main factors affecting knowledge transfer. The blocks where journals of library science lie belong to this subgroup.

If the power of the discipline where the journals of the block lie strengthens a lot, the knowledge power of the block and that of its professionally close block will also increase. So the knowledge potential difference narrows and these blocks form the strong-strong subgroup. At this time, the block will use the advantage of close professions to transfer knowledge to professionally distant blocks by discipline superiority. Different from the strong-weak subgroup, where a certain block occupies knowledge authority status, the roles and status of a certain block and its professionally close block in the strong-strong subgroup gradually converge, so cohesion subgroups of knowledge transfer who have favorable interaction form. Informatics journals in blocks 2 and 3 during the period of 2008 to 2012 belong to this kind of subgroup.

5.0 Conclusion

We used the cross-citation data of journals in library and information science in CSSCI to analyze the network structure and pattern of knowledge transfer network in library and information science based on journal cross-citation. And we reach the following conclusions.

From the aspect of the structure of the journal cross-citation knowledge transfer network in library and information science in China, we can see that journals’ knowledge transfer activity is frequent. The knowledge strength, the force of discipline integration and the knowledge gap between journals all affect the interactive model of knowledge transfer. From the out-degree and in-degree of the nodes in the network, journals in library and information science can be divided into knowledge output-based journals, knowledge sucking journals, and knowledge balanced journals. Knowledge output-based journals usually publish plentiful articles with hot topics by strong pioneers so the chances for them to be cited are abundant and their knowledge power and discipline integration are superior. Articles in knowledge-sucking journals are less specialized and have little impact on
other journals. Knowledge-balanced journals are always the junctions of knowledge; they realize the supplement and reference of discipline knowledge and their active interaction promotes the rapid development of disciplines. Comparing the data from 2003 to 2007 and that from 2008 to 2012, we find that these types of journals remain relatively stable though small variations exist.

According to the blocks’ traits in the journal cross-citation knowledge transfer network in library and information science in China, the websites during the periods of 2003-2007 and 2008-2012 are all divided into four blocks by professional bias and citations, which are respectively the core block, the strong-edge block, the weak-edge block and the periphery block. From 2003 to 2007, the power of library science was strong, so the block where its authoritative journals lie was cited a lot and effectively annexed the status of knowledge authority. During this period, the network of journal cross-citation knowledge transfer in library and information science in China reveals the core-periphery structure. From 2008 to 2012, informatics developed rapidly and library science got stuck. The block in which journals of informatics lie replaced the core status of library science and became the knowledge source in the field of library and information science. During this time, the status and roles of the two blocks where journals of informatics lie gradually converged, and the core-periphery structure of the entire network was weakened. This shows that with the exchange of professional power the knowledge gap and interactive model will change accordingly. Thus the hierarchical structure of the knowledge transfer network of cross-citation in library and information science journals in China evolves towards a fuzzy core structure and a distinct periphery structure.

Based on the traits of nodes and blocks in the network of journal cross-citation knowledge transfer in library and information science in China, we construct the pattern of its knowledge transfer. The knowledge transfer activity of journals of library and information science shows the evolving tendency from weak-weak subgroups to weak-strong or strong-weak subgroups and finally into strong-strong subgroups. The blocks should grasp their own developing condition, and prompt the formation of knowledge transfer cohesive subgroups which have good interaction.

At present, it is still a brand new idea to study the process of knowledge transfer by cross-citation. It shares many similarities with citation identity and citation image so they can learn from each other. In this paper we have tried to serve as a modest spur to induce someone to come forward with other, related valuable contributions. We based our research on the citation environment of CSSCI, so the results of analysis might have certain limitations. We hope that follow-up research can show more comprehensive, intuitive and dynamic journal knowledge transfer activity by collecting more intact journals’ citation information in library and information science and applying multiple analysis methods, so as to provide more sufficient and valuable information for the academic circle (periodicals circles, scholars, institutions, etc.) of library and information science.

References


Freitas, Juliana Lazzarotto, Gabriel Junior, Rene Faustino and Bufrem, Leilah Santiago. Theoretical approximations between Brazilian and Spanish authors’ production in the field of knowledge organization in the production of journals on information science in Brazil. Knowledge organization 39: 216-23.


Development of an Information Support System for Yogic Science using Knowledge Organization Systems†

Vinod Kumar B.L.* and Khaiser Nikam**

*/** Department of Studies in Library and Information Science, Manasagangotri, University of Mysore, Mysore, Karnataka 570006, India
* <vinodkumarbl@gmail.com> / ** <khaiser.nikam6@gmail.com>

Khaiser Nikam, is the former Head of the Department of Library and Information Science, University of Mysore. She obtained her Ph.D., M.Lib.Sc., B.Lib.Sc. and B.A. from the University of Mysore, and an MBA from Karnataka State Open University, Mysore. Her specializations are user studies, knowledge organization, and corporate and public libraries. She is a member of several academic, advisory boards and selection committees and has served as member of the editorial board of four Indian periodicals in library science.

Vinod Kumar B.L. is the former chief librarian of the Vydehi Institute of Dental Sciences and Research Centre Library, Bangalore and deputy librarian of Swami Vivekananda Yoga Anusandhana Samsthana Deemed University Library, Bangalore. He obtained his Ph.D. from the University of Mysore, M.Phil. from Annamalai University, Annamalainagar, M.Sc. and B.Sc. from Kuvempu University, Shankaraghatta, and a post-graduate diploma in yoga education from Kaivalyadhamla, Lonavla. His research interests are in the fields of knowledge organization systems, digital libraries, and yogic science.


Abstract: This paper deals with the design and development of an information support system for yogic science using specially designed knowledge organization systems such as a yoga glossary and yogic thesaurus. A machine-readable Sanskrit-English bilingual glossary, thesaurus for yogic science is developed using Greenstone Digital Library software, and also there is a web portal for the yogic science community, which includes a list of all major yoga institutes, research centers, libraries, glossaries, thesauri, yoga subject term visualization maps, Google groups, forums, online digital repositories, and online public access catalogs related to the discipline of yoga.

Received: 12 April 2014; Revised 25 June 2014; Accepted 28 June 2014

Keywords: yoga, yogic science, information support system, knowledge organization system, glossary, Sanskrit-English thesaurus

† The authors are grateful to Prof. A. Neelameghan, Sarada Ranganathan Endowment for Library Science (SRELS), Bangalore. We shall always cherish with very high respect and gratitude to the invaluable time and efforts spared by him for this purpose. We also thank S.K. Lalitha, SRELS, Bangalore, for extending her unselfish assistance to create the GSDL collection. We thank Paul Matheu, MultiTes USA for providing the MultiTes Pro BETA version. We thank for Prof. K.S.Raghavan for his help. We thank for Dr. K.N.Prasad for his encouragement from the beginning of this research work.

1.0 Introduction: need for the study

Swami Vivekananda Yoga Anusandhana Samsthan (S-VYASA) is an academic and research centre for yoga and allied sciences and has links with a number of institutions in India and abroad. The digital library of S-VYASA consists of an online catalogue of information resources—books, reports, conference proceedings and full texts of research papers, theses and dissertations are accessible on their website at www.svyasa.edu.in. Yoga is an ancient discipline and many of the early works are in Sanskrit, e.g. Patanjali’s Yoga Sutras (PYS), Gheranda Samhita, Yoga Vasistha, Hatha Yoga pradipika, Shiva Samhita, etc. Even today researchers in Yogic science publish in San-
skrit on the subject. In recent times, however, there are books, reports and papers on yoga, its application and effects on human health in English and in some other languages of Latin origin. In this context the need for a vocabulary management tools was ascertained. Hence we undertook this research for the design and development of an information support system using knowledge organization systems (KOS) such as glossaries and thesauri.

There are several yoga teaching, practicing and research centres, but very little information is available about their activities and contributions to society. There are several yoga teachers spread all over the country and abroad but details of their expertise and wisdom are not well known. There are several publications on yoga and its application, in the form of books, journals, research papers, conference proceedings, dissertations, monographs, audio and video cassettes and CDs, but there is no comprehensive documentation centre. Hence there is a need to bring together all items published and further regular updating is very essential to maintain continuity in the citation of relevant documents on yoga.

In this paper we discuss the need for designing an online digital library on yoga and related aspects. The aim of this paper is to indicate how a library similar to other online digital library on yoga and related aspects. The aim of this paper is to indicate how a library similar to other digital libraries in the world on yoga can be designed. We show some examples of the institutes which have taken up similar initiatives as those planned here. The existing online digital repositories of yogic science that are listed in the web portal created by using Google are shown (https://sites.google.com/site/yogaandalliedsciences/).

As a suitable example for design and development of a yogic science digital library we discuss the networked consortia for sharing of resources in yogic science at the National Information Centre for Yoga and Allied Sciences (NICYAS). As an introduction we discuss some basic concepts relating to yogic science. Yogic science information resources are published in a variety of formats and in order to enable the library-and-information centres to serve the varied needs of their users it is essential that information is shared among the various Yogic science organizations. This will help in cutting down library expenses. No study of this nature has been undertaken so far in the field of yoga in India.

Yoga is one of the six orthodox systems of Indian philosophy. It was collated, co-ordinated and systematized by Patanjali in his classical work, the Yoga Sutras, which consists of 185 terse aphorisms. In Indian thought, everything is permeated by the supreme Universal Spirit (Paramatma or God) of which the individual human spirit (jivatma) is a part. The system of yoga is so-called because it teaches the means by which the jivatma can be united with, or be in communion with the Paramatma, and thereby attain liberation (moksa).

In India, research and academic libraries in the field of yoga are building digital libraries. The yogic science fraternity requires access to digital information resources. Yoga is the very foundation not only of Indian culture but also for other cultures of the world. Yoga is an ancient discipline. It is recognized as an important and valuable heritage of India. The word yoga is derived from the Sanskrit root yuj meaning to bind, join, attach and yoke, to direct and concentrate one's attention on, to use and apply. It also means union or communion. It is the true union of man's will with God.

2.0 Use of multiple knowledge organization systems (KOS) together

An information system is designed to store, retrieve information and provide different services to meet the needs of different classes of users (researchers, students, practitioners, the general public). It needs user-friendly interfaces, good indexing and retrieval facilities. It is well recognized that methods and devices to manage the terminology used in the system are necessary both at the data entry stage and in formulating search expressions to secure adequate relevance and recall. A knowledge organization system such as schemes for subject classification, thesauri, taxonomies, term nets, etc. alone is not adequate. A combination of two or more such devices can provide better facility and satisfaction to users of an information system.

3.0 Prior literature

Some of the previous studies in the field are those of Hudon (2001), Jorna and Davies (2001), and Kwassnik and Rubin (2003), which have dealt with problems of multilingual thesaurus design and development. Earlier studies mostly deal with terms of Latin origin. There are very few studies dealing with terms of non-Latin origin. Neelameghan (2001), Neelameghan and Raghavan (2005), and Raghavan and Neelameghan (2008) describe their work of preparing bilingual and multilingual thesauri, F-thes (Sanskrit, Farsi, English) and G-themes (Tamil and English). A comprehensive and well-organized information support system in the field of yoga is very much needed with special emphasis on KOS for management of vocabulary of the field.

A yogic science project proposal for the creation of a National Information Centre for Yoga and Allied Sciences (NICYAS) at Prashanti kuteram was conceived by Neelameghan and Rajashekar in 1999. Yoga is practiced and taught in many centres of the world and many papers are generated. Some of the institutes we list have already initiated networking and exchange of information.
with many of these centres in India and abroad. The vision to create NICYAS is intended to avoid duplication of work and regular updating of the information related to yoga and allied science to disseminate information to every corner of the world.

4.0 Methodology

The principal objectives of this study are:

1. To design and develop a bilingual glossary for the yoga domain.
2. To design and develop a bilingual thesaurus for the yoga domain.
3. To create a knowledge map for the yoga domain; and,
4. To hyperlink terms in the thesaurus to corresponding terms in the glossary.

The completion of the thesaurus and preparation for the glossary required for the design of an information system in yogic science was carried out in a step by step manner. The design of the thesaurus involved the following eight steps:

Step 1: Study of the subject by examining the literature available on the subject, moving progressively from reference sources – dictionaries, glossaries, encyclopedias, etc., to the more detailed material, such as, papers in periodicals, proceedings of seminars, conferences, symposia, etc.

Step 2: Creating a corpus of terms.

Step 3: Identifying user needs from yoga library user’s profiles.

Step 4: Selection of terms from the corpus in consultation, personal interaction, discussion with subject specialists, research officers, scientists, divisional deans, faculty members of the yoga university, library users in the yoga university residential campus and the users of yoga information systems. Terms were collected and the relationships among them were discussed to construct the thesaurus.

Step 5: We recorded each term in the format chosen as most convenient. The layout of the ultimate product was decided and the thesaurus was constructed accordingly.

Step 6: Thesaurus construction started with WINISIS software, later MultiTes Pro BETA version was used to create a bilingual Sanskrit-English glossary interlinked to the thesaurus.

Step 7: The customized collection format for Greenstone Digital Library Software (GSDL) was prepared by the Sarada Ranganathan Endowment for Library Science, Bangalore (SRELS) and same format was used for creating a glossary and thesaurus using GSDL.

Step 8: We created a hierarchical map of yoga terms using Microsoft Office Word SmartArt™ graphics to visually communicate information for better visualization of yoga subjects; this was uploaded at https://sites.google.com/site/yogaandalliedsciences/home/yoga-terms-visualization/.

5.0 Sources of concepts and terms

The corpus of terms was gathered from existing dictionaries, glossaries and lexicons. Relevant terms from the OM database and the F-Thes glossary were transferred1 (Nellameghan and Raghavan 2005). A collection of over 5000 books, peer-reviewed journals such as International Journal of Yoga (IJOY), International Journal of Yoga Therapy, research reports, theses and dissertations were used for identifying core concepts. Some selected secondary sources were also consulted (see appendix).

6.0 Software used for designing the information system

In this work we used three software programs; namely WINISIS, MultiTes Pro BETA version, and GSDL software.

6.1 WINISIS

The construction of a thesaurus started with WINISIS software, CDS/ISIS 1.5 for windows from UNESCO 2003, because it is easy to install and use. The relevant terms were transferred from the OM database and the F-Thes glossary, which were developed using WINISIS. Excellent hyperlinking was possible within and outside the database, but was not Unicode compatible; later we moved to MultiTes Pro BETA version.

6.2 MultiTes Software

MultiTes Pro BETA version was provided by Paul Matheu, MultiTes, USA. The bilingual Sanskrit –English glossary was created using MultiTes Pro software, because it is Unicode compatible, data can be imported from Notepad, it is easy to create customised fields and to generate alphabetical reports that can be uploaded to the Internet. However, MultiTes Pro is not open source and not freeware. Later we moved to GSDL software.
6.3 Greenstone Digital Library software

In the meantime SRELS was working on GSDL and the same customised collections format was used for creating a glossary and thesaurus. GSDL was developed by the University of Waikato, New Zealand and promoted by UNESCO. Easy to install, customisation also is possible. It imports data from WINISIS. The collections can be exported to CD/DVD, but take more time for building larger collections and updating. GSDL was used for building bilingual thesauri (Kumar and Nikam 2011) as it was Unicode compliant and allowed hyperlinking within the thesaurus as well as to external sources.

7.0 Scope and Limitation of the Study

The present study covers a broad area of yoga and its branches. But only the yoga glossary and thesaurus are used as they are helpful in finding translated words, near equivalent words and transliterated words in both Sanskrit and English languages. The glossary and thesaurus cover the needs of the users of yogic sciences information systems only. The limitation is that the system will be especially useful for students and teachers of yoga as it deepens the knowledge about the yoga.

The study also covers different types of yoga systems beneficial to scholars, researchers, teachers and practitioners of Yoga. The present work will open a new vista in research on the rich and highly variegated tradition of yoga as handed down by the nātha-siddhas and munis to the advantage of their posterity. This will also go a long way toward reestablishing the fact that the knowledge of the siddhas and munis is authentic and authoritative as it dawns from intuition. They combined ‘parikśa’ (mental investigation) with discovery by intuition to produce systematic knowledge.

Further information system development is limited to the Indian environment. Whenever exact translated words were not available in English, near equivalent and transliterated terms for the original Sanskrit terms were included in the study. We have not done evaluation with users, because the system is developed based on S-VYASA Yoga University users’ needs extracted from library users’ profile.

8.0 Outcomes of the study

8.1 Interface features

A bilingual glossary containing 1924 terms for the yoga domain was designed and developed. A Sanskrit-English bilingual glossary was developed using MultiTes Pro v2011.02.20ux BETA version. The display of alphabetical generated glossary terms shown in Figure 1 as well as the glossary are uploaded for wider usage at http://www.svyasade.com/library/inc/YG/alpha.htm and https://sites.google.com/site/yogaandalliedsciences.

![Figure 1. Yoga glossary showing the description for the term “Yama”](image-url)
A yoga bilingual thesaurus called Y-Thes for the yoga domain was developed using WINISIS software and later exported to GSDL. Display of Y-Thes in WINISIS and the Sanskrit-English yoga glossary as linked to the yoga thesaurus are shown in Figure 2.

A Unicode compatible software, GSDL, was used for building the bilingual thesaurus as it allows hyperlinking within the thesaurus as well as to external sources and search options using an onscreen keyboard with Sanskrit language.

8.2. Example of display of search terms

Sanskrit terms can be accessed by pressing the “Sanskrit Term” button in the navigation bar. This displays a list of terms in alphabetical sequence. Clicking on any of the displayed hyperlinked terms, say यम (yama), will retrieve matching records in the pre-selected databases (Figure 3). English descriptors can be accessed by pressing the “English Descriptor” button in the navigation bar. This displays a list of terms in alphabetical sequence.

![Figure 2. Yoga thesaurus with Sanskrit-English yoga glossary hyperlink](image1)

![Figure 3. Display of thesaurus in Sanskrit](image2)
8.3 Search facility in GSDL

In the Yoga Thesaurus search facility, users can type a Sanskrit term using an on-screen keyboard (Figure 4). For keying in search terms in languages other than English a Unicode compliant keyboard is necessary. The terms may be dragged from the list and dropped in the search box if necessary.

8.4 Hierarchical map of yoga terms

We created a hierarchical map of yoga terms using Microsoft Office Word SmartArt™ graphics to visually communicate information for better visualization of yoga subjects. The map was uploaded at https://sites.google.com/site/yogaandalliedsciences/home/yoga-terms-visualization/ and this is shown in Figure 5.

8.5 Term translation practices

Equivalent English words are not available for the traditional Sanskrit origin yoga words and equivalent Sanskrit words are not available for the English yoga words. For many Sanskrit terms or concepts exact English equivalent terms are not available; similarly, for the recent works about yoga in English, Sanskrit equivalent terms for some of the specialized English terms are not available. In such cases hypertext linking to transliterated terms and scope notes or descriptive definitions was used (Figure 6).
9.0 Conclusion: proposed advantages and future study

The Sanskrit-English bilingual glossary will help improve understanding of yoga concepts for users. The Sanskrit-English bilingual thesaurus will help for better understanding of the relationship and navigation among concepts in the yoga. The knowledge map for the yoga domain is created to improve understanding of the concepts and relationships in Yoga. The information support system will facilitate the exchange of knowledge between modern English yogic practitioners and the traditional Sanskrit origin saints/sadhakas.

Goals for future study include creating a classification scheme for yoga and allied sciences. Alternatively, if a definition is available in the target language, one could develop a coding scheme (preferably a classification scheme) for the domain of the thesaurus; in this case the search term might be a code that the system can pick up whatever the language. Coining new terms in the target language will be the task of subject-language specialists when there is no suitable term for a concept in one or more of the languages of the thesaurus. Finally, we hope to undertake discussion with library professors, yoga scholars, yoga practitioners, and subject experts for feedback for further development of the system.

The accessibility of different information resources by a person sitting in a remote and isolated corner of the world through the Internet is going to reduce the burden of establishing duplication of institutional work in different parts of the world and in different countries. This is going to benefit the poor and underdeveloped nations in a big way by reducing the burden of investing huge amounts in designing systems such as the Yogic Science Digital Library. The system permits searching terms in Sanskrit, English and navigation through hypertext linking to equivalent terms, hierarchical terms, and associative terms in a thesaurus. WINISIS provides good hypertext linking facilities among terms in the thesaurus database as well as hyperlinking of terms in the thesaurus to lists, texts and other files outside the thesaurus residing in the same computer as the thesaurus database or in another computer in a remote location. MultiTes Pro software is Unicode compatible and suitable for Indian languages such as Sanskrit/Devanagari. GSDL is a suite of software for building and distributing digital library collections. It assists in organizing information and publishing it on the Internet or on CD-ROM. One of Greenstone’s unique strengths is its multilingual nature, Unicode compliance. The visualization maps will help to understand the concepts and their relationships. The hierarchical tree map structure will help aid understanding of the concepts and relationships among them and help lead to the creation of a depth classification scheme and thesaurus. The combination of a yoga thesaurus, a yoga glossary and a subject map is more useful for end users.

Note

1. The multilingual system was named for the Sanskrit mantra “OM” signifying sacred importance.
References


Neelameghan, Arashanipalai and Rajashekar,T.B. 1999. Project proposal; creation of National Information Centre for Yoga and Allied Sciences (NICYAS) at Prashanti kuteeram. Bangalore: VYASA.


Appendix 1.

Source texts for glossary and thesaurus construction


The Philosophy of Language and Knowledge Organization in the 1930s: Pragmatics of Wittgenstein and Ranganathan†

Gustavo Silva Saldanha

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 155, Rua Lauro Muller, Botafogo 22 290 160 Rio de Janeiro, RJ, Brasil, <gustavosaldanha@ibict.br>


Abstract: This paper proposes a historical-epistemological study of knowledge organization focused on the 1930s. The main question is the development of a thought-oriented pragmatics of language in the philosophical scene of the period. In distinct foci, the Austrian philosopher Wittgenstein and Indian philosopher Ranganathan discussed epistemological development under the premises of a pragmatics of language. The aim is to propose analysis of the dialogue between the visions of Ranganathan and Wittgenstein in the construction of the possibilities of knowing from the use of language within the 1930s. The method adopted is the “panoramic presentation” of Wittgenstein, based on the analysis of concepts to understand the construction of a grammar within a community of language use. As main results, we draw attention to the potential correlation between the development of the philosophy of language in the 1930s and the construction of an epistemology for the organization of ordinary knowledge. Knowledge organization needs to review recognition of the pragmatic construction throughout the twentieth century as a theoretical assumption for the understanding of a notion of knowledge from the point of view of pragmatics.

Received: 12 May 2014 Accepted: 14 May 2014

Keywords: knowledge, language, philosophy, use, Ranganathan, Wittgenstein

† The author wishes to thank Bruno Nathansohn, of the Instituto Brasileiro de Informação em Ciência e Tecnologia, for the paper’s translation into English.

1.0 Introduction

In 1930 we find the production of both atomistic and pragmatic conceptions of language. On the one hand, there is the search for a relativistic postulate, and on the other hand there is an essentialist perspective on language, which is placed on the revision of philosophy itself as a practice of the knowing. In other words, “knowledge” is observed under the prism of language; however, this viewpoint originated from two clearly distinct views. We have pragmatic empiricism, which refers to an open plan of the construction of language through ordinary experience, and simultaneously there is neopositivist metaphysics, which searches for rationality based on the idealism of a language without cultural barriers.

In the same decade, the Viennese philosopher Ludwig Wittgenstein, who was the most prominent thinker in the neopositivist vein of the Vienna Circle, was transforming his philosophical proposal, which was originally developed in England. Wittgenstein takes leave from the logistic reflection of language, and draws closer to a discussion of the multiple characteristics of the production
of meaning. On the other hand, the Indian mathematician Shiyali Ramamrita Ranganathan started his philosophical work about librarianship when he was in the city of Madras, after he had stayed in the United Kingdom for a period. Far from identifying a parallelism according to the perspectives of each philosopher, it is noticed that there is an exercise of reflecting both on knowledge and on knowledge organization, which comprises respectively the perspectives of both Wittgenstein and Ranganathan in a convergent context.

Acknowledging the relevance of philosophical studies for knowledge organization (KO), we propose, thus, a historic-epistemological study focused on the decade of the 1930s. The pivotal issue is the development of thought aimed at the pragmatics of language in the philosophical scenario of that period. This thought takes ordinary language as a founding element of sense. The idea of usage is recognizable as structural, and only by it is it possible to conceive meaning. Within different perspectives, the Austrian philosopher Wittgenstein and the Indian philosopher Ranganathan discuss epistemological development according to the background of a pragmatics of language.

We approach the production of sense from the concept of use. Ranganathanian pragmatics makes possible withdrawal from the western-modern notion of “knowledge.” Since the philosophy of language was established in the ideas of the Indian philosopher, we can think in terms of “ordinary knowledge,” taking the practice of organization on the informational level as responsible for a plural notion of epistemology. This positioning draws Ranganathan closer to Wittgensteinian thought, and to the profound development of philosophical studies of ordinary language in the period.

The main objective of this work is to propose an analytical dialogue between the perspectives of Ranganathan and Wittgenstein, on the level of the construction of possibilities of knowing, from the use of language in the 1930s. As specific objectives, we try to define pragmatics as guidance for the construction of knowledge in the 1930s, recognizing the philosophy of language according to Ranganathan, and debating the conditions of the transformation of KO, which is provoked by the pragmatic philosophy of the Indian philosopher.

The method adopted is a panoramic presentation of Wittgenstein, based on the analysis of concepts that conceive of the construction of a grammar inside of a language-using community. Through the panoramic presentation we can identify presuppositions that support effective communication among members of a group, including (for example) a scientific community. As main results, we see the potential for correlation between the development of the philosophy of language in the 1930s, and the construction of an epistemology of KO. Besides the reconstruction of the notion of “knowledge” itself is the issue of information studies, which was started by Ranganathan, as well as possibilities for new meanings for the practice of KO from the point of view of pragmatics. As a conclusion, we highlight the need for a philosophical review of KO, the recognition of the construction of pragmatics through the twentieth century as a theoretical assumption for KO, and the understanding of a notion of knowledge based on the perspective of usage, according to the pragmatic point of view of language.

2.0 On the path of language

Wittgenstein (1979) understands knowledge as an endless walk through contexts and contextualizations, knowledge as action, conferring upon an individual or his social group direct importance in the construction of senses for the world that surrounds it. In this manner, interaction among routine actors emerges from the knowledge of each specific community, because each group of individuals understands the world differently, and recreates this world according to its own perspective. Wittgenstein (1979, 57) worries about language as an enormous risk, and a permanent challenge for the construction of knowledge. He states: “the routine position of contradiction or its position in the ordinary world: this is the philosophical problem.”

In this manner, Wittgenstein’s pragmatism is a theory that interprets action as a sphere in permanent discontinuity from the plan of routine, finding in the contradiction between the deliberative confrontation among individuals their subjective perception concerning the social, eminent topic of investigation. This focus, based on the second phase of the Viennese philosopher’s theory, points to a wide notion of pragmatics, which has an objective to understand not only the uses of words in language, but also to discuss preoccupation with the definition of ‘truth’. This focus also recognizes and contextualizes discursive changes within different spheres of communication on the basis of social relations and cultural manifestations, contributing to the explanation of dilemmas of rationality, of scientificity, and of commensurability within knowledge in general, as well as within scientific knowledge, supplying alternatives for the construction of an ethics that discerns alterity within its routine arena.

Such Wittgensteinian pragmatism finds a clear approximation with the Ranganathan vision about the the process to know, and about the process to organize, and the products of the process to know. When the Indian philosopher debates the range of his philosophical proposals for librarianship, he highlights the need for contingent thought, which is established according to the vision of each culture that takes possession of contents produced as
a reflection of knowledge, as well as comprising a vision about the need for transformation of the manner of perceiving the ways in which each bibliographic artifact can be seized, its material-symbolic contingency basic to the understanding of such appropriation. What Ranganathan (2009) realizes as a call of books for all is explained by need from a local point of view, recognizing each community, on different levels: of genre, of profit, of class, and of territoriality. Just like that, each group of individuals is included in the democratic vision of access to books: men and women, urban and rural residents.

Because of the contemporary issue about access to knowledge and the emergence of knowledge in the epistemology of information science, there is demand for relevance of approaches to the philosophical reflection about pragmatics. Ludwig Wittgenstein's thought allows a pragmatic-oriented reading, extending the potential of Ranganathan reading, for instance, given the following aspects:

- The viewpoint of a philosophy of information that is based on a systematic observation of ordinary language;
- The idea of a relationship among science, society, and culture that can be applied within information science;
- The change from the point of view of the scientifi city of information science—from the question: “what is science?” which is unfolded in “social science or humanities?” to the question: “when and how are we useful and responsible in the condition of specialists?”; and,
- The debate about the long line of thought in information studies, which is unilaterally structured in a representational philosophy of language.

Perceiving the Indian philosopher's thought as contemporary with these reflections, we can recognize the revolutionary role of Ranganathan in the context of those highlighted questions.

The distinction between essentialism and pragmatism in philosophy is also found in the base practices of KO. It is a matter of what is a deemed backdrop philosophy presented in two directions: representational philosophy, and pragmatics, or pragmatic philosophy, as two scenarios of identity in this area. The notion of a backdrop transmits the idea that those are the two scenarios in which the entirety of the area dedicated to organizing knowledge is always positioned. In the case that we want to observe it from another backdrop, we truly escape from its possible nucleus; that is, studies oriented to knowledge organization, representation, transmission and preservation.

On the one hand, the representational tradition takes language as means and, therefore, as representation. Here the professional who works with the organization of knowledge as a mediator is concerned with the accurate representation of signs, using formal logic, and technology as an ideal model to optimize re-representation and its dynamics; here, meaning is viewed as an ideal representation of a term. On the other hand, the pragmatic tradition presents language as action and therefore as transmission. Here the professional works in the organization of knowledge as interpreter, concerned with contextual transmission of signs, presenting analysis of narratives and of discourses, as a method to understand the possibilities of the organization of knowledge. Meaning is viewed as the use of a term, or its experience. Just like that, pragmatism shares from an anthropological point of view of reality, that is, we can understand the world and its social relations in order to organize its heritage from deconstructing understanding of social uses, which gives meaning to artifacts, and to words.

In this latter tradition, we follow both the “second Wittgenstein” and that presented in Ranganathan, be it a philosopher of librarianship or a theorist of classification. This approach allows us to rethink the construction of the studies of information, mainly in the context of the acknowledgement of language as a unity, not only strategically, but also as a foundation in the area.

3.0 Ordinary language philosophy in Wittgenstein

Ludwig Wittgenstein, marked the history of philosophy with his atomistic view about the pragmatic possibilities of comprehension. Wittgenstein tried to present in his second theoretical phase, centrally represented in his Philosophical Investigations, and peripherally by compiled works in the Brown Book and Blue Book among others, that the signification of words is a result of their use in discourse. Here he demonstrates that the occurrence of terms out of sense is a disenchantment. It is true that this existence is a mistake of language. The use of the term is what confers its meaning.

In the early days of his reflections, however, Wittgenstein believed in an a priori order in the world. Thus, in the beginning of his arguments, the world had a fixed structure (Monk 1995, 126-27). These reflections are theoretical marks of its first discursive phase, defined by the edition of the Tractatus Logico-philosophicus, written between 1917 and 1918, and published in 1921, while the philosopher was fighting in the First World War. The concern with terms and their meanings is, in this work, argued by Wittgenstein, because of the logical foundations of searching for an essence for the uses of language.

As Oliveira (2002, xxiii) stated, “logic creates ... the frame for structuring our knowledge of the pre-existing world.” Thus, the logic investigation means “the investigation of the whole regularity; out of the logic everything
is aleatory, accidental” (2002, xxviii). Bertrand Russell’s preface itself in the work corroborates this logic of searching (2002, 2-3): Wittgenstein “is concerned with the necessary conditions for a precise symbolism, e.g., a symbolism in which a phrase means anything that is defined.” In other words, “every function of language is to have sense,” and it “only satisfies this function in so far as it is approached of ideal language postulated.”

The appreciable differences into the direction to other perceptible views in the philosopher’s later writings that make that his interpreters divide his works correctly into the first and the second Wittgenstein keep some open elements by the argument on the *Tractatus Logico-philosophicus*. Thus, to Wittgenstein, philosophy is a “critics of language” (Wittgenstein 2002, 53), and his object is the “clarification” of thought. Philosophy, therefore, in the Wittgensteinian thought, “is not a doctrine, but an activity, in general” (Wittgenstein 2002, 63), an activity of “enlightenment.”

At the end of the War, the philosopher faced obstacles to edit his *Tractatus*, so he left for a small Austrian village, called Trattenbach, where he gave classes for children (Monk 1995, 183-84). During the 1920’s, whilst the ideas from the *Logic-Philosophical Treatise* were reverberating, studies in the Viennese Circle conferred substantial emphasis on his work from 1929; at this same time, Wittgenstein’s book was accepted as a thesis by Moore and Russell (Monk 1995, 250), and the concern about language led the philosopher to construct and edit a dictionary based on a “primitive grammar,” but not official. It is his approach to pragmatism. Here, there is a philosophical acknowledgement of how languages can be used by communities under multiple forms, and all of them can be configured as perfectly valid, since they are coherent with the routines of use.

Still in the 1920’s, Wittgenstein dictated to a typist his thoughts which were later published as *Philosophical Observations*, the work that made distinctions of different (Monk 1995, 268). Here, Wittgenstein wrote (2005, 57): “grammar proportionates to language the necessary degrees of liberty.” It is the principle of a thought that is oriented to the dynamics of language games, to words as action and not as representation; it is dedicated to language game=phenomenology, its description is in the search for social rules that constitute grammar (Wittgenstein 2005, 37).

In the first half of the 1930s, Wittgenstein offered classes to a select group of students. The compilation of material from his classes was recorded in the publications known as *Blue Book* and *Brown Book*. These two works expand the pragmatic discussion started in the previous decade by the philosopher, an argument that continued until the philosopher’s final writings in the collection of texts entitled *On Certainty*, written in the last year of his life. In the *Blue Book* are additional reflections on the notion of “family resemblance,” which is a possible substitute for the concept of “essence,” presented in the *Logic-Philosophical Treatise*. While the essence of the “young” Wittgenstein was based on an “anxiety of generalization,” according to the family similitudes, the philosopher conferred a relativist view on philosophy and its language. In the *Brown Book*, Wittgenstein intensified his philosophical method, deepening the concept of “language games” (Monk 1995, 304-10).

The *Blue Book*, which was dictated to the students during the school year 1933-34 in Cambridge, had started from a crucial question (Wittgenstein 1992a, 25): what is the sense of a word? His answers to this question were oriented along a pragmatic order, but it is not formalist as in the work *Tractatus Logico-philosophicus*. That which can “animate the signals,” Wittgenstein reveals (1992a, 30), “is its use.” In this manner one can (Wittgenstein 1992a, 43), “agree with something that was done or said by means of presenting a way that leads to this action,” and in some cases “it means to describe the way which was used;” in others, “it means to describe the way itself, and that is in conformity with certain accepted rules.”

The *Brown Book*, recorded by Wittgenstein’s students in the school year 1934-35, will deepen the notion and the method proportioned by language games. For Wittgenstein (1992b, 14), language games are not “incomplete parts of human communication.” What governs these language games are the “rules,” imagining the term “rule” in its vulgar use, that is, its social use (1992b, 25). Wittgenstein also tries in this work to explore the notion of “primitive language” using more arguments, or “primitive words,” that is, as an anthropological viewpoint it apprehends language as a whole in its collective construction, coming upon the concept of “family resemblance”.

At the end of 1930s, Wittgenstein prepared the text that would constitute the work that marks what several authors acknowledge as the core of the second phase of his thought, the *Philosophical Investigations*. Nevertheless, because of the Second World War, he could not finish the work (Monk 1995, 369). The philosopher would not testify to the edition of this work, however, it was accessed by readers in a provisory edition that Wittgenstein left in 1949. (Monk 1995, 483)

4.0 Ranganathan and the “unfinished library”

Shiyali Ramamrita Ranganathan (1892-1972) was born in the State of Madras (Tamil Nadu), in India, and may be considered one of the more original thinkers in librarian-
Generalizing from librarianship in five main deductions, sustained by inductive analysis, are the principles of “facets” of the universe. Thus, Ranganathan (2009) succeeds in using traditionally intertwined philosophical methods by an approach that clearly we would treat today, as incomplete “postmodernism.” The first law he created, from which the other laws follow, finds its genesis, according to his narrative, in an “advanced point of time,” in 1928, when he got back to India. After his studies in England, he worked in the library, and his ex-professor, Edward B. Ross, had told him: “You want to say, the books are for being used; you want to say that this is your law.” From this point starts Ranganathan’s library “cosmology” (Ranganathan 2009, 3).

With the emphasis on preservation, the Indian librarianship philosophy, which is represented by a mathematician, is focused on access as a main practice of librarianship. Imprint is stated in accordance with the development of North American thought lines, and of Otletian documentation, and of information science as a whole. Ranganathan (2009, 7) demonstrates to us perseverant accordance with the dictum “the books exist for being preserved.” This would usurp the first law, which is “the books exist to be used.” The Indian reminds us that, before the press, the copy of Mahabharata carried a full life and, in this manner, negligence with the first law was justifiable.

Thus, the focus on accessibility is stated in Ranganathan’s “library” (2009), as the place of the institution in the core of a community: “when they saw me to suggest a place in the market street, which went through the heart of the town” (Ranganathan 2009, 11), and the opening in the gradually extended times is initially answered, to what the Indian treats on the first law’s “magic of mantra.” By this law, the proposed concept of library here only exists in the process of use: “A modern library cannot exist without readers” (Ranganathan 2009, 42).

While the third law points to a wide rhetorical investigation on the possibilities of publicity applied in the library, the fourth law allows us to think about the library as a metaphor of the “logologic town,” to save the reader’s time means to create an infrastructure that allows the dynamics of circulation of “connoisseurs” and “knowledge.” What counts here are the measures about the spending of time, free access, classified arrangements for bookcases, signaling along the pathways where information can be found, cooperative cataloging, and the construction of a reference service specialized for the user.

This last point would be subdivided into two: fast reference service, and long range reference service. The first would be represented by accessible information kiosks for orienting and instructing the readers, besides giving simple consultations; the second would respond with applied contact between the reader and the reference ser-
vice, which indicates to an attendance of the librarian until the bookcases with the user (Ranganathan 2009, 226).

On the fifth law, Ranganathan (2009, 263) takes the vital principle of libraries to “be an instrument of universal education that gather(s) and diffuses freely all of teaching resources and disseminates knowledge with their supports.” The “big library” for Ranganathan (2009, 48) is in an inversion of the traditional filia of librarianship thought: the focus still is the friendship of knowledge, but this is not in the books, but in the user, precisely, in the dialogue with the user. Because of that, the librarian shall be “friend, philosopher and leader” for the user: the librarian shall serve the community.

The “big library” focuses on the mantra “education for all,” the focus of the second law; a book for each reader: that is, democracy. His “supreme wisdom,” affirms Ranganathan (2009, 59), is in the art of the war strategy: “The second law will not accept a defeat. At the end, it will win. This is our faith.” It would reply by breaking differences of treatment between men and women, between rural men and urban men, between children and adults, between blacks and whites, in every class of workers, searching for affirmation of equality. It is here a clear principle of contextuality, each region responds to a correspondent interest. The “unlimited democracy” defended by the second law responds, finally, to the unrestricted right to teaching and to entertainment (Ranganathan 2009, 92).

In the case of the second law in the Ranganathan thought, we are interested basically in the concept, which is not translatable for the Western world: digvijaya. In medieval India, the term would mean the process of conquest, in the sense of spreading influence on other peoples. In general lines, it would be treated as a military campaign and, by extension, the victory of this campaign took place in a vast territorial space. The digvijaya is responsible for bringing the “democratic novelty announced by the indefatigable Second Law” (Ranganathan 2009, 94).

The librarianship digvijaya only can be thought, to the Indian philosopher, if accepted by the librarianship movement in the United States, in the 19th century, when it became “the land of libraries” (Ranganathan 2009, 95). Its construction would involve a political action centered in multiculturalism. Its discursive project should result in negotiation with the situation of a librarianship legislation liable to make the second law’s mantra possible (Ranganathan 2009, 114), “If exists an important lesson, which the history of the Second Law’s digvijaya clearly highlights, is the responsibility of the minister of education because of the supplying of books for all” (Ranganathan 2009, 137).

Ranganathan (2009, 49) concludes his explanation on the first law calling upon another mantra, since the “Mr. Sri Krishna celebrated words: your right is related to the action, never related to the fruits.” Ranganathan’s mantra seems to sound repeated, as an eternal and precise bell inside the web, “Never forget that in the libraries, the books are gathered to be used, prepared to be used, kept to be used and offered to be used” (Ranganathan 2009, 41).

On the second law, that is, from its “sketch of the world conquest expedition,” a prayer indicates: “Bring the knowledge to whose doors need it/to doors that need it, and to teach everybody to comprehend what is right! Not even the distribution of every land is compared with that form of service” (Ranganathan 2009, 137). The digvijaya, “the march” for the conquest of the world by libraries, according to Ranganathan (2009, 303), “is virtually a short history of the movement through libraries in different countries.”

5.0 The books are not for use; the books are the use itself: Ranganathan and Wittgenstein, language and knowledge in 1930s

The philosophy of language and its deep relevance for information science, as analysed by Blair (1992), finds in Ranganathan a very close perspective of the ordinary language studies in the 1930s, highlighting the Wittgensteinian work. Wittgenstein’s concern always was oriented to language problems and their representation communication, such that the philosopher’s work gained deeper dimensions for analysis in several areas of investigation in information science. Wittgensteinian inserts are examples undertaken by Novellino (1996) in the study of the transfer of information, and in González de Gómez (1993) about epistemological questions of the representation of knowledge. More than this, in the historic reflection of what we treat today as the area of “philosophy of information” we can identify an extremely fruitful perspective of argumentation between both Ranganathan’s and Wittgenstein’s pragmatics. One of the stronger questions in this perspective would be in the concept of use, a revolutionary notion for the current contexts of the epistemological reflection of the area.

The notion of use is presented as one of the fundamentals of Wittgenstein’s philosophy, and its interaction with the connection between theory and practice in Ranganathanian KO is objective. To Wittgenstein (1979), philosophy must think in a practical form, it must answer the question, what must not be thought to be about an immediate perspective. Thus, it must previously approach ordinary philosophy of primitive languages, and practice also as reflection. His philosophical construction is over against the assumptions of KO studies as an area of understanding the uses and products of knowledge. In the perspective of Ranganathan (2009), we never can think of the construction of knowledge without observing the practices of appropriation of the users’ knowledge: a clear view of his first law.
Unfolding the notion of “sense” from the question of use, obtained by pragmatics, the category “sense” in Wittgenstein, and in information science resized by a Ranganathanian reading is one of the instances that demands more investigation and exhaustive debate. To Wittgenstein (1979), one of the interrogative formulations, among those that define all the doubt of his philosophical course, is: what does it mean? It will be from this question, from which the philosopher will start his path toward reaching a pragmatic approach.

Likewise, Ranganathan (2009) did not succeed in developing his philosophy before establishing the general deduction based on the notion of use: it is because of the mantra of the first law, and only because of it, that we can imagine the extent of KO practice. We cannot determine a philosophy of librarianship, in its general character, without an *a posteriori* viewpoint, which is established because of use. The sense, from the notion of library to the definition of contents of each book, only can be given as the pragmatic strength of the appropriation of knowledge.

Because of the understanding of this point of pragmatic inflection about knowledge, it is possible to comprehend what is behind the first law: The book only can “exist” as the appropriation process of itself. Before the mantra “books are for being used,” it is necessary to understand the wide Ranganathanian pragmatic view that the books are use itself. It would be impossible, for the librarian of Madras’s philosophical thought, to argue conversely. The concept of book cannot be taken by librarianship without access and appropriation. In other words, to the philosophy of librarianship the concept of “book” does not exist if this is fallen apart from the notion of “use.” This places Ranganathan in the sphere of the ordinary language philosophy of the 1930s, demonstrating its similitudes with the second Wittgensteinian view.

### 6.0 Conclusion

We demonstrate how pragmatics, today a common discourse in the epistemology of information science, may receive consideration originating from other demarcations of our thought, highlighting the need for a philosophical revision of KO. That is the case of the Ranganathanian perspective about our practice. His perspective of “use,” and his possibility of “signification” concerning the contingent appropriation of knowledge are contemporary with important philosophical analysis, like those undertaken by Wittgenstein.

An important critical revision between the expression “organization of knowledge,” and the notion of “scientific knowledge,” is highlighted, in this last case, to think of the core of reflection of information science in its epistemological construction in the 20th century. It is notorious that the concept of “organization of knowledge” dates back to the Anglophone tradition of library science in the 19th century. It reflects, in a certain measure, a typical way of knowing, and a way of conceptualizing knowledge in the 18th century, based on positivism and in an empirical view of the exercise of human reflection.

If the scientific weight of the term “knowledge” is withdrawn, we notice that, according to both pragmatics and Wittgenstein’s thought, Ranganathan demonstrates that the idea behind the process of knowing pervades all kinds of ordinary knowledge. If knowledge is blurred with empirical science, after the 18th century, Wittgensteinian thought and Ranganathanian philosophy openly question the multiplicity of knowledge, with the pragmatics of the 18th century, and not the logicist unity of knowledge.

Likewise, the expression “organization of ordinary knowledge” responds through a philosophic-historic argument: if we think in terms of “organization of knowledge,” we would tend to approach fundamentally the 20th century. That is the moment when this concept emerged, and, in fact, it is based on the discursive scenery of information science. Conversely, a significant issue to think of organization as the foundational concept in Ranganathan, would be approached closer to the idea of organization of ordinary knowledge. The use of the term ordinary knowledge completes this justification: we have here the idea of knowledge as a socially constructed practice, and based on the diversity of communities, and on the complexity of differences. It is not only a means, like the case of the academic context. It is a kind of knowledge, which is in the process of the use of language, which is experienced by each group of individuals in different spaces: only from this perspective could we think about an epistemology of the area, today predicated as originating from information.

### References


Emerging Discursive Formations, Folksonomy and Social Semantic Information Spaces (SSIS): The Contributions of the Theory of Integrative Levels in the Studies carried out by the Classification Research Group (CRG)

Maria Aparecida Moura

School of Information Science, Federal University of Minas Gerais, Brazil, Av. Antônio Carlos, 6627, Pampulha 31270-901, Belo Horizonte, MG – Brazil
<mamoura@eci.ufmg.br>, <cidamoura@gmail.com>

Abstract: This paper focuses on the discursive formations emerging from the Social Semantic Information Spaces (SSIS) in light of the concept of emergence in the theory of integrative levels. The study aims to identify the opportunities and challenges of incorporating epistemological considerations in the act of acquiring knowledge into the consolidation of knowledge organization and mediation processes and devices in the emergence of phenomena. The goal was to analyze the effects of that concept on the actions of a sample of researchers registered in an emerging research domain in SSIS in order to understand this type of indexing done by the users and communities as a classification of integrating levels. The methodology was established by triangulation through social network analysis, consensus analysis and archaeology of knowledge. It was possible to conclude that there is a collective effort to settle a semantic interoperability model for the labeling of contents based on best practices regarding the description of the objects shared in SSIS.

Received: 12 June 2014; Accepted: 14 June 2014

Keywords: discursive formations, folksonomy, Social Semantic Information Spaces, integrative levels, Classification Research Group

1.0 Introduction

We must question those ready-made syntheses, those groupings that we normally accept before any examination, those links whose validity is recognized from the outset. We must oust those forms and obscure forces by which we usually link the discourse of one man with that of another; they must be driven out from the darkness in which they reign—Foucault
grated to sharing, production, management and mediation. This trend is pursuant to the fact that the informational subject has become a source of information in itself, as the actor is required to comprehend, systematize, manipulate and dynamize informational flows in Social Semantic Information Spaces (SSIS).

The SSIS are characterized as technological environments that link people, knowledge and computer resources. Because they are grounded on social connectivity and semantic technologies, they stimulate the coexistence of specialized communities, organized around information sharing by means of themes and concepts. These semantic spaces arise in the form of private or public social platforms, inside which profiles of individual researchers and investigation groups are generated, sharing different information quality standards based on distinct research interests and insights about the world.

Within this context, the term folksonomy was created by Vander Wal (2004), as a neologism associating the term “folk” with the term “taxonomy,” with the purpose of emphasizing classification agreements on the content shared in digital environments. The purpose of this tool is to generate taxonomies used to categorize and retrieve the contents resulting from users’ experiences and social networks in the web. The term folksonomy carries additional concepts, such as: ethnic classification, collaborative labeling, social bookmarking, social labeling, narrow folksonomy, popular taxonomy, and social classification, among others.

Here we assess the theoretical aspects of the concept of “emergence” within the theory of integrative levels and their development in the studies of the Classification Research Group (CRG). The goal was to analyze the effects of that concept on the actions of a sample of researchers registered in an emerging research domain in the website Academia.edu (http://academia.edu/), in order to understand how researchers’ representations shared in SSIS affected the linguistic behavior of the users in contexts of scientific information sharing. How, and to what extent, do the labels shared in such contexts match those in the systems of information representation adopted in instances of access mediation?

2.0 Emerging discursive formations

According to Foucault (2010), discourse is a set of statements that may belong to different fields but are nevertheless submitted to common functional rules (clinical, medical, psychiatric speech). Discourses should be handled in the instance in which they are played, once they do not represent a place of ease where other questions may arise (structure, occurrence, systematics and transformations). For Foucault, discursive formation (2010, 133) is:

A body of anonymous, historical rules, always determined in time and space that have defined for a given period, and for a given social, economic, geographical or linguistic area, the conditions of operation of the enunciative function.

Contemporary discursive formations had inevitably been consolidated under the aegis of information and knowledge disclosure and sharing. The information system of the objects is permanently modified and slowly deposited into (medical, economic, social) discourse. Pursuant to that, the conceptual systems produced in such contexts are oftentimes heterogeneous or incompatible derivations, making it always necessary to analyze the way concepts arise and their disclosure in a given context. Understanding discursive formation, within the context, may help to identify the structural relationships and, consequently, support the acquisition of strong clues about the proposition of relationship systems that are better associated with emerging knowledge.

For the knowledge organization area, proposing information retrieval instruments in context without understanding the dynamics of discursive formation in a given domain of knowledge has become even more complex. This happens mainly because the concepts refer to the description of a field where the principles are set out. These are formed as a bundle of relations (not an isolated object, an individual work or an area of knowledge at a given moment), where the context, the discursive regularities and coercion, the theoretical choices and the historicity are taken into consideration, and happenings, transformations, mutations and processes are articulated.

Discursive formation is voiced at three levels: surfaces of emergence, authorities of delimitation and grids of specification. Surfaces of emergence highlight the circumstances of instantiation, individual differences and degrees of rationale, conceptual codes and the types of theories. The authorities of delimitation refer to the identification of institutions that adjust the domain and are responsible for distinguishing, assigning, naming and implementing the signification processes as object. Finally, the grids of specification refer to the grouping schemes and classifications adopted in a specific context of discursive formation.

Foucault (2010, 28-29), however, affirms that, in the scientific context:

There exists another principle of rarefaction of a discourse, complementary to the first, to a certain extent: the author. Not, of course, the author in the sense of the speaking individual who pronounced or
wrote a text, but in the sense of a principle of grouping of discourses, conceived as the unity and origin of their meanings, as the focus of their coherence. This principle is not everywhere at work, nor in a constant manner: there exists all around us plenty of discourses which circulate without deriving their meaning or their efficacy from an author to whom they could be attributed: everyday remarks, which are effaced immediately; decrees or contracts which require signatories but no author; technical instructions which are transmitted anonymously. But in the domains where it is the rule to attribute things to an author — literature, philosophy, science — it is quite evident that this attribution does not always play the same role. In the order of scientific discourse, it was indispensable, during the Middle Ages, that a text should be attributed to an author, since this was an index of truthfulness. A proposition was considered as drawing even its scientific value from its author.

From a philosophical point of view, emergence is the process by which a phenomenon arises from a combination of causes, while it may not be considered a result of their respective effects.

For Goldstein (1999, 50), emergence refers to the instantiation of innovative and coherent structures, standards and properties during the process of self-organization in complex systems. Its major characteristics are:

- Coherence or co-relation: emergents appear as integrated wholes that tend to maintain some sense of identity over time. That coherence spans and correlates the separate lower level components into a higher level unit.
- Global or macro level: since coherence represents a correlation that spans separate components, the locus of emergent phenomena occurs at global or macro level;
- Dynamic level: emergent phenomena are not pre-given wholes but arise as a complex system that evolves over time. As a dynamical construct, emergence is associated with the arising of new attractors in dynamical systems (i.e., bifurcation).
- Ostensive level: emergents are recognized by showing themselves, i.e., they are ostensively recognized.

Emergence is an important concept in the theory of integrative levels, which influenced the studies of the CRG.

3.0 The Classification Research Group and the theory of integrative levels

According to Gnoli (2011, 30), the structures of knowledge organization involve influences deriving from different factors. Among these factors, Gnoli highlights the ontological, epistemological and pragmatic dimensions. The ontological dimension deals with actual phenomena that are studied. The epistemological dimension deals with the purposes, the cultural, disciplinary and methodological aspects that are considered and discussed in the organization of knowledge. Finally, the pragmatic dimension involves material aspects related to documents, collections and information users.

Founded in 1948-1952, the CRG dedicated itself to the study of classification systems aiming at contributing to the development of documentary research. The focus of its studies involved, above all, the study of a multiple sided assessment of the relational operators and the theory of integrative levels. According to Spiteri (1998), the theoretical work of the CRG embraced the study of facet analysis, relational operators and the theory of integrative levels. Employing the theory of integrative levels, the CRG construed the process of indexing as a complex and ontological relationship in the information producing environment and in the context of research. That importance given to the context of emergence of phenomena contrasted with the centrality of literary warrant adopted at that time.

Given the wide experience of the members of the CRG in the treatment of specialized information in different contexts, the group disagreed with the assumption of a finite universe of knowledge underlying the most traditional classification systems adopted at that time. The major limits of the classification systems referred to the difficulty to maintain and update systems due to changes in the production of knowledge, notably related to interdisciplinary and integrating dialogues pursuant to these processes.

As early as 1950, Farradane postulated on the support of epistemological considerations about the very process of knowledge acquisition (Rivier 1992, 63). From that point of view, the group proposed the establishment of semantic relations a posteriori, and received harsh criticism for abandoning, to a certain extent, Ranganathanian principles.

According to Foskett (1962), apud Spiteri (1995), the appropriation of the theory of integrative levels carried out by the CRG presented the following interpretation:

The world of things develops from the simple towards the complex by accumulation of new and divergent properties and that at certain points changes occur which transform the “entity” from a member of one group or class into a member of a
new group. The new entity has properties of its own, characteristic of the new level of organization within it, and it behaves in a similarly new and characteristic manner.

In accordance with Spiteri (1995), influenced by James Feiblemann (1965), the CRG proposed a set of categories that could apply to the entities, namely: physical entities, chemical entities, non-living heterogeneous entities, artifacts, biological entities, men and mentifacts. Currently, the possibilities to identify, monitor, record and segment the informational practices and the distinct user profiles in digital collaborative environments encourages the recovery of issues proposed by the CRG, with the purpose of identifying the objective potentials of incorporating the epistemological considerations of the process of knowing, organizing and mediating information upon the emergence of the phenomena.

4.0 The research

The systematic approach of human-social phenomena involves, in accordance with Domingues (2004), description, explanation and interpretation based on abstractions of the real world. Pursuant to that, the qualitative research developed had the purpose of describing, explaining and interpreting a contemporary phenomenon within the context of knowledge organization. In order to achieve that goal, the group chose to triangulate the methods and theories according to which the theory of integrative levels was articulated, as well as the archaeology of knowledge, with the purpose of understanding the process of signification in digital environments and its repercussion in the representation and retrieval of information in dynamic contexts. By means of netography and the analysis of social networks (ARS), academic profiles, social aggregations and classifying practices of the Academia.edu site, users were identified and analyzed based on the emergent themes e-Science, e-Research, cyber-infrastructure, and science 2.0.

From a theoretical point of view, e-Science is the term that represents the new trend of scientific issues, according to Hey and Trefethen (2008, 15):

The next generation of e-Science problems range from the simulation of complex engineering and biological systems to research in bio-information technology, protoemies and pharmacogenetics. In many of these instances researchers need to combine the expertise of other research groups and access specialized resources often distributed across the globe .... Similarly in the social sciences the data sets that need to be analyzed are frequently so large and widely distributed that both the memory storage as well as the computational power of individual workstations are inadequate and the use of distributed computing resources will become the norm.

The website Academia.edu (Figure 1) is an open science digital platform dedicated to sharing articles and docu-
ments developed at different stages of the research among academic practitioners. This platform was launched in September, 2008. With that objective, the researchers are able to create academic profiles in the social platform and offer institutional information focused in their research interests. That information generates an individual site with the researcher’s profile. Based on this profile, the researcher is able to follow other people interested in the same research themes and share their digital libraries and academic productions.

The site Academia.edu is based on social webs. The social network connects the users by means of the content and notes generated by these users. From a perspective of knowledge classification, these researchers share the emerging terminology, which is consolidated within the investigation domains, and provide a classification that reflects the social context and the dynamics of the studies carried out therein.

Relevant terms were selected as they related to the thematic groups available in the Academia.edu platform, namely: e-Science, cyber-infrastructure, e-Research and Science 2.0. At that stage, 429 profiles of users were found, combined with the proposed themes, with 124 distributed terms. Using the software Visual Anthropac—a freeware program used to assess the cultural domain, we sought to develop a structure for a free list composed of tags created by the users in each of the groups, and identify the projections found in the articulation between the lists obtained (Table 1).

The list of projections clearly referred to the current common belief set forth by the groups upon articulation of the proposed themes; however, oscillations still occur between naming the groups and their tags. There are tag-naming groups concurrently integrating all other groups assessed. From our point of view, that is evidence of a high degree of semantic articulation.

We also sought to identify the thematic interests of eight researchers who were part of the e-Science interest group. In that sense, a list was prepared with the specific terms of each of the researchers, compared to the projec-

<table>
<thead>
<tr>
<th>TAG</th>
<th>PROJECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>cloud_computing</td>
<td>0.707</td>
</tr>
<tr>
<td>computer_science</td>
<td>0.444</td>
</tr>
<tr>
<td>computer_science_education</td>
<td>0.424</td>
</tr>
<tr>
<td>computer-based_learning</td>
<td>0.404</td>
</tr>
<tr>
<td>and-learning</td>
<td>0.391</td>
</tr>
<tr>
<td>digital_culture</td>
<td>0.389</td>
</tr>
<tr>
<td>date_mining</td>
<td>0.372</td>
</tr>
<tr>
<td>digital_humanities</td>
<td>0.372</td>
</tr>
<tr>
<td>distributed_file_systems</td>
<td>0.355</td>
</tr>
<tr>
<td>document_genres</td>
<td>0.342</td>
</tr>
<tr>
<td>dynamic_load_balancing</td>
<td>0.332</td>
</tr>
<tr>
<td>Education</td>
<td>0.306</td>
</tr>
<tr>
<td>heterogeneous_distributed_systems</td>
<td>0.279</td>
</tr>
<tr>
<td>high_performance_scientific_computing</td>
<td>0.262</td>
</tr>
<tr>
<td>information_retrieval</td>
<td>0.214</td>
</tr>
<tr>
<td>new_media</td>
<td>0.212</td>
</tr>
<tr>
<td>internet_&amp;_society</td>
<td>0.19</td>
</tr>
<tr>
<td>open_source_software</td>
<td>0.18</td>
</tr>
<tr>
<td>parallel_architectures</td>
<td>0.163</td>
</tr>
<tr>
<td>resource_allocation</td>
<td>0.11</td>
</tr>
<tr>
<td>social_computing</td>
<td>0.077</td>
</tr>
<tr>
<td>software_engineering</td>
<td>0.057</td>
</tr>
<tr>
<td>web_2.0</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Table 1. Projection Indicators
tions in the group of tags, with the purpose of identifying similarities and distinctions among them. The result in Figure 2 was achieved.

The observation allowed us to assert that the list obtained represents the concept of e-Science informally agreed upon by the group, as well as to identify the area of interest and of application that prevails among the members. That informal semantic pact is also reflected in the documents shared by the group. Based on the incidence of the terms adopted by the group, it can be observed that, within the assessed group in the Academia.edu website, e-Science is considered a grid of high-performance computers in clouds, focusing on the improvement of research quality in different domains. In that group, 86 terms were found, six of which showed a frequency between 25% and 75%, and indicated that they refer to more consistent terms of the proposed theme. All other 82 terms indicated different approaches of the relevant theme, evidencing a probable semantic or ontological relationship. Among the terms are: digital_humanities, digital_signature, archaeology, archaeological_informatics, authentication, archaeological_graphics_&_illustration, distributed data mining, resource allocation, virtual machines, and information society research.

5.0 Conclusions

The study conducted allowed the observation of a collective effort to settle a semantic interoperability model for the labeling of contents based on best practices regarding the description of the objects shared in SSIS. The recovery and the negotiation of senses related to the concepts that compose a given discursive formation in collaborative environments worked concurrently as surfaces of emergence, authorities of delimitation and grids of specification surface, allowing more expeditious establishment of significations pacts, both in the labeling and in the information mediation and recovery processes. It was noticed that the formation of emerging online social networks and collaborative environments has been fostering a larger verticality of the treatment of discourse in its generating instance, and strengthening collaboration initiatives among the users, regarding the treatment of emerging concepts (identification, delimitation and specification). In addition, the social semantic information spaces, by allowing the identification of discourse succession and co-existence methods, as well as strengthening of the authors associated with them, represent good perspectives of consolidation of procedures that can represent dynamically the emerging knowledge shared in a network. Such changes in the information and knowledge sharing methods require an ontological and semiotic approach in order to understand the possibilities of social classification and the pacts of significations that arise in the social semantic information spaces. It is believed that this theoretical and technological approach may help to improve semantic tools, classifications, navigation taxonomies and methodologies for the construction of a

![Figure 2](image_url)
References


SNOMED-CT as Standard Language for Organization and Representation of the Information in Patient Records†

Virgínia Bentes Pinto*, Camila Regina de Oliveira Rabelo** and Igor Peixoto Torres Girão***

Universidade Federal do Ceará, Departamento de Ciências da Informação.
Av da Universidade, 2762 Benfica 60020-080, Fortaleza, CE, Brazil
*<vbentes@ufc.br> / **<camilaregina.rabelo@gmail.com> / ***<igorpeixoto310@gmail.com>

Abstract: The Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), such as the Medical Subject Headings (MeSH) and the Health Sciences Descriptors (MeSH) is a standard for handling, organizing, representing and retrieval of information in the health context. It is structured, among other things, in 19 categories: clinical diagnosis/disease, procedures, observable entities, body structure, body, substance, biological and pharmaceutical products, sample, physical object, physical force, event, geographical or environmental location, social context, stages and scales, special concepts and qualifiers. We present research results carried out with patients’ medical records in the Walter Cantidio University Hospital, at Federal University of Ceará. The line guiding this study seeks to answer the following question: what is the contribution of these categories to build a representation of the patient’s medical records at the Department of Medical Records and Statistics (SAME), at the Walter Cantidio University Hospital (HUWC)? The objective of the research is to study the contribution of SNOMED-CT for the representation of information within those records. It is therefore an exploratory study supported by neofunctionalist method and content analysis, the physical structure of digitized records was analyzed at the SAME of the HUWC. Then we analyzed a corpus of two patient records with nine volumes, about 4000 pages corresponding to 777 Mb. The results and conclusions show that the hierarchical categories of SNOMED-CT may bring contributions to the representation of the charts, as it is a robust terminology based on ontology, contemplating the essence of the information recorded in these documents. Regarding the physical structure of the chart shows some similarities, and hence can contribute to information retrieval with higher added value, since it allows the use of pre and post-coordination as well as natural language, synonyms and acronyms.
Keywords: SNOMED-CT, patient records, terminology, information

† The authors thank the National Council of Scientific and Technological Development (CNPq; Brazil), for financial support.

1.0 Introduction

The concern about a terminology, aimed at a normalization, whose purpose would be to obtain a language able to facilitate the process of communication between citizens, has taken on massive proportions, mainly in the context of science and technology (S&T). We believe that, in these contexts, the first concepts in this field are developed in chemistry, botany and zoology taxonomy. Recently, terminological ideas started to migrate into engineering taxonomy with Wüster Eugen, of the Vienna School, and D.S. Lotte and Drezen, founders of the Russian school. This new reality has consolidated especially because of the development of studies of lexical terminology.

In the context of library and information science, we can consider that the first efforts in connection with the establishment of terminological aspects started with the bibliographic classifications, as well as the subject heading lists. However, with the emergence of the thesauri in the late 1950s, these ideas started to become more accepted, with the systematization of the work of the Information Center of the United States Department of Defense (ASTIA).

Regarding the healthcare field, we can highlight the first medical bibliography called “De claris medicinae scriptoribus” published in 1506 (XVI century), whose author is Dr. Synphorien Champier (Figueiredo and Cunha 1967). In 1879, the Medical Indexing Service – USA, began to produce Index Medicus, and later on the Database of the U.S. National Library of Medicine (NLM), the Medical Literature Analysis and Retrieval System (MEDLARS), whose online version is called MEDLINE (REIS, 1979). Another venture that cannot be faded into oblivion is the publication, in 1954, of the Subject Heading Authority List by NLM. Then, the Unified Medical Language System (UMLS), the Systematized Nomenclature of Medicine and Index Medicus, in 1986. In addition to these initiatives arises, in 1974, the first version of the Systematized Nomenclature of Medicine (SNOMED). In 1970 SNOMED was released, which, in 1993, was renamed SNOMED International, and in 2000, renamed SNOMED RT. From the union between SNOMED RT and Clinical Terms Version 3 (CTV3) of the National Health Service (NHS), from the United Kingdom, the Systematized Nomenclature of Medicine arises—Clinical Terms (SNOMED-CT), a bilingual reference terminology for representing clinical information accurately and unequivocally, in order to enable communication between health professionals worldwide.

In Brazil, Pedro Luiz Napoleão Chernoviz idealized the standardization of terminologies in the healthcare field. It is what made him publish, in Paris, 1870, the Dictionary of Popular Medicine and Science for Accessory Use of Families. The Regional Library of Medicine also deserves attention for participating in the construction and management of Health Science Descriptors (DeCH), published in 1987.

All these initiatives contributed significantly to the continuous qualitative progress of the terminologies in the healthcare field, with the ultimate goal of standardizing the terms, facilitating communication and making it more efficient between members of the multi-disciplinary healthcare teams, as well as between them and patients, so as to prevent communication noise between these agents involved. From observing this phenomenon, our interest in investigating it was aroused, making us ponder about the following question: what is the contribution of the categories of SNOMED CT to develop the representation (indexing) of the patient records in the Medical Records Service and Statistics (SAME), of the University Hospital Walter Cantidio (HUWC) of the Federal University of Ceará (UFC). This research has the overall aim to study the applicability of SNOMED CT to represent (index) the information along these medical records, regarding the relation between them and the terms present in the categories of SNOMED CT. Within this overall objective there are some specific objectives, as follows:

– Know the physical and logical structure of SNOMED CT and its categories to represent information in the clinical context of the healthcare field;

– Study the physical and logical structure of the patient records so as to make the concepts recorded in these documents, in order to synchronize them with the SNOMED CT terminology.

– Compare these categories, applying them to the representation of the information in the patient records in HUWC-UFC.

Here are, therefore, the important aspects of this exploratory research based on the neo-functionalist method and content analysis, taking, as its object of study, a corpus composed of two patient records, totaling nine volumes and containing about four thousand pages equivalent to 777Mb. The result showed, however, that the categories of SNOMED-CT would be applied to the indexing of these documents, even if their categories are
not shown, maybe because the terms studied are only related to the terminology of nephropathy.

2.0 Patient records

The literature on the history of medicine notes that recordings on the health status of the patients were part of the doctors’ concerns ever since the first centuries before the common era, as the example of Imhotep and Hippocrates. Hippocrates, in the fifth century BCE shows the importance of getting these records, claiming that they “should accurately reflect the course of the disease and indicate its possible causes” (Marin, Massad and Azevedo Neto 2003, 9).

According to Moutel and Baret (2002, 8), until the late eighteenth century, the semantics of the patient records was related to the doctor, only from the nineteenth century is it that the notion of patient records appears in healthcare organizations. “Physicians based their observations, and consequently their notes, in what they heard, felt, recorded in chronological order. Therefore, these stories are focused on documenting a patient’s medical history.” Since 1931, in the United States, a record of the quality of the patient records has become an ethical requirement in hospitals. But only after 1970 was it that patient records came to have a central place in medical practices and other healthcare professionals’ practice, as they include (Moutel and Baret 2002, 9):

Medical, social and administrative data ... administrative record that was a long time out of patients’ reach, to ensure respect of medical confidentiality before others, but it is a controversy that sometimes a whole service could know the a patient’s health status that the patient him/herself ignored.

In Article 1 of Resolution 1.638/2002 of the Federal Council of Medicine (Brasil CFM 2002) of Brazil, patient records are defined as:

A single document consisting of a set of information, signals and recorded images, extracted from facts, events and situations about the patients’ health and assistance they are provided with, done in a legal, scientific and confidential way, used to enable communication between members of the multi-professional healthcare team and continuity of care provided to the individual.

Moreover, Bentes Pinto (2006, 3), shows the particularities and specifications in patient records, stating in her reflections that these stories are:

A document that contains all registered information concerning a patient, whether they be for identification purpose, or socioeconomic status, or health status (observations made by healthcare professionals, x-rays, prescriptions, test results, diagnosis from specialists, notes on the patients’ improvement written by the multidisciplinary healthcare team: doctors, nurse, psychologist ... with respect to the observed progress), administrative matters, etc. Actually, it is the physical memory of the patients’ history, being, therefore, essential for the communication between members of the healthcare team, as well as between the team and patients; being also essential for the continuity, security, effectiveness and quality of the patients’ treatment. Thus, also, helping the management of healthcare institutions.

According to Bentes Pinto (2006), the patient records have two different structures: physical and logical ones. The first contains the informational categories concerning the patient, health care providers and hospitalizations and the efforts of the health care providers. In this group, we have the patient records, data on initial clinical examinations, further requirements and their respective results, final diagnosis(es), treatment(s) undergone by the patients, their daily improvement, data on nutrition, social service, psychological care, prescriptions on drug doses to be used, pre and post-operative evaluation, recovery, surgical annotations, antimicrobial monitoring, anesthesia newsletter, classic control system and summary of discharges in a recovery room. Also, the expense for hospitalization, discharge reports, prescription with guidelines to be followed by the patient, consumption in the operating room and death notifications.

While in the second, logical structure, there is a description of the information itself, about the patient him/herself, i.e, identification, socioeconomic and administrative data, such as: full name, identity number, CPF (Individual Taxpayer Registry), education, religion, home and work address, age, complexion, descent, city of birth, nationality, marital status, spouse’s name, number of children, liability statement, social service observations, profession, place of work, spendings on social security; anamnesis (main complaints, history of current illness, personal and family history, previous morbidity history, services, addiction, diet), physical examination and diagnostic hypotheses; reports from nursing personnel (graphs of temperature, pulse and respiration-TPR, blood pressure-BP and water balance), observations from social, psychological and nutritional service.
3.0 Terminological languages for representing information in the healthcare field

3.1 Considerations of terminological languages

To speak of terminological languages, it is necessary, first of all, to understand what a language is in itself. There are some concepts about language. Some consider it as the biological expression and, therefore, there is the need for good physiological condition of the subject so that communication occurs, idea defended by, among others, the philosopher Humberto Maturana. Others, such as Wittgenstein, state that language is a means of social interaction, including documentary and computer languages. Therefore, regardless of the concepts, every language has the purpose to communicate. From this perspective, we cannot understand it apart from the social context in which it is uttered. For that matter, Echeverría (2007, 50) tells us that the language “is born of the social interaction among human beings. Consequently, language is a social phenomenon, not biological,” once there is “an interaction between different particular human beings” developed in a “consensual domain” in which “the participants in a social interaction share the same system of signs ... to name objects, acts or events in order to coordinate their ordinary activities.”

However, it is not our intention to discuss all these possibilities here; we will focus on exposing some looks at documentary language, found in the treatment, organization and presentation of information present in the patient records. That is important because these documents are set as the main channel of communication and information between the multidisciplinary healthcare team and between the team and the sick person (Bentes Pinto 2006).

Documentary languages (DL) also known as terminology, like any other, are designed and intended to facilitate communication between professionals of the same specialty. They result from the documentary boom that occurred from the nineteenth century because of the large number of documents that permeate the scientific and technological world, and the necessity to reduce linguistic inconsistency caused gradually by the course of natural language, considered polysemous by nature. Due to this polysemy and intended to reduce communication noise in different fields, areas or disciplines there was a call for a standardization of terminology so that the informational flow is achieved more clearly between specific communities.

In the literature of library and information science, it is considered that the structure of documentary languages was built in the late nineteenth century with the publication of the Dewey Decimal Classification (DDC) in 1876, and the Universal Decimal Classification (UDC) by Paul Otlet, in 1895. In 1914, the Library of Congress of the United States published the Library of Congress Subject Headings (LCSH) which was used as the basis for many other lists. According to the ANSI/NISO Z39.19 -2005 (ISO 2005), a thesaurus is a “controlled vocabulary organized in an known order in which the equivalence, homographic, hierarchical and associative among terms are displayed clearly and identified by standardized indicators of relations.” The construction of these languages is initially given by the organization of terms classified according to the categories of knowledge. Actually, they are constructed based on the terminologies of each specialty. According to Cabrè (1998, 70), terminology:

Is a transdisciplinary science because terminological products are pieces of linguistic representation on which should be built any field of scientific knowledge so as to acquire, develop and transfer specific knowledge from any domain (which means, in the field of law, medicine or physics, for example, that the discipline of terminology plays a pivotal role in providing knowledge of transporting terms that mediate in communication, as an identifier of underlying rules in connection with the creation of and relation between terms, and as a method and ability to work.

Terminologies are configured as systems of standardized terms used by professionals in a specific area of knowledge. They are the basis for the construction of documentary languages. Usually when we refer to those languages, we are referring to a specialized language. Therefore, this language may not exist otherwise, only through the use of specialized terms inserted in the given context, since each terminology is valid in the professional extent to which it relates.

A documentary language is a set of signs representing the terminology of a specialty using standardized concepts and is intended to represent, express or describe the contents of the documents, facilitating information retrieval. As Urdicain explains (1996, 307), “the documentary Language is a system ... of standard symbols that facilitate the formal representation of the content of the documents to allow retrieval, manually or automatically, of the information requested by users.” Thus, they have similar specialized thesauri, which map the terms/concepts of very specific areas, such as the Medical Subject Headings (MeSH), built by the National Library of Medicine (NLM), that present a set of descriptors standardized for thematic representation or indexing documents in the healthcare field. To us, thematic representation of the information is understood as the result from the act of reading, identifying, choosing and/or translating informational items concerning the topics discussed in a document.

Initially, documentary languages were limited to cover various fields of knowledge, which were universal and
The Systematized Nomenclature of Medicine (SNOMED) is an international and multilingual clinical terminology to be used in the making of electronic patient records. It covers the clinical content and can be used for indexing documents. Documentary languages are the way in which the representation or indexing of information are made. For a broader understanding of that matter, the following section will present SNOMED-CT, and give further explanations about its main features.

3.1.1 SNOMED-CT

The Systematized Nomenclature of Pathology (SNOP), which started in 1965, published by the College of American Pathologists (CAP). In 1974, the first edition of SNOMED was published, renamed SNOMED International in 1993. The agreement between CAP and NHS was signed in 1999 and, thenceforth, with changes in the logical structure, was born SNOMED Reference Terminology (SNOMED-RT) in 2000. In 2001, the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) was published as a result of the union between SNOMED-RT and Clinical Terms terminology of the National Health Service (NHS) from the UK. In 2002, the UMLS National Library of Medicine (NLM) donated a license for the use of SNOMED-CT in Spanish. In 2007, IHTSDO bought SNOMED-CT's intellectual rights.

The structural organization of SNOMED CT consists of 19 high ranking categories of concepts, descriptions and correlations. These categories are presented below, in Table 1:

SNOMED-CT is set as one of the terminologies with more scope within the healthcare field, for being a multiaxial ontology, which stands out because of the possibilities offered in its structure, in consequence of the large amount of existing terms in its vocabulary. The latest update was published in 2008, and contains in its database of terminology more than 311,000 concepts related to its 19 hierarchical axis and countless subclasses. Thus, the relationship between these classes follows the semantic axis (IHTSDO, 2012). The main difference between the SNOMED-CTs is the concern about indexing, retrieval and semantic interoperability of all data records of the patient records and their peculiarities that had not been thought of in previous terminologies studied. SNOMED-CT includes: The “tables” of concepts, descriptions and relationships; cross-references to CIE-10; MC with epidemiological and/or statistical purposes; SNOMED-CT Technical Reference Guide; a browser that allows to navigate the terminology; and 2 updates: January and July (English Ed.) SNOMED-CT can be used for storing laboratory reports, image storage, anatomy pathology reports, protocols, database of autopsies, support system for decision making, coding the news about healthcare in the press, and tissue banks in electronic medical records. Therefore, “it is an extensive clinical terminology that provides clinical content and expressivity for documentation and clinical reports, which can be used to encode, retrieve and analyze clinical data” (SNOMED-CT 2012, 15).
4.0 Method and material

This research features an exploratory study aimed to allow us to obtain new insights into the SNOMED-CT and its application for indexing of patient records on paper. In order to delve into this topic, we reviewed the relevant literature on each section of the subject. Neofunctionalist method and content analysis were some of the tools that we made use of to interpret reality. We did not find any research related to this present work that aims to study the contribution of the SNOMED-CT terminology for the representation of information in patient records, whose recording is done manually (printed). However, we found some research that includes the use of SNOMED-CT for electronic patient records.

Neofunctionalist theory aims to understand the language of the representation of information, which has the function of improving communication between the members of multidisciplinary health teams and between those members and the patient, with the least amount of noise possible. Such a method is an innovation of functionalism and appears in sociological research and in the administration field to explain facts about the world. This states the “things” written on patient records, for instance, the health professionals that attended that patient. Hence, we think of these stories as having a critical role in communication. If they are not seen this way, their importance will never be acknowledged. In neofunctionalism, the patient records are not only thought of, but in addition, they are observed as to how and if they are fulfilling their function, and, if they are not fulfilling their role, what outcome this might produce.

Regarding content analysis, this theory allowed us to map the concepts/terminology of the patient records and allowed us to map the categories of the SNOMED-CT for electronic patient records.

Table 1. Source: IHTSDO. SNOMED CT, 2013

<table>
<thead>
<tr>
<th>High-level categories of SNOMED-CT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Finding: it refers to concepts related to an observation, evolution, clinical status, e.g.: Tuberculosis, anemia, and low cardiac output.</td>
</tr>
<tr>
<td>Pharmaceutical/biologic product: they are related to their therapeutic mechanism, e.g.: Substance: Generic ingredients – dye, hematoxylin. Products: Clinical Drug - name, strength, dosage form.</td>
</tr>
<tr>
<td>Environments/geographical locations: places in the healthcare centers or hospitals, e.g.: ICU, trauma center.</td>
</tr>
<tr>
<td>Procedure: activities E.g: complete blood count, physical examination, asthma vaccine</td>
</tr>
<tr>
<td>Qualifier value: Contains some of the values for attributes. E.g.: Laterality, weight.</td>
</tr>
<tr>
<td>Social context: They involve the conditions and social circumstances significant for basic care, e.g.: social contexts, religion, ethnic groups,</td>
</tr>
<tr>
<td>Observable entity: E.g: temperature, heart rate, gender</td>
</tr>
<tr>
<td>Record artifact</td>
</tr>
<tr>
<td>Situation with explicit content, e.g.: no nausea</td>
</tr>
<tr>
<td>Body structure: anatomical structure</td>
</tr>
<tr>
<td>Physical object: natural or manufactured. E.g.: catheter, latex gloves</td>
</tr>
<tr>
<td>Staging and scales: E.g.: Glasgow, staging system, gynecological malignancies, FIGO</td>
</tr>
<tr>
<td>Organism (Agency), animals, plants and microorganisms: protocols, mandatory reporting. E.g: anthractis Bacillus, lichen</td>
</tr>
<tr>
<td>Physical force: are the mechanisms involved in traumatic injuries. E.g.: combustion car accident</td>
</tr>
<tr>
<td>Linkage concept: e.g.: a) Assertive ratio b) Attribute</td>
</tr>
<tr>
<td>Substance: they are related to pharmaceutical or biological products. E.g.: chemicals, adverse drug reactions</td>
</tr>
<tr>
<td>Events: do not refer to procedures, e.g.: floods, earthquakes</td>
</tr>
<tr>
<td>Special concept (Concepto especial): Incluye los conceptos inactivos manteniendo relacion con el concepto activo Special concept: Includes inactive concepts relating to active concept.</td>
</tr>
<tr>
<td>Specimen: refers to samples taken for examination or analysis. E.g.: prostate specimen by needle biopsy, CSF (Cerebral Spinal Fluid) sample</td>
</tr>
</tbody>
</table>

Table 1. Source: IHTSDO. SNOMED CT, 2013
To compose the corpus of this research, we selected two patient records. The two patients were hospitalized in the 1990s. The criteria used to choose these documents is that they are composed of nine volumes and contain about four thousand pages, the equivalent of 777 Mb. After this selection, the stories were digitized, to begin the process of content analysis. We carried out a comparative study of the structure of these records and the offers of the proposed categories in the SNOMED-CT terminology. After that, we will extract some information from the patient records to compare to the categories referred to in SNOMED-CT.

5.0 Result and discussion

As presented in Table 2 below, the clinical terminology SNOMED-CT has nineteen categories, namely: clinical finding, pharmaceutical/biologic product, environments/geographical locations, procedure, qualifier value, social context, observable entity, record artifact, situation with explicit content, body structure, physical object, staging and scales, organism, physical force, linkage concept, substance, events, special concept, and specimen. These terminological labels are intended to cover all of the aspects related to the nature of the information on the state of health of a patient.

Thus, the results of the empirical research in the patient records show that, although our research has been done in handwritten records, we see that SNOMED-CT categories can be applied to the treatment, organization, retrieval and management of information of these documents in the information systems of health centers and hospitals, both in the public health sector and the private sector (Table 2).

According to the results of the empirical research conducted, we understand that 11 out of 19 the SNOMED-CT categories were found in the first patient record, totaling 42% of the terms. Conversely, in the second patient record, the data is reversed: we identified that 11 of these categories are found in these records, constituting 58%. It is sufficiently interesting, once our analysis is based on printed records. We also understood that this result can be the consequence of the little knowledge we have of the health field terminology used in SNOMED-CT, and also for some difficulties in translating some terms into Brazilian Portuguese, since SNOMED-CT was designed to be used in the United States even though its intention is to expand its range overseas.

<table>
<thead>
<tr>
<th>Categorias Terminológicas del SNOMED-CT</th>
<th>Patient record 1</th>
<th>Patient record 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical finding/ Disease</td>
<td>Chronic Kidney Disease(CKD)</td>
<td>Bearer of CKD (Chronic Kidney Disease)</td>
</tr>
<tr>
<td>Procedure</td>
<td>Dialysis</td>
<td>Hemodialysis</td>
</tr>
<tr>
<td>Observable entity</td>
<td>Headache</td>
<td>Pale both cutaneous and mucous membrane</td>
</tr>
<tr>
<td>Body structure</td>
<td>Vesical wall</td>
<td>X</td>
</tr>
<tr>
<td>Organism</td>
<td>Candida (fungus)</td>
<td>X</td>
</tr>
<tr>
<td>Substance</td>
<td>Heparin in drops</td>
<td>Intravenous</td>
</tr>
<tr>
<td>Pharmaceutical / biologic product</td>
<td>Lasix</td>
<td>Rocaltrol</td>
</tr>
<tr>
<td>Specimen</td>
<td>Urea</td>
<td>Urine</td>
</tr>
<tr>
<td>Physical object</td>
<td>X</td>
<td>Double catheter</td>
</tr>
<tr>
<td>Physical force</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Events</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Environmental or geographical locations</td>
<td>Sickroom; Fortaleza, Ceará</td>
<td>X</td>
</tr>
<tr>
<td>Social context</td>
<td>Housewife</td>
<td>X</td>
</tr>
<tr>
<td>Staging and scales</td>
<td>Discharge condition</td>
<td>Clinical evolution</td>
</tr>
<tr>
<td>Special concepts and qualifiers</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Linkage concept</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Qualifier value</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Special concept</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Record artifact</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Situation with explicit content</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 2. Association between SNOMED CT and the patient records. Source: Results of empirical research
Similarly, it was not possible to find the categories in which we would be able to retrieve information related to family history, i.e., information about the diseases in the family. For example, in one of the patient records, we took notice of the physician’s observation: “mother with hypertension.” This information cannot be put into any of the categories, since it is not related to the patient’s diagnosis; however, it refers to his/her mother and is eventually a very important piece of information. Another remark we can make is that we did not find, among all the categories in SNOMED-CT, any in which it was possible to classify the specialty in which the patient had been diagnosed, or for which he had been being treated, for instance: nephrology/nephropathy, which is a very important representation (indexing) in any patient records.

6.0 Conclusions

In this research, we seek to know and understand the terminology SNOMED-CT and its use for the representation (indexing) of information in patient records on paper. The result of this empirical research shows that this terminology is structured similarly to other documentary languages, with hierarchical relationships and synonymy, among other things. Therefore, in our point of view, the SNOMED-CT terminology can be used for the indexing of patient records, whether they be electronic, digital or on paper. The clinical terminology of SNOMED-CT is a language that provides innumerable possibilities for representing (indexing) the information in patient records, since it has labels for a large number of categories and all of their relationships. Moreover, this terminology has some other categories related to lab reports, image attachments, specific protocols, construction and maintenance of databases and system supports for decision making and dissemination of information about health needs of a city, state or country, and genome studies, among other things.

Through SNOMED-CT, you have a reference terminology in the health field, which facilitates the exchange of information and is also a tool for decision making and for information retrieval. Therefore, it is a language that processes, organizes, mediates and manages the information. We would like to highlight that in this study, we listed the classes and categories belonging to the SNOMED-CT terminology. However, we had some minor setbacks because, although we applied for a short-term license only for this research, it was not granted. Therefore, our empirical research was done on the online pages of SNOMED-CT. For that reason, we had some trouble understanding the categories, not in terms of language, but because of the difficulty understanding to what the concept referred.

References


The Use of Classification in Archives as a Means of Organization, Representation and Retrieval of Information

Fernanda Ribeiro

Universidade do Porto, Faculdade de Letras and CETAC.MEDIA, <fribeiro@letras.up.pt>

Abstract: In this paper we discuss the interest in and the use of classification in archives, both in terms of the intellectual organization of information in an organic-functional, functional or thematic hierarchical structure and at the level of information representation and retrieval of the informational content. Starting by looking at the concept of “classification,” we go on to analyze the information access tools relative to Portuguese archives. This enables us to understand how classification has been used to establish the organizational schedules that provide context to the informational production of the entities (persons, families, corporate bodies) that generate the archives, and how such schedules represent (or not) with scientific objectivity the informational reality being considered. We then discuss the usefulness of classification in archives as a tool for the representation and subsequent retrieval of the informational content, thus enabling an across-the-board access to information, alongside the access by creators and authors that the organic-functional schedules provide. Comparisons, analogies and differences regarding the use of indexing languages in other information systems, especially libraries, where subject classifications have been in widespread use since the end of the 19th century, are also established.

Keywords: classification, archives, organization, information

1.0 The concept of classification in knowledge organization

Anyone consulting a dictionary or encyclopedia of a general character will find a simple definition for the concept of classification, which can be summarized in a few words: intellectual act or operation which consists of grouping elements that have something (a characteristic, a property) in common to form a set or class. To classify therefore means to form classes of elements with affinities and simultaneously to distinguish them from other classes, whose elements lack such characteristics. The system formed by classes and subclasses that are interlinked is formalized through a classification scheme, which is also called a “classification plan,” “classification framework” or “classification table.” The classical classification schemes are based on the principle of logical division, devised by Aristotle in antiquity. It was later taken up by other philosophers and scientists including Francis Bacon, who developed the famous classification of the sciences in the 17th century, and Leibniz and Comte, who also made important contributions in the field of classification.

The classifications used for the organization of information date back to the pre-classical civilizations and had the function of either the arranging and physical ordering of documentary media or the intellectual grouping of their content or meta-informational representation in accordance with certain criteria (authors and producers, titles,
subjects, etc.), with the ultimate goal of allowing the searching and finding of documents and or information. Definitions of classification identical to that set out above are also formulated in specialized dictionaries of information science. By way of illustration we may mention two paradigmatic works here that represent the dominant thinking in this area of knowledge: the *Diccionario enciclopédico de la información y de la documentación* (Cacaly 1997) and the encyclopedic *Diccionario de Ciencias de la Documentación* (López Yepes 2004). In these two dictionaries, the entries on classification trace its temporal and structural evolution, and we can draw some interesting conclusions from these reviews. First, we realize that it is from the nineteenth century, with Jacques-Charles Brunet, that we can truly speak of bibliographic classification—*dans le domaine bibliographique, c’est Brunet qui en 1804 instaure la première classification importante* (*Diccionario enciclopédico*, 136)—and that the great encyclopedic classifications (enumerative and, later, faceted) appeared from the end of that century and were widely used in libraries and documentation centres throughout the twentieth century. Second, it is clear that the discussion on the issue of classification contains no reference to its use in archives, and there seems to be a consensus that this is an issue for libraries and and documentation centers, with archives being excluded from the area of knowledge organization (KO).

In Cacaly’s *Diccionario enciclopédico*, there is no reference to archives in the various entries on classification and library classification. In contrast, in López Yepes’ *Diccionario enciclopédico* there is a specific entry entitled “clasificación de archivos” from which it appears that the classification in this area is understood more as establishing a category system according to various criteria (nature of the producer of information, geographic scope of the archives, age or type of the documents kept, purpose of the service they provide) than as an intellectual operation to organize and represent information or to arrange and sort documents by placing classification clearly outside the field of knowledge organization. Let us look at the content of the entry, which offers more comments (López Yepes 2004, 321):

*Clasificación de archivos (A). Según quien produce el fondo documental: públicos y privados. Por su categoría o ámbito administrativo: locales, provinciales, regionales, de Comunidades Autónomas, y Generales. Según la edad de los documentos: oficina o gestión, central, intermedio e histórico. Por su finalidad: administrativos e históricos. Dependiendo el número de fondos documentales que custodien: singulares y múltiples. Algunos autores clasifican los archivos por su soporte o temática: cartográficos, audiovisuales, especiales, etc.*

Interestingly, in the *Dicionário do Livro* (Faria and Pericão 2008), the problem of classification is addressed comprehensively with regard to the various areas where it is used in relation to documentation and information—librarianship, archival science and heritage protection of documents—fitting unambiguously into the field of knowledge organization, and it does not exclude the application of classification in archives from this field. There are fifty-three entries in this dictionary which start with the word “classification,” “classifier,” “to classify” or “classifiable” but it is within the generic entry simply entitled “classification” that reference is made to its use in the archives. After a generic definition (an “ordered set of concepts systematically distributed in classes, forming a structure; structuring of concepts into classes and subdivisions to express the existing semantic relationships between them”—Faria and Pericão 2008, 258), the authors present some statements that are more operative, clearly identified with knowledge organization, such as: assignment of a classification system’s indexes to data or documents to facilitate their indexing, arrangement and retrieval; documentary language based on structured representation of one or more areas of knowledge in classes and in which the notions and their relationships are represented by notation symbols. And finally, they conclude with the entry where archives are mentioned (author translation):

*In archival science, the intellectual component of organizing archives, which consists of the development of a framework or a plan, table or code based on criteria of organizational structure, or chronological, geographical, thematic, alphabetical criteria, or criteria of logical and systematic relationships grouped according to their similarities or differences, enabling the location of the series or items within the plan, scheme, etc.*

From the examples listed it is possible to understand that there are distinct interpretations of the concept of classification, though its connotation with the field of librarianship is clearly dominant if we think of the classification in the area of knowledge organization. To corroborate this statement, we can also see the entry on classification by Wellisch in the *ALA World Encyclopedia of Library and Information Services*, in which we see very clearly the librarianship perspective mentioned above. The author states the concept of classification quite clearly and outlines its applications in information organization in these terms (Wellisch 1986, 200):

*It is important to distinguish between three different meanings of the term classification in library practice: in its most literal and basic sense it is the act of classifying or making a classification scheme (the resulting scheme often being called a classification for*
short); second, it is the act of classing or assigning class marks to documents that indicate subject content; third, it is the resulting physical arrangement of documents (books or other materials) on shelves or the related but not necessarily identical arrangement of documents surrogates (catalogue entries) in a classified subject catalogue. The first of these, making a classification scheme, is obviously the fundamental one, the others being dependent on it. In the same entry, Wellisch particularly lingers over the matter of classification schemes, stating their constituent elements (tables, notation and alphabetical index) and examines in particular the best known systems of library classification: Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC), Cutter’s Expansive Classification, the Library of Congress Classification (LCC), Bliss’s Bibliographic Classification, and Colon Classification.

The angle from which we started to discuss classification in archives, as an instrument for the organization, representation and retrieval of information, emanates naturally from the integrated and unified approach that is inherent to information science as it is approached and practiced at the University of Porto. We do not assume an archive and library dichotomy, and therefore the technical component that is developed in the context of information services and systems, whatever their organic framework (institutional, organizational, family, personal), can only be understood as a set of procedures for the processing of information, regardless of the sphere of application. Thus, classification is seen as an intellectual and technical operation, which translates into a categorization and systematization for organizational purposes and a formal representation with respect to information retrieval. The final section of this text comes back to this idea and develops it further.

2.0 The use of classification in archives

In any classic archival manual we invariably find a chapter on classification, perceived as one component of archival organization. The English speak of arrangement, the French of classement, the Spanish of clasificación-ordenación, thus it is a mandatory subject in the training and daily work of archivists.

Herrera (1989) devotes a chapter to classification, which he deems an inseparable operation from ordering, two essential steps in the organization of archives (181):

Clasificar y Ordenar son dos actividades dentro de una más amplia que podemos llamar Organización, perfectamente diferenciadas y esenciales, en aras a la conservación de los documentos, de una parte, e indispensables para inventariar y catalogar, de otra.

And later on in the same book (187):

La clasificación siguiendo el principio de procedencia precisa de su plasmación material en un esquema o cuadro que no es otra cosa que el andamio para sistematizar cada fondo en sus secciones y series. Habremos de distinguir siempre dos niveles o estadíos: el primero se identifica con la estructura o funcionamiento de la institución (por ejemplo, su organigrama; órganos o funciones) y corresponde a las secciones y subsecciones o bien grupos o subgrupos; el segundo nivel equivale a las series documentales, es decir a los testimonios de actividades derivadas de aquella estructura.

We infer from this that we are not talking about content and information from the standpoint of knowledge organization. It is, rather, the use of classification from an organic-functional perspective, which seeks to mirror the structure and activity of the entity that produced the information, though it is nonetheless a very important aspect of the implementation of classification.

This view of the Spanish author is not far from what is postulated by the American archivists, based on the writings of Schellenberg and followed in several countries, including Brazil, by established authors such as Bellotto. Adopting the position of the Society of American Archivists (SAA), Bellotto (2004, 140) discusses the notion of arranjo (a term used in Brazilian means arrangement) and accepts the definition from the SAA glossary: Arranjo é o processo e o resultado da organização de arquivos, documentos e manuscritos de acordo com princípios arquivísticos consagrados, particularmente o da proveniência, respeitando-se os seguintes níveis: arquivo, fundo, grupo ou seção, série, conjunto lógico dentro da série e documento.

Another reference work in the field of archivistics is the manual by Couture et al. (1999). Ten years after Herrera, these authors take a very similar position on the function of classification in archives and also advocate an organic and functional perspective (221):

L’objectif de la classification est d’assurer que tous les documents sont rattachés à leurs fonds d’origine et qu’à l’intérieur de ce dernier ils sont classifiés dans l’ordre correspondant à celui que leur a donné leur créateur… C’est seulement dans ce contexte que le plan de classification fournit la presque tangible de l’application du principe de respect des fonds d’archives. C’est par ce plan que seront identifiées les limites externes et la structure interne des fonds.

However, Couture et al. also consider that the classification plan, the material expression of the classification as an intellectual operation (221):
Constitue le premier instrument de repérage intellectuel des documents, facilitant de la sorte les démarches de recherche à des fins tant administratives qu’historiques. Son existence stabilise les processus de gestion de l’information en proposant un cadre de référence et en facilitant la confection ultérieure d’instruments de recherche spécifiques, tels que guides, répertoires, inventaires et index.

We glimpse here an approach to the librarianship view, through the assumption that classification is also a tool that facilitates access to information in the same way that the Universal Decimal Classification does in the context of libraries and documentation centres.

Although in some authors, like those just mentioned, we find a perspective that is less custodial and more focused on access, one that does not view classification as a mere instrument of the material arrangement of documents, the dominant idea on classification nonetheless remains the one that is delivered by traditional archivistics and legitimized through various dictionaries of terminology in this area. For example, just see the succinct definitions of classification either in the Portuguese Dicionário de Terminologia Arquivística, or the Brazilian one:

Classificação—componente intelectual da organização, que consiste na elaboração e aplicação de um quadro ou plano de classificação (Alves, et al. 1993, 20).

Classificação—sequência de operações que, de acordo com as diferentes estruturas, funções e atividades da entidade produtora, visam a distribuir os documentos de um arquivo (Dicionário 1996, 16).

Considering this consensus on the definition of classification, focused on its use to provide the arrangement of archival funds, it would be natural for such a conceptualization to actually be reflected in the organization of archives and the design of tools to access information. However, we seriously doubt that it is like that, judging by what has been the practice throughout the twentieth century, as we have proven by the analysis of access tools relating to Portuguese archives published between 1889 and 1996.

In a study published a few years ago (Ribeiro 2003), 526 access tools of various types were identified, which, after analysis of their internal structure, revealed a lack of rigor in the organic-functional representation of documents and archives subject to analysis, description, classification and indexing. Indeed, even then we felt that an important conclusion to be drawn after analyzing the types of tools for accessing information in Portuguese archives that had been published, was a clear tendency to favor the production of instruments to serve the interests of research, especially historical research. This perspective necessarily led to the development of instruments that were rather compartmentalized in light of the full reality of the archives to which they related and an appreciation of the documents as sources of information, rather than as components of a functional-organic whole, i.e., the system to which they belong.

The statement that any archival information access instrument should make the context of production of documents intelligible is something that archivists accept as undeniable. However, in practice, the tools produced do not follow this principle. Many of the instruments analyzed are limited to the sorting of description units according to a logical criterion (alphabetical or chronological, in most cases), which, in addition to not translating the organic and functional structure of the archives, in fact distort the archive’s own reality, because the understanding of the context in which the archival units were produced and

---

**Figure 1. Information access tools of Portuguese archives**

![Graph showing information access tools of Portuguese archives](image_url)
organized is lost. The correct representation of archival units analyzed for the purpose of an information access tool can only be achieved if the internal structure of this instrument has the organic and functional classification as the primary criterion for sorting such units. The classification codes are both the elements used to order the description units and also work as access points and allow a proper contextualization of these same units.

The information access tools for Portuguese archives that we chose as the study subject were also analyzed in terms of their internal structure, with various criteria for sorting descriptive information being identified: alphabetical (including subcriteria such as anthroponymical, geographical, ideographic and onomastic), systematic (generic classification, thematic or organic-functional classification), chronological, numerical and topographic. Of these criteria, attention is drawn to the systematic one, based on the use of classification, that is, whenever the internal structure of a tool is based on the organization of groups of description units, listed under items (or classes). Such classification may correspond only to a logical criterion of arrangement without any other purpose (classification in the general sense), or it can result from a systematic order of subjects (thematic classification) or it may represent the organic structure and the archival functions (organic and functional classification). Depending on the sort criteria identified from 443 information access tools it was possible to identify the sort criteria that we mentioned before, as shown in Table 1.

<table>
<thead>
<tr>
<th>SORT CRITERIA</th>
<th>NUMBER OF INSTRUMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alphabetical</td>
<td>54</td>
</tr>
<tr>
<td>- Anthroponymical</td>
<td>15</td>
</tr>
<tr>
<td>- Geographical</td>
<td>43</td>
</tr>
<tr>
<td>- Ideographic</td>
<td>4</td>
</tr>
<tr>
<td>- Onomastic</td>
<td>14</td>
</tr>
<tr>
<td>Classification</td>
<td>74</td>
</tr>
<tr>
<td>- Thematic classification</td>
<td>10</td>
</tr>
<tr>
<td>- Organic-functional</td>
<td>38</td>
</tr>
<tr>
<td>classification</td>
<td>(Total: 122)</td>
</tr>
<tr>
<td>Chronological</td>
<td>88</td>
</tr>
<tr>
<td>Numerical</td>
<td>20</td>
</tr>
<tr>
<td>Topographic</td>
<td>83</td>
</tr>
</tbody>
</table>

Table 1. Sort criteria in information access tools in Portuguese archives

In addition to these data it must be noted that in 83 tools it was not possible to identify any sort criterion of the description units, since their organization inside the tools appeared to be totally random, lacking any kind of logical ordering.

These results are undoubtedly very interesting and indicated a few conclusions. The fact that only 38 instruments (7.2%) have an internal structure corresponding to the organic structure itself and to the functions of the archives to which they relate reveals the inadequacy and inaccuracy of most of the access instruments to archival information that have been published in Portugal. This may be explained, first, because most of the instruments are not the responsibility of archivists but of historians and other researchers, who aim to make available information that they have knowledge of and can be disclosed, and so the lack of archival criteria in such instruments is natural; in addition, the chronological limits of the instruments cover times when archivistics was relatively undeveloped from the technical point of view, and far less developed in theoretical terms. However, after 1980 there were 187 instruments and in these, too—corresponding to the period of greatest development of Portuguese archives—we find a strikingly poor representation of the reality in question.

We also concluded that nearly all kinds of instruments were spread among different sort criteria, although some had greater expression than others. Thus, the catalogues were mostly chronological and topographic; guides were organized in greater amounts according to a classification criterion (but not organic and functional), although the alphabetical and geographical were also quite well represented, the onomastic, anthroponymical, topographic and chronological indexes were the most represented; inventories were the instruments in which an internal structure of an organic-functional type appeared more often, but other classifications and alphabetical, geographical or topographic sort criteria were also used to establish their structure.

In 2013 we again collected, by sampling, information access tools of Portuguese archives published between 2010 and 2012, with the purpose of confirming whether the findings of our previous study could be deemed out of step with current reality. Of a set of 30 information access tools we found that only 9.6% have an organic-functional classification, on the basis of their internal structure. Though this is slightly higher than in the previous study, this figure still reveals the insufficient rigor with which the information access tools represent the archival reality that they choose as their subject for analysis.

Will this finding be justified by the paradigmatic view that still dominates Portuguese archivistics, in which mastery of the technique outweighs a scientific, reflexive and problematizing approach? We believe that concern with “doing” research tools without a previous analysis of the archival reality, which is based on application of the scientific method, has led to access to information being an end in itself and not the final step of a process that requires rigorous knowledge of the object that will be formally represented, using metadata, so that the organic context in which the documents and information were produced and
used becomes intelligible to users who yearn to be able to efficiently access the information they seek.

3.0 Classify to organize and arrange versus classify to represent and retrieve

From the foregoing, we can easily conclude that the use of classification in archives and libraries has been a separate issue and had different purposes, because, in fact, it derives from perspectives that are also different, about how the organization of information is viewed. In archives, particularly in historical ones, the vision is still quite custodial and very much focused on the arrangement of documents; in libraries, the issue of access to information began to be valued early on and therefore classification has been used more in the representation and retrieval of information. They are two distinct but complementary functions that can be performed in any information system, promoting better performance and a more efficient service for users.

The two key dimensions that classification may have can be briefly described as follows:

– organizing and arranging (physically) documents
– representing and retrieving information.

The first of these dimensions in historical archives has been mainly concerned with the intellectual organization of the documentation, resulting in the development of classification tables that supposedly represent the organic and functional structure of the producing entity. However, as mentioned above, that is not always the case because most of the time thematic or functional classification tables are designed and the context of informational production is not rendered intelligible. Using classification to physically arrange documents and give them a physical and systematic organization is not done at all in historical archives, obviously because the main concern is the utilization of space in the stacks and the fonds are not freely accessible.

By contrast, in current records, the lack of classification tables is striking, but on the other hand, “classification schemes” are normally used, intended not only for the intellectual organization of documents but their physical arrangement, too, in files or folders (physical, placed on shelves or electronic, stored in the computer). These schemes are usually thematic or functional and apply mostly to correspondence and “business processes,” leaving out a lot of other documentation generated and accumulated by producing entities.

Libraries, meanwhile, use classification for the purpose of organization and storage very differently, favoring the physical systematization of documents for the implementation of open access, thus avoiding the need for catalogues as tools to mediate between users and information. The organization of collections for free access was and is increasingly the key reason for the use of large universal classifications such as DDC or UDC, despite the recognized obsolescence of such classification schemes. Unlike archives, libraries use the classification to organize and arrange documents according to their informational content and not according to their organic provenance. But this difference does not undermine the use of classification as a central tool for the organization of knowledge, and in any case its use is fully justified.

Considering the second dimension of application of classifications—representing and retrieving information—again we detect substantial differences between current practice in libraries and what is usual in the archives. The emergence of library classification dates back, as we have seen, to the early 19th century with the appearance of the famous Brunet Classification, but it is especially from the development of the DDC in 1876 that we can talk about classification as a language of representation of the informational content or, as we would say today, an indexing language for categories.

The literature on indexing languages in general is abundant and there is undisputed consensus among experts. For illustrative purposes we can cite various classic authoritative reference works (Chaumier 1982; Maniez 1987; Skype 1991; Campos 2001; Gil Urdiciain 2004). There is also plenty of literature on the specific case of classification both on individual languages (DDC, UDC, LCC, Bliss, Colon) and on the theory of classification and general questions on the subject (e.g. see, Simões 2001). There is a long tradition of systematic catalogues in libraries and documentation services, and from the 1950s this began to be supplemented by other types of information access tools based on vocabulary and combinatory languages, of which thesauri are the most typical expression. Subject indexes, databases, subject catalogues with the use of vocabulary languages are widely used and are the absolutely essential tools for the representation and retrieval of the thematic content of documents. In the web world, classifications and thesauri have evolved into other forms of expression, such as ontologies, taxonomies, concept maps, and folksonomies, to make the semantic web a reality that always meets user needs, which today the hermetic language of librarians dominates less than ever; users want to search in their natural language, “Googling” and navigating through cyberspace.

And in the archives? How do we get access to thematic content? Which languages are used and what is the role of classification here? Nearly two decades ago we undertook a study on the use of indexing languages in archives (Ribeiro 1996) and at the time there was virtually no literature on
the subject. From the survey we conducted on Portuguese archives we concluded that subject indexing was virtually nonexistent, in fact in line with what was concluded a few years earlier from the work developed in Canada by the Bureau of Canadian Archivists (1992, 35):

In the archival world, Lawrence Dowler suggests, there is some uncertainty “about the value and effectiveness of subject indexing; that is, [archivists] tend to think it is important, but have very little information about either how it is currently used or how it might be used.” Others have argued that subject indexing is of limited value to archivists, because, whereas books and articles are about something, archival records are part of personal or administrative transactions and, therefore, do not lend themselves to being subject indexed. Archival records are something, as distinct from being about something.

The same conclusion also reached by the participants in the seminar held in 1998 in San Miniato, Italy (Pierulivo and Cerri 2000), a pioneer meeting on the discussion of subject access in the archives.

After about fifteen years, it is legitimate to ask ourselves how the processing of information in archives has evolved and wonder how the integration of archives occurred in the digital and networked society in which we operate. There are no known synthesis studies that analyze this issue, but what we can say is that over the past two decades and under the uncontrollable effect of the internet archivists find themselves pressured by users—who do not want to go to the archives reading rooms to get the information they need—to make their contents available on the web. On the other hand, increasing interest by users focuses on obtaining information on subjects, regardless of whether it is kept by a library, an archive, a museum, a memory institution of any kind. And it is also undeniable that searching such a wide variety of information sources is only effective if the metadata related to content is created, that is, indexing languages (which include classifications) remain essential regardless of whether the types or formats used are more or less interoperable.

It is therefore much more about representing and retrieving information than about organizing and arranging documents and, in this new post-custodial paradigm, access to information is at the heart of the work and concerns of information professionals. Thus, it is clear that it no longer makes sense to use classification and other languages differently in archives and libraries, and there is no justification for the training of professionals who will work in these two types of information services (or even other bodies such as museums, documentation centres and information management services within a wide range of organizations) to differ, far less with regard to the technical component of the organization and representation of information. This is also a natural consequence resulting from training in information science, taught at the University of Porto, which assumes the epistemological unity of the area without artificial distinctions dictated by the rationale of professional contexts (David 2008; Castro et al. 2011). This unified education aims to train information professionals and managers able to work in any organizational context, information flow generator and, as such, embraces in an integrated manner all the technical component devoted to knowledge organization.

Notes

1. There is also an entry entitled clasificación de fondos documentales de archivo and another on clasificación funcional de series documentales (López Yepes 2004, 325).
2. On this issue, see Ribeiro (1998) to show to what extent the use of classification in archives differs from what is common practice in libraries and documentation.
3. The study cited was the topic of Ribeiro (2003).

References


ISKO and Knowledge Organization’s 25th Anniversary: The Future of Knowledge Organization and ISKO Panel Discussion

Reported by
Rebecca Green*


The main idea of this panel was to create a platform for discussing knowledge organization in the past, present, and future within ISKO. During the panel discussion the following three questions were asked: 1) What is knowledge organization (KO)? 2) What changes do you foresee in the future that will prove to be the most challenging for ISKO? 3) What is your ideal picture of what the ISKO of the future could be? How do we get there?

1.0 What is knowledge organization?

Rebecca Green: I will lead us off with two insights. First, according to the ISKO charter, “It is the aim of the Society [ISKO] to promote research, development and application of all methods for the organization of knowledge in general or of particular fields by integrating especially the conceptual approaches of classification research and artificial intelligence. The Society stresses philosophical, psychological and semantic approaches for a conceptual order of objects.”

Second, in Dewey the rule of application instructs us to class a work on, say, a thesaurus of architecture—that is, the making of a thesaurus applied to architecture—with other works on architecture. But developing a thesaurus on architecture doesn’t make the developer an architect. Against that backdrop, what is knowledge organization?

Claudio Gnoli: What is KO as seen from the perspective of other people? There is a lot of work nowadays that actually is KO but is called with other names by the com-
munities of, e.g., ontologists, taxonomists, terminologists, topic maps experts, information architects, etc. Unfortunately KO is often not identified as a field in itself, maybe because it deals with such basic logical components of knowledge (classes, hierarchies, terms, etc.) that people take them for granted.

We as ISKO should include these communities in the discussion, but in order to achieve this we should adopt some common, consistent terminology, quite like Ranganathan did within his own works when writing about devices, canons, etc. I for one am trying to adopt the term dimensions of KO (that is the ontological, epistemological, pragmatic, etc., dimensions) in the same sense as previously used in Tennis's and Hjørland's papers, although arguing different things.

Ingetraut Dahlberg: When we founded ISKO we needed a new name, other than Society for Classification, since we had just left the German society with this name because it became a society of mathematicians. We thought of Bliss’s two books but considered the expression “organization of knowledge” too long in the name of a society and changed it—according to the German way of expressing such combinations—into knowledge organization. To our astonishment, it was accepted worldwide. However, we just meant it as another name for order activities in classification. Our journal International Classification had a current bibliography of relevant titles. The classification system for this bibliography remained the same when we renamed the journal Knowledge Organization. I referred to this and explained its structure in my articles of KO 2006 and 2014. I consider knowledge organization as a subdiscipline of the science of science with application fields not only in the information sciences but also for all subject fields (domains) needing taxonomies (classification systems of objects) and other fields like statistics, commodities, utilities, weapons, patents, museology, etc. According to science theory, every domain has its own area of objects and of methods and processes, next to other relationships. In knowledge organization one expresses the objects by “all kinds of knowledge” and the methods by “organization” in the sense of creating order of the given kinds of knowledge and its activities.

Peter Ohly: ISKO’s declared aims have to be seen historically and structurally. In its origin it emerged from library science cataloging. The forerunner GiKl (German Society for Classification) was founded in contrast to the DGD (German Society for Documentation) as a society with more theoretical and methodological orientation and with less stress on documentation praxis. When at least in this society the statisticians got the majority, ISKO was founded by the non-statistician part as a society with less orientation to business informatics. Thus far ISKO has still today problems to get connections to more computer-oriented neighbor fields, like knowledge management.

Furthermore ISKO has problems to claim a focus that is not already occupied by other established scientific neighbor communities, e.g., artificial intelligence, neuroscience. Thus ISKO has to define explicitly its boundaries to know with whom and how to interact and not to compete with. Such restrictions for ISKO as a society do not apply in the same determination to KO as a theoretical field, as a university discipline, or as a journal focus.

The main problem for ISKO as a society is that it does not attract a well-defined profession: “Knowledge organizer” or “Semantic worker” are not established professions. Accordingly the community of KO must be more moderate and realize that it has its main application and acceptance in library science. I would describe the focus of KO more general as arranging of knowledge, instead of ordering, classification or organization. How far extraction, connection, reasoning, or interpretation of knowledge should be also included in the focus of ISKO should carefully be considered and optionally be denied. Surely the economic dimension of valuable information, the scientific background of the applied field, and the sociological aspects are missing in the definition of its approaches. The latter is important for studying the acceptance of systems as well as the social dynamics of use and misuse, not at least in social software models.

Wiesław Babik: There are many definitions of knowledge organization. Although they are not always fully coherent, after summing up they define the content and the range of this notion. The main difficulty in defining knowledge organization—in my opinion—is the fact that all the definitions are based on two other notions which are very often defined only in an intuitive way: the notions of knowledge and organization. This way we produce a kind of etymological definitions. The answer for the question asked at the beginning depends on the way we understand what knowledge is and what organization is. Besides, one should always remember that the subject (topics) of knowledge organisation is composed of knowledge and information seen as raw material for knowledge, but understood from the perspective of its organization (viewpoint). This locates knowledge organization among other scientific disciplines concerned with various aspects of knowledge. Finishing I would like to express my belief that knowledge organization is a very important inter- and multidisciplinary domain, indispensable to science, education and research, which is also practiced within modern information science.
Amos David: The way knowledge is acquired, represented, managed and exploited has changed with the connected world and the functionalities associated. I would suggest that these issues related to knowledge in the digital world should be included as an object of research.

Vera Dodebei: Although cultural approaches may be included in the semantic conceptual frame, ISKO should increase discussions on cultural aspects, considering, for example, the connected societies.

Grant Campbell: Taking off from your observation that we can create a glossary of architecture, for example, without being labeled architects, I suggested that KO’s greatest strength lies in its position outside of domains and its capacity for working between them. Whereas disciplines and domains tend to focus on cultivating their own terms and their own practices and traditions and canonical texts, KO has a more itinerant role: rather like the traveler who travels from town to town, and in exchange for a meal by the fire, relays to the inhabitants stories of other places.

Dagobert Soergel: Among the ISKO membership there is a great deal of expertise in principles of knowledge organization. But most ISKO members come out of a library and bibliographic systems tradition. This presents somewhat of a barrier to bring this expertise to the much wider arena where it is applicable and where it would be beneficial. Documentary information is just a part of the information landscape. There are other applications of high importance, such as, to name a few,

- electronic health records (EHR) connected to patient treatment and medical research on the one hand and medical billing on the other hand
- scientific data
- research networking systems
- business information systems
- linked data as a format

To enable transfer of ISKO expertise into these wider application areas and the associated communities requires a re-orientation. ISKO members need to work in other areas as illustrated, for example, in the paper by Maria Lopez-Huertas. We need to invite speakers from other communities to ISKO conferences (and actively solicit contributed papers which would be subject to the review process). At the same time we need to make sure that there is a sufficient number of papers of interest to members of other communities.

Claudio Gnoli: ISKO resources that can help a more clear identification of what is covered by KO are 1) our journal, 2) the online KO literature, 3) the dictionary-glossary of KO that we are planning to publish on our website. Also when browsing mailing lists in order to feed the online list of coming KO-related events I often have to decide whether they are so or not, and I especially refer to topics listed in their calls for papers: I take that a KO event, independently from the technical means and carriers it addresses, should have to do in some way with the subject content of documents, in Buckland’s broad sense of document [see his keynote at this conference] so clearly not just in libraries as feared by Soergel. I agree that developments of the semantic web should be considered and included in KO. While most current work in the semantic web is concerned with technical interoperability, KO should contribute as for conceptual interoperability (conceptual mapping, SKOS, OWL, etc.).

2.0 What changes do you foresee in the future that will prove to be the most challenging for ISKO?

Vera Dodebei: I see a number of questions we must address: 1) How will KO researchers consider the challenge of accessing memory data (traces) and heritage representations from collective knowledge in the internet? May we consider forgetting the default for memory? 2) What may be a sustainable world considering the information and knowledge society? 3) How can we keep our collective memory knowledge safe in an unstable space always in movement? To this I raise one possible answer: we must be in connection with theories from multiple fields, especially in the domain of art, history, anthropology, archeology.

Amos David: The most important challenge that I see is the temptation of being attracted by technologically-driven research.

Ingetraut Dahlberg: If we agree that knowledge organization is a scientific discipline in its own right, we need to develop it accordingly and start with elaborating its roots. Eugen Wuester, the Austrian Terminologist in the early thirties, had paved the way by developing a series of standards for concepts (DIN 2330, etc.). I developed a new theory of the concept, published in German in the first issue of *International Classification* in 1974 and in English in the volume of the FID/CR Conference in Bombay in 1975. In later publications I extended it further, calling a concept a unit of knowledge. According to this theory a concept needs to be analyzed into its characteristics; by finding the same or similarly expressed characteristics in different concepts, true systems of concepts
can be created into the two hierarchical, and the complementary and functional relationships. This I have shown in many publications already in the seventies, the last one in German in 2008 and in English somewhat later in KO 2009. It is also contained in an abridged way in the article on “A Systematic New Lexicon of All Knowledge Fields” with the theories and principles of the ICC (KO 2012). I would like to add that I learned very much from the mathematician and librarian S. R. Ranganathan and consider him still fundamental for all of our students.

Peter Ohly: KO goes back to known principles of concept formation, as applied in the developing of indexing systems. But instead of seeing it mostly in the field of cataloging it must be more open to realize that there are new applications, new knowledge sources and quite other applications, like virtual knowledge generation, mobile devices, decision making, evaluation indexes. We often speak of literacy and think of KO literacy for the users of other disciplines. Instead we should realize that we have permanent need for understanding new upcoming techniques and thinking in neighboring fields. Openness can be reached by tutorials, workshops, and co-operation that broaden our thinking in neighboring fields. Openness can be reached by tutorials, workshops, and co-operation that broaden our understanding and applicability of neighboring disciplines, specialized areas, and other cultures.

Grant Campbell: Two challenges will confront KO in the future, as it will other disciplines: 1) The problem of cultural and community differences: Canada is currently undergoing considerable stress from its earlier catastrophic treatment of indigenous peoples, in particular around education and, by implication, organizations and structures of knowledge. KO will be urgently needed by many countries and cultures that are trying to negotiate the demands of different cultures, and trying to step outside the usual “one size fits all” approach to enforcing uniformity. 2) The need to develop sustainable economies and cultures will require major changes in KO: changes that will enable all of us to think more easily and clearly in terms of sustainability. ISKO 2016’s announced theme couldn’t be more appropriate or timely.

Jill McTavish: One challenge I see for ISKO’s future is how to better incorporate new, different, and upcoming voices. It’s much along the line of what Grant Campbell and José Augusto Chaves Guimarães have said. If one were to examine the discourse of our conference there would be only a few voices that dominated throughout. Why don’t the students feel comfortable commenting? Why are there only a handful of people that felt comfortable commenting throughout the conference? How can those that feel comfortable talking try to create spaces for others to speak, and how can ISKO structure different voices into its programming? For example, a student representative could have been on both of the panels offered at ISKO 2014. ISKO 2016 could also offer partial conference scholarships to new students—perhaps through an essay contest, lottery, or something.

Wiesław Babik: I am absolutely convinced that today we should promote a network approach to knowledge organization, both in its theoretical and conceptual dimensions as well as in practical ISKO activities. For the last 25 years we have been able to observe an explicit tendency toward automation, globalization and socialization of information and knowledge creation processes, including knowledge organization and ISKO activities. ISKO has noticed those changes and has been trying to take them into consideration in its deliberations and activities. But there are also some threats that should be noticed. In such a situation, both ISKO and KO should be more human-oriented and sustainable development driven, and these should be the most important challenges for the next 25 years. Knowledge organization and ISKO cannot forget about human beings, which seems to be quite common in the present world.

3.0 What is your ideal picture of what the ISKO of the future could be? How do we get there?

Amos David: ISKO has a sound foundation. This should be maintained and reinforced. ISKO conferences constitute an excellent forum for scientific exchanges and Knowledge Organization has a very good reputation. To maintain its level of recognition, the community should remain focused on scientific objects rather than technology-dependent issues.

Laura Ridenour: I need to make a disclaimer: I don’t have a library background, and I hope I won’t be lynched for what I am about to say. I would like to suggest that knowledge organization needs to consider an open access model of publication to provide access for people who may be interested in KO, but are not part of the community. My reasoning is that Knowledge Organization (the journal) is not widely accessible; knowledge organization literature is both difficult to locate and misindexed in databases such as LSTA, usually placed under knowledge management; (related to my first point) that in order to be relevant, we must contribute to scientific literature in a forum in which more people are able to access our bodies of work; and we need to collaborate with individuals in other research specialties.
Wiesław Babik: In my opinion both ISKO and Knowledge Organization will benefit from the implementation of the idea of information and knowledge society, because this process demands high quality information and knowledge. Knowledge organization can help to regain control over a chaotic information world, both literally and metaphorically, especially when supported by knowledge and information ecology. This new domain has already found its place in academic teaching within study programmes connected with information and in research. In ISKO activities one can separate three basic levels: international, national and local. All of them should be developed in a balanced way. This should be better represented by the ISKO structure. The structure of the Polish chapter could be seen as an interesting example.

Peter Ohly: ISKO should become a virtual institute where from everywhere one can be connected with specialists in KO and get their advice. This should not only be performed in a KO-pedia style but also as an e-science, where projects are performed virtually with scientists, coming as well from other disciplines. KO managers should not only know what is the best KO but as well be able to explain and elaborate the differences, strengths and weakness of special KO approaches in special applications.

Dagobert Soergel: Ideally, ISKO would develop into a society that covers KO issues in a wide range of applications, with keen attention to common principles, and that attracts people focusing on KO from many communities, serving as a common meeting point for the transfer of basic knowledge and of reusable modules in the development of KO systems. Along similar lines, ISKO should get involved in formulating information literacy standards so these standards incorporate not just surface skills in searching for information but deeper understanding of principles of knowledge organization that will make people much better searchers. Within KO courses in library and information science programs, we need to make students keenly aware of the wide range of KO applications, the areas, and widely used systems, such as

- CYC Ontology
- WordNet
- Gene Ontology (GO)
- SnoMed
- ICD-10 (The International Statistical Classification of Diseases and Related Health Problems, tenth revision)
- NAICS (North American Industrial Classification System)
- HS (Harmonized Commodity Description and Coding System. World Customs Organization)
- Bloom: A taxonomy for learning, teaching, and assessing: A revision of Bloom’s taxonomy of educational objectives
- AAT (The Art and Architecture Thesaurus)
- Iconclass

This extension of the range should also be pursued for the journal KO. A wider range and higher quality of submissions would ideally elevate the recognition and prestige of the journal, increase readership (readers from many different communities) and elevate KO into a first-tier journal. Finally, it would be useful to create a list of associations, conferences, and separate listservs that deal with KO and also repositories for KOS. Some I can think of are:

- ASIST SIG-CR: Organizes a one-day workshop at ASIST Annual Meeting
- IAOA (International Applied Ontology Organization): Organizes FOIS (Formal Ontology for Information Systems) conference
- NKOS
- Ontolog-forum: Organizes the yearly Ontology Summit
- Yearly ontology matching workshop
- ICBO (International Conference on Biomedical Ontologies): every other year.
Book Review
Edited by Hanne Albrechtsen


Never would I have ever thought to write a review about a book from a computer scientist. However the title provoked me, I wanted to know what ‘they’ know about our field of knowledge organization. And here it goes.

For this handbook of 570 pages, the compiler, Miguel-Angel Sicilia from the University of Alcala, Spain, assembled 20 contributions from 36 authors of 13 countries (Austria, Belgium, Brazil, Chile, Ecuador, England, Germany, Greece, Ireland, Italy, Mexico, Spain, and the USA). He divided this collection into 4 sections: (I) four (plus one) contributions on metadata, (II) four on ontologies, (III) eight on applications in special domains, and (IV) three on technologies for interoperability, metadata extraction and metadata and ontology storage.

Metadata—an expression coined already in 1968 by Philip Bagley (Wikipedia) is being understood as data about data, and according to Sicilia is “a generic term of any kind of meta-information.” With the creation of the Web, “metadata has been brought to the hearth of the architecture of cyberspace.” The expression media for metadata today are XML and RDF. In the language of KO experts, metadata may be understood as concepts: for information scientists as descriptors; for taxonomists as taxa; and for librarians as subject headings. In the meantime the latter adopted already ‘metadata,’ apparently for its simplicity and possible collaboration with computer people, the same happened—but in a much wider sense—with the term ‘ontologies’ used for classification systems, taxonomies, thesauri, vocabularies and the like, together they are termed by computer people ‘semantics.’ Although absurd, as ‘ontology’ is a subdiscipline of philosophy, this expression in the plural has made in the meantime a triumphal procession into all domains working towards organizing knowledge elements into systems—with the exception of the taxonomists.

With his first contribution, Chapter I.1, on “Metadata Research: Making Digital Resources useful again,” A.-M. Sicilia set the frame for the handbook and called ‘metadata’ a research discipline, and by trying to define it more extensively, referred to different forms of metadata today. Incidentally he deplores “the lack of a pertinent professional society or organization”—apparently he is not as yet aware of the existence of ISKO, the International Society for Knowledge Organization. Finally he refers to the newly started “Linked Data ... as a simple but extremely powerful concept (sic!) that overcomes limitations of microdata and XML-based harvested collections.” They provide “the fundamental properties to build a coherent global system of metadata in the coming years ... they relate useful information to other information ... with RDF as implementation technology.” (For an example of this possibility see the reference at the end.)

Chapter I.2 deals with “Metadata Typology and Metadata Uses;” by Eva Mendez from Madrid and Seth van Hooland from Brussels. They provide an overview of how metadata can be interpreted differently and give a typology of metadata schemes and elements and how they can be deployed by user communities. The authors of Chapter I.3, M.D. Lytras from Greece, A.-M. Sicilia (Spain) and C. Cechinel from Brazil wrote about “The Value and Cost of Metadata” considering as metadata only those data which are in digital form and which are a subject to be used through communication technologies, managed through software to enhance information retrieval as an alternative or complement to text-based IR algorithms and formal bibliographic records. Their value and costs are dependent on the functions which vary in the many different sciences using them; they may increase usefulness and create benefits. Table 1 shows examples of metadata systems and their typical characteristics with their quality and evaluation policy as well as sources of value. In Table 2 some metadata cost examples are given.

Chapter I.4 by X. Ochoa from Ecuador is about “Metadata Quality,” presenting the main theoretical approaches to the definition of metadata quality, so far poorly researched because of inherent complexity and subjectiveness. A selection of strategies and tools to evaluate the quality of metadata is given.

Chapter I.5 on “Ontologies in Systems Theory” by E. Curras from Madrid should belong in the next section. It aims at applying systems theory to ontologies, which are
to be understood as “complex, conceptual, empirical classification systems,” where the “principal and primary node is the word.”

Section II, as said before, to be meant on ontologies, starts with Chapter II.1 by L. Papaleo from Italy and gives an “Introduction to XML and its Applications.” The abstract says: “Extensible Markup Language (XML) is a meta language for defining new languages. Its impact on the modern and emerging web technologies has been and will be incredible.” I must confess, this article was most informative and helpful for me, also as it contains the explanation of very many respective abbreviations. It explains what XML is, its origin in SGML and the application of the latter in HTML. Further headings are: The birth of XML, Syntax of XML documents, Name spaces, Structuring XML documents, Element declarations, XML schemas, Rendering XML documents via CSS (Cascading Style Sheet), Transforming and rendering XML documents.

Chapter II.2 by S. Arroyo from Spain and K. Siopraes from Austria in “Ontologies and Ontology Languages” give a general overview on approaches to model metadata by controlled vocabularies, taxonomies, thesauri and ontologies. These systems are explained in order to understand their later possible change into a form that is interpretable for the computer. For this, so-called ontology languages are necessary, the most relevant ones are: RDF(S), Resource Description Framework, a general purpose language to represent information about resources in the Web; OWL, Web Ontology Language, which exists in three versions: WSML, Web Service Modeling Language, which has four versions; SKOS, Simple Knowledge Organization System which stands for “all kinds of controlled vocabularies describing the minimal set of classes and properties that are necessary to express knowledge in simple structures such as taxonomies” and finally OIL built on RDFS, but superseded by the OWL family of languages.

Chapter II.3 by P. G. Picazo and J. Tramulas, both from Spain, deals with “Topic Maps.” These have already become an ISO standard, reflecting the importance of this tool to describe and organize digital information resources. Chapter II.4 has three authors from Germany: T. Bürger, E. Simperl and Ch. Tempich: “Methodologies for the Creation of Semantic Data.” They provide an overview of knowledge engineering methodologies for knowledge-based systems as well as their tool environments supporting their application. Industrial take-ups of these semantic technologies seem possible.

For Section III with applications of metadata and ontologies in eight special domains M. G. Sicilia, the compiler, selected some rather unusual but also very voluminous domains. Here we just mention the authors and their titles: Chapt. III.1: M. E. Prieto. Mendez (Spain), V. H. Menendez Dominguez (Mexico) and Ch. L. Vidal Castro (Chile): “Metadata and Ontologies in E-Learning.” Chapt. III.2: G. Colomb., D. Merico, M. Güntel (Italy): “Metadata and Ontologies for Health.” Chapt. III.3: N. Palavisinis, N. Manouselis (Greece): “Agricultural Knowledge Organization Systems: An Analysis of an Indicative Sample.” Chapt. III.4: E. Blanco (Spain): “Metadata and Ontologies for Bioinformatics.” Chapt. III.5: F. Sarton, S. Bandini (Italy): “Metadata and Ontologies for Mechanical Objects: Design and Manufacturing.” Chapt. III.6: L. Santos-Santos, T. Aguado-Goméz, (Spain): “Metadata and Ontologies for Emergency Management.” Chapt. III.7: D. Kanellopoulos (Greece): “Metadata and Ontologies for Tourism.” Chapt. III.8: T. Bürger (Germany), M. Hausenblas (Ireland): “Metadata Standards and Ontologies for Multimedia Content.” In almost all of these articles it was astonishing for me that the work of information scientists in elaborating thesauri and classification systems had apparently been very useful instruments for the work of these authors. In the case of Chapt. III.2 “Health,” the authors gratefully acknowledged the work towards a Unified Medical Language System (UMLS), especially also for the definitions of their terms!

The three chapters of the following Section IV appeared to me as most valuable for every newcomer into this new knowledge-field for information scientists, taxonomists and librarians. Indeed, their authors with their references reveal a familiarity with “our business.” Chapt. IV.1: R. Eito-Brun, (Spain): Technologies for Metadata Integration and Interoperability. Chapt. IV.2: K. Golub, H. Muller, E. Tonkin (England): Technologies for Metadata Extraction. Chapt. IV.3: M. Parmelee, L. Obst (USA): Technologies for Metadata and Ontology Storage.

Doubtlessly, this Handbook of Prof. Sicilia is a laudable undertaking and must have been really hard work. There was apparently no time left for a subject index or a list of abbreviations with their interpretation. Laudable also is the number of references to each article. I took the “pain” to count all of them. If I am not mistaken, I found altogether 963! A really rich little bibliography of this field!

The volume was written by and for computer science people (with two exceptions). In some cases they cited at the end the grants by which they were able to do the investigation. Not only with regard to the last three chapters but in some way also for the others, it might serve our community quite well. Unfortunately the whole undertaking came to appear only in January 2014 and although I ordered it already in January, I received the review copy in April. And it must have taken five years to assemble, write and check all the contributions, visible in the year indications of the references, the latest ones are from 2009! Therefore a number of important developments since then could not be included.
It seems to me, a new way of possible acquaintance and collaboration between knowledge organization and knowledge engineering communities ought to be taken into consideration, perhaps by textbooks written from each side for the other side. The three last chapters of the Handbook would pave the way.

Reference


Ingetraut Dahlberg
Bad König, May 29, 2014


Reference citations within the text should have the following form: (Author year). For example, (Jones 1990). Specific page numbers are required for quoted material, e.g. (Jones 1990, 100). A citation with two authors would read: (Jones and Smith, 1990); three or more authors required for quoted material, e.g. (Jones 1990, 100). Specific page numbers are re- quired for quoted material, e.g. (Jones 1990, 100). A citation with two authors would read: (Jones and Smith, 1990); three or more authors would be: (Jones et al., 1990). When the author is mentioned in the text, only the date and optional page number should appear in parenthesis – e.g. According to Jones (1990), …

References should be listed alphabetically by author at the end of the article. Author names should be given as found in the sources (not abbreviated). Journal titles should not be abbreviated. Multiple citations to works by the same author should be listed chronologically and should each include the author’s name. Articles appearing in the same year should have the following format: “Jones 2005a, Jones 2005b, etc.” Issue numbers are given only when a journal volume is not through- paginated.

Examples:


Illustrations should be kept to a necessary minimum and should be embedded within the document. Photographs (including color and halftone) should be scanned at a minimum resolution of 600 dpi and saved as .jpg files. Tables and figures should be embedded within the document. Tables should contain a number and title at the bottom, and all columns and rows should have headings. All illustrations should be cited in the text as Figure 1, Figure 2, etc. or Table 1, Table 2, etc.

The entire manuscript should be double-spaced, including notes and references.

Upon acceptance of a manuscript for publication, authors must provide a wallet-size photo and a one-paragraph biographical sketch (fewer than 100 words). The photograph should be scanned with a minimum resolution of 600 dpi and saved as a .jpg file.

Advertising

Responsible for advertising: ERGON-Verlag GmbH, Keesburgstr. 11, 97074 Würzburg (Germany).
© 2014 by ERGON-Verlag GmbH.
All Rights reserved.

KO is published bi-monthly by ERGON-Verlag GmbH.
- The price for the print version is € 259,00/ann. including airmail delivery.
- The price for the print version plus access to the online version (PDF) is € 289,00/ann. including airmail delivery.
Scope

The more scientific data is generated in the impetuous present times, the more ordering energy needs to be expended to control these data in a retrievable fashion. With the abundance of knowledge now available the questions of new solutions to the ordering problem and thus of improved classification systems, methods and procedures have acquired unforeseen significance. For many years now they have been the focus of interest of information scientists the world over.

Until recently, the special literature relevant to classification was published in piecemeal fashion, scattered over the numerous technical journals serving the experts of the various fields such as:

- philosophy and science of science
- science policy and science organization
- mathematics, statistics and computer science
- library and information science
- archivistics and museology
- journalism and communication science
- industrial products and commodity science
- terminology, lexicography and linguistics

Beginning in 1974, KNOWLEDGE ORGANIZATION (formerly INTERNATIONAL CLASSIFICATION) has been serving as a common platform for the discussion of both theoretical background questions and practical application problems in many areas of concern. In each issue experts from many countries comment on questions of an adequate structuring and construction of ordering systems and on the problems of their use in opening the information contents of new literature, of data collections and survey, of tabular works and of other objects of scientific interest. Their contributions have been concerned with

1. clarifying the theoretical foundations (general ordering theory/ science, theoretical bases of classification, data analysis and reduction)
2. describing practical operations connected with indexing/classification, as well as applications of classification systems and thesauri, manual and machine indexing
3. tracing the history of classification knowledge and methodology
4. discussing questions of education and training in classification
5. concerning themselves with the problems of terminology in general and with respect to special fields.

Aims

Thus, KNOWLEDGE ORGANIZATION is a forum for all those interested in the organization of knowledge on a universal or a domain-specific scale, using concept-analytical or concept-synthetic approaches, as well as quantitative and qualitative methodologies. KNOWLEDGE ORGANIZATION also addresses the intellectual and automatic compilation and use of classification systems and thesauri in all fields of knowledge, with special attention being given to the problems of terminology.

KNOWLEDGE ORGANIZATION publishes original articles, reports on conferences and similar communications, as well as book reviews, letters to the editor, and an extensive annotated bibliography of recent classification and indexing literature.

KNOWLEDGE ORGANIZATION should therefore be available at every university and research library of every country, at every information center, at colleges and schools of library and information science, in the hands of everybody interested in the fields mentioned above and thus also at every office for updating information on any topic related to the problems of order in our information-flooded times.

KNOWLEDGE ORGANIZATION was founded in 1973 by an international group of scholars with a consulting board of editors representing the world's regions, the special classification fields, and the subject areas involved. From 1974-1980 it was published by K.G. Saur Verlag, München. Back issues of 1978-1992 are available from ERGON-Verlag, too.

As of 1989, KNOWLEDGE ORGANIZATION has become the official organ of the INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO) and is included for every ISKO-member, personal or institutional in the membership fee (US $ 55/US $ 110).

Rates: From 2013 on for 6 issues/ann. (including indexes)
€ 295.00 (forwarding costs included) for the print version resp. € 289.00 for the print version plus access to the online version (PDF). Membership rates see above.

ERGON-Verlag GmbH, Keesburgrstr. 11, D-97074 Würzburg; Phone: +49 (0)931 280084; FAX +49 (0)931 282872; E-mail: service@ergon-verlag.de; http://www.ergon-verlag.de

Founded under the title International Classification in 1974 by Dr. Ingegerdt Dahlberg, the founding president of ISKO. Dr. Dahlberg served as the journal’s editor from 1974 to 1997, and as its publisher (Indeks Verlag of Frankfurt) from 1981 to 1997.

The contents of the journal are indexed and abstracted in Social Sciences Citation Index, Web of Science, Information Science Abstracts, INSPEC, Library and Information Science Abstracts (LISA), Library, Information Science & Technology Abstracts (EBSCO), Library Literature and Information Science (Wilson), PASCAL, Referativnyi Zhurnal Informatika, and Sociological Abstracts.