I Simposio Internacional sobre Organizacion del Conocimiento, Bibliotecologia y Terminologia:

An Editorial

Richard P. Smiraglia, Editor-in-Chief

Knowledge Organization (KO) as a domain is evolving rapidly and its boundaries are being pushed amoeba-like in every direction as a consequence. All that readers of this journal need to do to find evidence of this evolution is to look at the journal itself, which moves from quarterly to bi-monthly with this issue. While the peer-review system serves a gate-keeping function on the intension of the domain, making certain that articles appearing in this journal align with accepted conceptual tenets, the system simultaneously serves a different function (perhaps we can align it with the opening of gates) for KO conferences, where it is the extension of the domain that is constantly being probed and tested by new research. Gate-keeping is an important function for any domain, which is why peer review is a hallmark of ISKO's regional and international conferences as well as this journal. So it is even more impressive to consider these two functions together, which at once serve to intensify the core concepts of knowledge organization and simultaneously to stretch their application into new corners of the knowledge domain. It is a sort of inspiration-expiration dichotomous action, solidifying the core on the intake and pushing the boundaries (or axes, as Tennis (2003) has called them) on the outgo. Indeed, the new "Forum: Philosophy of Classification," and occasional feature beginning with this issue, which has been generated by Birger Hjørland, chair of ISKO's newly active Scientific Advisory Council, is an example of this dichotomous action. For further examples we can turn to the contents of regional and international KO conferences, which provide interesting temporal glimpses of this evolutionary process.

In prior editorials I have applied domain-analytic tools to the proceedings of several KO conferences (Smiraglia 2008; 2007; 2006), and in a 2009 paper I used the same tools to create snapshots of KO from both North American and non-North American perspectives. At base the central research questions always are the same: what are the parameters of the intension and extension of KO as reflected in the particular conference? Sub-questions, of course, arise around clustering of research fronts and the reach (or impact, if you will) of segments of the common KO literature. In this editorial I will present a summary analysis of the August 2007 Mexico City conference

"I Simposio Internacional sobre Organizacion del Co-nocimiento, Bibliotecologia y Terminologia," whose proceedings were just available in print in 2009.

1.0 The Simposio

This conference, announced as the First International Symposium on Knowledge Organization, Library Sci-ence, and Terminology, was held in August of 2007. The conference was held at UNAM's Centro-Universitario de Investigaciones Bibliotecologicas, whose headquarters are in magnificent space at the top of a tower providing exquisite views of the rest of the university and the city to the north. Conference or-ganizer Catalina Naumis Peña graciously invited me to participate, and I was delighted to do so (the photo shows the neighboring parallel tower Torres II Hu-manidades). I was very graciously allowed the option to present in English. My ISKO colleague (then presi-dent) Maria-José López Huertas was gracious enough to sit with me at the lectern and translate for me so the audience could hear my talk in Spanish. Like most of the conferences I analyze in this space, I was a par-ticipant, which might color my interpretation. The conference was intense—42 papers over 3 days (38 of which appear in the proceedings). The geographical reach of the conference was impressive; papers came from authors affiliated with institutions in 7 countries, the majority from Mexico (Table 1). Twenty-one pa-pers came from North America, 13 from South or Central America, and 6 from Spain. Only one paper (mine, of course) was given in English; all the rest were given in Spanish and are in Spanish in the pro-ceedings.

Country of Origin	Number of Papers
Mexico	16
Spain	6
Brazil	5
United States	5
Columbia	4
Costa Rica	3
Uruguay	1

Table 1. Countries of origin

2.0 Citation analysis

Analysis began with citation counts per paper, per country of origin, and per thematic cluster. This was to see whether any geographic or cultural emphases might emerge. The mean number of references per

paper was 10.975, with a range from 4-47. The mean varied from country to country (Table 2). The mean number of citations per paper was highest in Uru-guay and lowest in Mexico. The means per broad subject area were calculated using the eight confer-ence panel titles. The mean number of citations per paper was highest for informatics and lowest for se-mantics and digital standards.

Country of origin	Mean	Thematic cluster	Mean
Uruguay	24	informatics in informa-tion retrieval	18.6
Costa Rica	16.3	representation and in-formation retrieval	13
Spain	15.83	translation for informa-tion transfer	12.6
Brazil	13.8	terminology in librari-anship	11.6
United States	9	terminological analysis	9.75
Columbia	8.25	terminology and differ-ent areas of knowledge	9.2
Mexico	7.75	standards and use of lan-guage in a digital envi-ronment	6
		semantics of information	5.6

Table 2. Mean references per paper per country and the-matic cluster

The mean number of citations per paper per country and per thematic group were tested using one-way ANOVA; in both cases there was no statistical sig-nificance observed, meaning the different means likely are the result of chance and therefore are not meaningful individually. One interpretation of the differences observed might be that they are depend-ent in every case on the particulars of the presenta-tions, and therefore reflect the overall variability in referencing observed for the entire conference, which hovers roughly between the mode of 5 and mean of 11.

A cross-tabulation of country with theme sug-gested independence, but it probably is wise to be suspicious of the role of chance because most cells have low numbers. There was relatively strong inter-est in terminology and ontology, and in representa-tion and language, with smattered interest in seman-tics, informatics, and digital standards. The contribu-tions from Brazil were ontologically and representa-tionally focused, from Columbia and Costa Rica the other thematic areas were the focus (Figure 1).
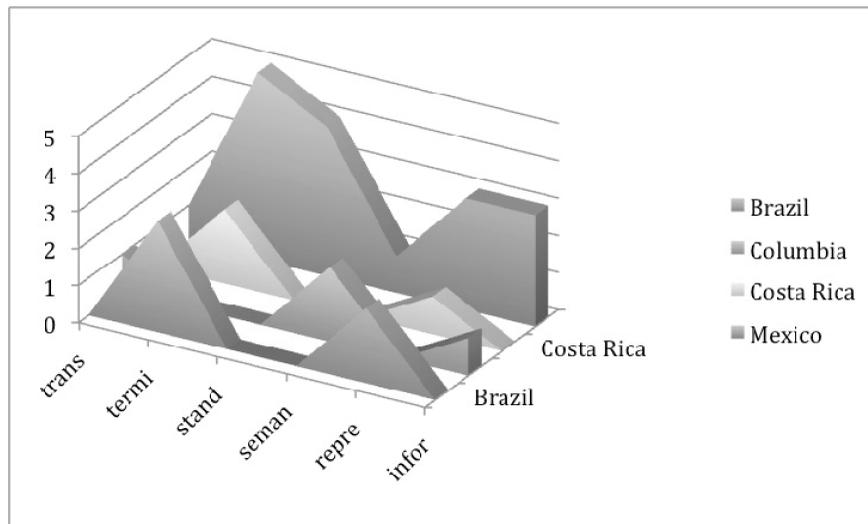
Figure 1. Themes per Country of origin

Age of cited works reveals clues about the immediacy of the domain, primarily by showing whether there is rapid absorption of new research or alternatively, whether there is a large quantity of classical material in the theoretical base. Typically information studies and its subdomains (including knowledge organization) resemble social scientific domains, falling somewhere between the two extremes. Papers cited by authors of the Mexico City conference had a mean citation age of 7 years, which places this group within the usual social-scientific range. Price's Index (the percentage of citations to articles ≤ 5 years older than the citing article) was 44%, which is consistent with a "hard" science. Interestingly, only 409 of the 442 citations contained a date, roughly 7% of the citations were undated. The majority of these appear to be references to informational websites, which, in all likelihood, should not have been cited (because they are not, strictly speaking, citations to source material). The mean age of citation ranged from 2 years to 18 years. There were a number of interesting near-anomolies—for example, works by both Cutter and Dewey dated in the 1960s. For the most part older cited works occurred in papers by librarians reporting metadata or cataloging standards. There were no statistically significant differences in citation age by either country of origin or thematic grouping. That tells us that there is a fairly consistent social-scientific rate of absorption of literature for this domain.

3.0 Most cited authors

The next step was to discover the list of most-cited authors in the domain. Names of these individuals are indicative of the research front in a domain, but in this case they also likely will tell us something about the fit between KO in general and KO in Latin America. There were 442 citations in the 41 papers, which were arrayed alphabetically by author (three papers had no citations). When single-occurrence authors were removed from the distribution 150 citations remained, meaning 292 had been single-occurrence citations. The 50 remaining multiply-cited authors were arrayed in a frequency distribution, the upper tier of which is shown here (Table 3).

Hjørland, B.	8
Lima, Vania Mara Alves	8
Cabre, M. Teresa	7
Smiraglia, Richard P.	7
Kobashi, N .Y	6
Library of Congress	6
Lopez-Huertas, Maria J.	6
Naumis Pena, C.	4
Winkel, Lois	4
Aguilar-Amat, A.	3
Berners-Lee, T	3
Cintra, A. M. M.	3
Dubuc, R.	3
Lancaster, F. W.	3
Lara, Luis Fernanado	3
Moreiro Gonzalez, Jose Antonio	3

Table 3. Frequency distribution of authors cited more than once

6

Knowl. Org. 38(2011)No.1

R. P. Smiraglia: I Simposio Internacional sobre Organizacion del Conocimiento, Bibliotecologia y Terminologia: An Editorial

The table is somewhat surprising, given the breadth of authors cited in the entire distribution. For instance, from the traditional KO community we find Hjørland and López-Huertas. From the traditional cataloging and classification community we find Winkel and the Library of Congress. From the information retrieval or knowledge representation side we have Lancaster, from the Internet we have Berners-Lee. Lima, Cabre, Kobashi, and Naumis-Pena lead the list of frequently-cited Latin American authors. Interestingly, these clusters align with the subject clusters already observed.

3.1 Author co-citation analysis

Author co-citation analysis (ACA) is frequently used in domain analysis to help identify active nodes within a domain. ACA measures the perceptions of the authors who are most productive in the domain, about relationships among the researchers they cite, based on the assumption that there is some likelihood that two researchers who are co-cited might be working on similar problem sets. I have found two approaches useful in analysis of segments of the KO domain—1) ACA using a set of most-cited authors, with data derived from the entire KO domain via *Web of Science (WoS)*; and, 2) ACA using the same author set, but with data derived from co-citation counts among the authors in the conference at hand. Table 4 is an MDS plot of author co-citation using *WoS* based on the author set produced above.

What we see here is the perception of the KO domain at large—in other words, author co-citation at large—of the set of authors identified by Il Simposio contributors as key. One interpretation is that there are two large clusters, with Naumis Pena, Lara, Cintra, Moreira and Lima, at the left, and Hjørland, Smiraglia, Cabre, Dubuc, Lancaster and Kobashi in the larger cluster at the right. One Latin American, one not; one information science oriented, one librarianship oriented, but likely with the perceived research front represented by the Latin American cluster at the upper left. Another approach perhaps is warranted as well, in which we consider four clusters, two within each of the larger clusters. These represent indexing and terminology, vocabulary control, linguistics, and thesauri. In this manner we see more differentiation; either way we have a reflection, or complementary picture, of the thematic analysis from section 1.0 above—terminology and ontology, representation and language, semantics, informatics, and digital standards.

Table 5 is a plot of the same author-set as they are co-cited within the conference by participant authors. There is obviously much less co-citation, so some names drop out of the analysis.

Here we see more clearly perhaps how the domain perceives itself. Cabre's well-known and heavily cited classic work on terminology is clearly at some distance from the other two clusters, where there is some distinction between language and semantic issues (on the right) and translation and classification (on the left). Either way, we now have three pictures of the intension of the domain.
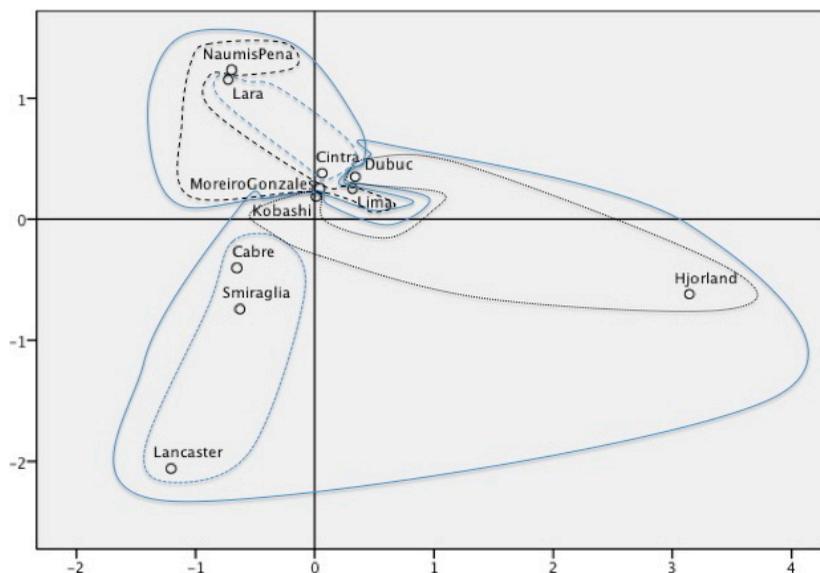

Table 4. ACA plot from *WoS*

Knowl. Org. 38(2011)No.1
R. P. Smiraglia: I Simposio Internacional sobre Organizacion del Conocimiento, Bibliotecologia y Terminologia: An Editorial

7

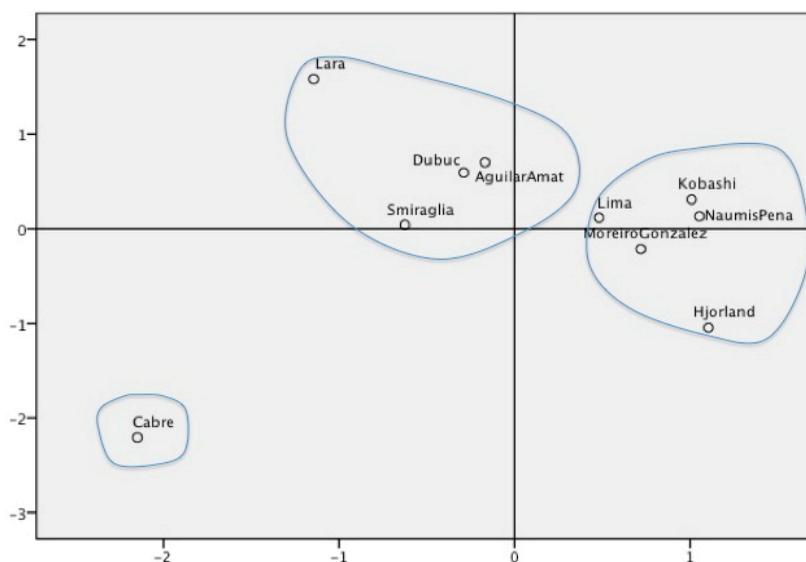*Table 5.* ACA plot based on inter-conference co-citation

4.0 Co-word analysis points to conclusions

All paper titles from the proceedings were entered into WordStat to analyze term co-occurrence. Table 6 shows the top of the frequency distribution, which presents further triangulation of the results we already have seen.

	FREQUENCY	NO. CASES	% CASES
INFORMACION	15	14	30.40%
RECUPERACION	8	8	17.40%
ESPECIALIZAR	5	5	10.90%
TERMINOLOGIA	4	4	8.70%
DOCUMENTAL	4	4	8.70%
ORGANIZACION	4	4	8.70%
INFORMACION_ RECUPERACION	4	4	8.70%
CONOCIMIENTO	4	4	8.70%
LENGUAJE	4	4	8.70%
CIENCIA	4	3	6.50%
ANALISIS	3	3	6.50%
CONTROLAR	3	3	6.50%
TERMINOLOGICO	3	3	6.50%
TERMINOLOGICOS	3	3	6.50%

Table 6. Frequency distribution of title keywords

That is, frequently occurring keywords are: information retrieval, terminology, documentation, knowledge organization and representation, and information sci-ence. Figure 2 is an MDS plot of the co-occurrence of these terms (Stress = .19; R^2 = .93).

What we see overall is a well-evolved, scientifically productive domain within the mainstream of knowledge organization. The Price's Index remarkably conforms to a hard science in its depiction of absorption of new material, which is unlike most analyses of other regional KO domains. What is most remarkable is the citation imprint, which is unlike those we have seen before, because of the large influence of Latin American authors. That is not surprising, but it is telling—we do not find those authors cited outside the region. There is quite heavy reliance on classic authors, e.g. Dewey and Cutter, and classic KOS, e.g. *LCSH*, but in translated editions, which again influences the domain's citation imprint.

With reference to our opening comments, we see that continued influence of gate-keeping through a double-blind refereed peer-review system. We see an evolution of a body of research that is linguistically bounded in Spanish and Portuguese, which is not highly cited in Western European or other North American publications, but which is, nonetheless, influential in the region. We see the incorporation of classical texts and therefore of their epistemic positions, which ground the domain in classical proportions within the defined extension of international KO. As for intensional axes, we see greater accommodation for knowledge representation, terminology, and natural language processing than is typically the case in KO at large.
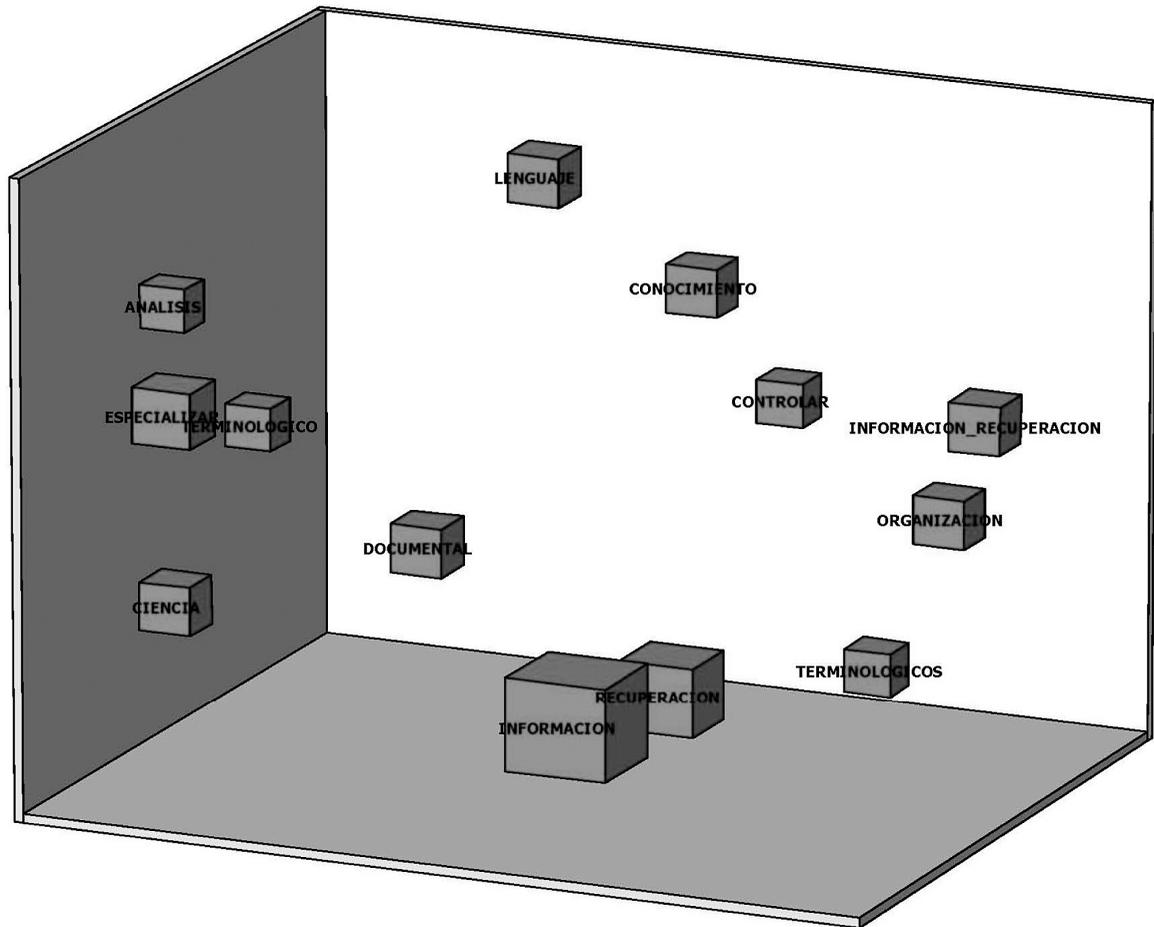
8

Knowl. Org. 38(2011)No.1

R. P. Smiraglia: I Simposio Internacional sobre Organizacion del Conocimiento, Bibliotecologia y Terminologia: An Editorial

Figure 2. MDS WordStat plot of title term co-occurrence

We congratulate our colleagues on a successful conference. But more importantly, we encourage them to continue to develop their particular epistemic brand of knowledge organization and especially to help us bring it more fully to the fore in global KO research.

(An excel spreadsheet developed as the basis of this research, including the conference program and all citations, can be found on my blog at: http://lazykoblog. wordpress.com/).

References

Smiraglia, Richard P. 2006. Whither knowledge organization? An editorial. *Knowledge organization* 33: 8-10.

Smiraglia, Richard P. 2007. A glimpse at knowledge organization in North America: An editorial. *Knowledge organization* 34: 69-71.

Smiraglia, Richard P. 2008. ISKO 10's bookshelf: An editorial. *Knowledge organization* 35: 187-91.

Smiraglia, Richard P. 2009. Modulation and specialization in North American knowledge organization: Visualizing pioneers." In Jacob, Elin K. and Kwasnik, Barbara, eds., *Pioneering North American Contributions to Knowledge Organization, Proceedings of the 2d North American Symposium on Knowledge Organization, June 17-18, 2009*, pp. 35-46 http://dlist.sir.arizona.edu/2630/

Tennis, Joseph T. 2003. Two axes of domains for domain analysis. *Knowledge organization* 30: 191-5.