

Richard P.Smiraglia, University of Wisconsin, Milwaukee

Facets as Discourse in Knowledge Organization: A Case Study in LISTA

Abstract: Knowledge Organization Systems (KOSs) use arrays of related concepts to capture the ontological content of a domain; hierarchical structures are typical of such systems. Some KOSs also employ sets of cross-conceptual descriptors that express different dimensions within a domain—facets. The recent increase in the prominence of facets and faceted systems has had major impact on the intension of the KO domain and this is visible in the domain’s literature. An interesting question is how the discourse surrounding facets in KO and in related domains such as information science might be described. The present paper reports one case study in an ongoing research project to investigate the discourse of facets in KO. In this particular case, the formal current research literature represented by inclusion in the “Library, Information Science & Technology Abstracts, Full Text” (LISTA) database is analyzed to discover aspects of the research front and its ongoing discourse concerning facets. A datasets of 1682 citations was analyzed. Results show thinking concerning information retrieval and the semantic web resides alongside implementation of faceted searching and the growth of faceted thesauri. Faceted classification remains important to the discourse, but the use of facet analysis is linked directly to applied aspects of information science.

1.0 Facets in KOSs

Most Knowledge Organization Systems (KOSs) use arrays of related concepts to capture the ontological content of a domain; hierarchical structures are typical of such systems. But synthesis, or the process of adding together complex descriptors to represent complex concepts, is also commonly employed in general bibliographic KOSs. Some KOSs also employ sets of cross-conceptual descriptors that express different dimensions within a domain. These are called facets, a term that is related to the flat surface on a geometric shape (Wikipedia), such as that of a diamond. In knowledge organization (KO) the earliest uses of facets involved the division of a set of entities into fundamental categories “which serve not only as the summa genera of classificatory structures but also as building blocks for establishing the internal structure of compound index terms or classes” (Dousa 2014, 41). In applied KOSs today, the term is used by taxonomists and information architects to describe any set of categories used for clustering and disambiguation (see e.g., Rosenfeld, Morville and Arango 2015; Hedden 2016). The use of facets in web search engines also seems to have become commonplace and is the subject of a recent increase in facet-related knowledge organization (KO) research (see e.g., La Barre 2006, Milonas 2010, Smiraglia 2017a).

The recent increase in the prominence of facets and faceted systems has had major impact on the intension of the KO domain and this is visible in the domain’s literature. An interesting question is how the discourse surrounding facets in KO and in related domains such as information science might be described. The present paper reports one case study in an ongoing research project to investigate the discourse of facets in KO. The complete study is reported in Smiraglia (2017b). In this particular case, the formal current research literature represented by inclusion in the “Library, Information Science & Technology

Abstracts, Full Text” (LISTA) EBSCOhost database is analyzed to discover aspects of the library-and-information science (LIS) oriented aspects of facets, faceted search, faceted classification and facet analysis as a research front and its ongoing discourse.

2.0 LISTA dataset

“Library, Information Science & Technology Abstracts, Full Text” (LISTA) (<https://www.ebscohost.com/academic/library-information-science-and-technology-abstracts>) indexes “600 core journals and ... 120 selective and priority journals” dating from 1960 to the present. Available to academic library users through EBSCOhost, LISTA is a core venue for information science literature, including much of the literature of KO.

For the present case study, the database was searched using the keywords “facet,” “facets,” “facet analysis,” and “faceted.” It was discovered that the database mixes the results such that the first three terms all can be retrieved in one dataset by using the keyword “facet” in the basic search feature. This search retrieved 1202 citations (searching was conducted on February 14, 2017). A second search using the keyword “faceted” retrieved 642 citations. One admitted limitation of this case study is the possibility that relevant research is not located under these terms in the database.

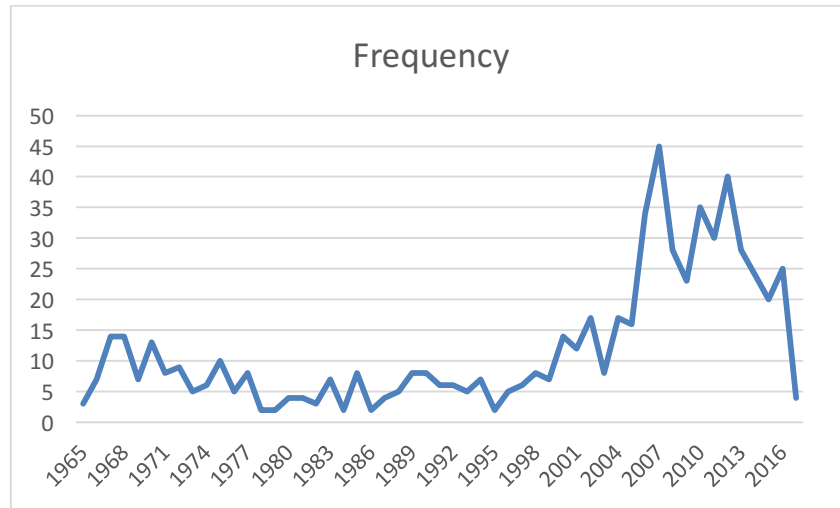
The two datasets were analyzed separately, and the combined de-duplicated list of 1682 citations was analyzed as a whole. The research is in progress as of the time of the submission of this abstract, therefore results are shown from all parts of the analysis..

3.0 Results

3.1 “Faceted”

There were 642 citations retrieved from LISTA using the keyword “faceted.” The chronological spread of the dataset is shown in Figure 1.

Figure 1. “Faceted” dataset in LISTA chronologically.



The dataset begins in 1965 and shows a moderate but decreasing rate of publication in the 1960s and 1970s, and then rapid increase from 2001 to 2013; the drop at the end likely is an artifact of the time required for publications to be indexed and enter the dataset.

There were 497 journals in the dataset; 116 had more than one article, and a much smaller group of 26 journals each had more than 5 articles. These are shown in Table 1.

Table 1. Journals with 5 or more articles in “Faceted” dataset.

Journals	Frequency
<i>Knowledge Organization</i>	33
<i>Cataloging & Classification Quarterly</i>	28
<i>Library Science with a Slant to Documentation</i>	19
<i>Information Processing & Management</i>	14
<i>Journal of Documentation</i>	12
<i>Information Today</i>	9
<i>Annals of Library & Information Studies</i>	8
<i>International Classification</i>	8
<i>SRELS Journal of Information Management</i>	8
<i>Library Hi Tech</i>	7

<i>Advanced Technology Libraries</i>	6
<i>Code4Lib Journal</i>	6
<i>D-Lib Magazine</i>	6
<i>Journal of the Association for Information Science & Technology</i>	6
<i>Library Journal</i>	6
<i>Online Information Review</i>	6
<i>Reference Librarian</i>	6
<i>SIGIR Forum</i>	6
<i>ALCTS Newsletter Online</i>	5
<i>Aslib Proceedings</i>	5
<i>DESIDOC Journal of Library & Information Technology</i>	5
<i>Journal of Internet Cataloging</i>	5
<i>Library Review</i>	5
<i>Library Trends</i>	5
<i>Nauchno Tekhnicheskaya Informatsiya Series 2</i>	5
<i>Scientometrics</i>	5

Although *Knowledge Organization* is the most populous journal, *Cataloging & Classification Quarterly* and *Library Science with a Slant to Documentation* are the next most populous—both of these are library practice oriented journals. *Information Processing & Management* and *Journal of Documentation*, which are more information science oriented, round out the top of the productivity cluster. Another 54 papers occurred in conference proceedings, most prominent were IFLA, ASIS and its SIG/CR Classification Workshops and various proceedings from the Indian Statistical Institute.

There were 528 first authors associated with the 642 papers. Sixty-six authors were associated with more than one paper, and a smaller group of 19 authors with three or more. These are shown in Table 2.

Table 2. Most productive first authors in “Faceted” dataset.

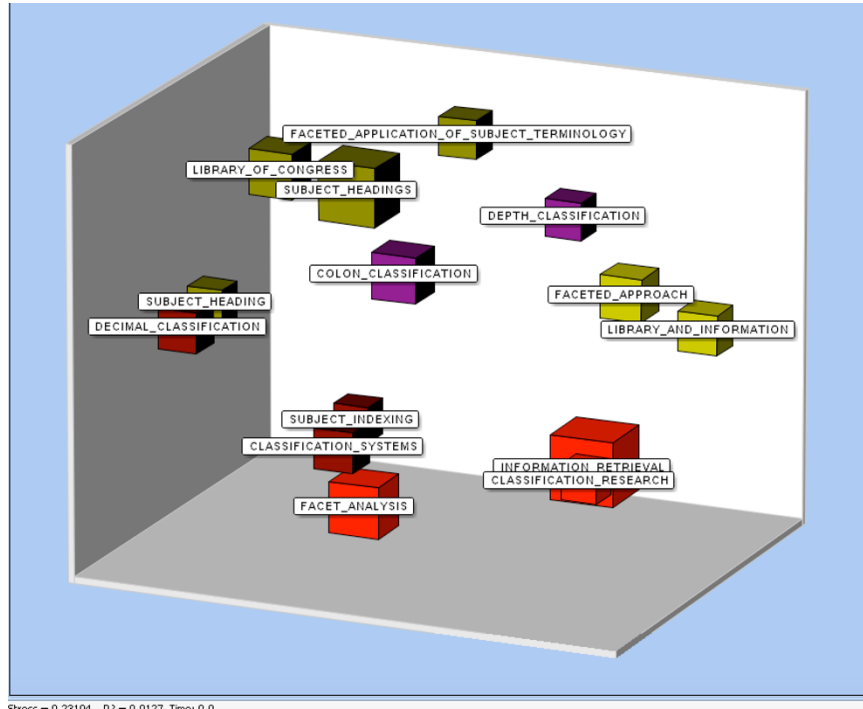
Authors	Frequency
Neelameghan, A	11
Broughton, Vanda	9
Ranganathan, S R	6

Anderson, James D.	5
Panigrahi, Pijushkanti	5
Spiteri, Louise F.	5
Williamson, Nancy J.	5
Aitchison, Jean	4
Breeding, Marshall	4
Gopinath, M A	4
O'Neill, Edward T.	4
Satija, M P	4

It is a bit surprising to find Ranganathan in this distribution, but the other names are familiar to the KO community. It is interesting to note the majority of these authors come from outside the United States, and almost all are generally associated more with library applications than with information science.

The titles and abstracts of the works in the dataset were entered into Provalis Research's ProSuite™ for co-word analysis. The WordStat™ module was used to generate frequency distributions of individual keywords and multi-word phrases. A multi-dimensionally scaled visualization of the most frequently occurring terms (both keywords and phrases) is shown in Figure 2.

Figure 2. WordStat™ MDS visualization of title terms (Stress = .23104 R²=.9127).



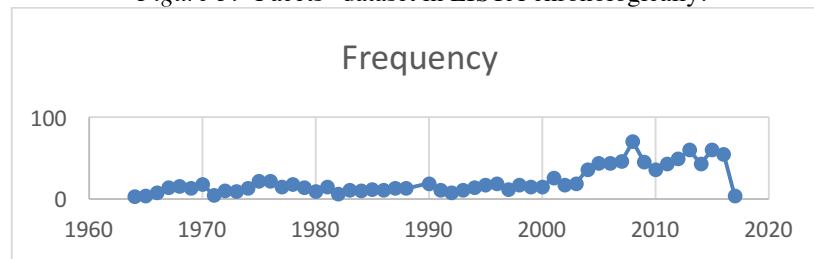
The colors identify four clustering concepts. At the core in bright red (near the bottom) are “facet analysis” and “classification research” associated with the most prominent term “information retrieval.” Considering the dimensionality of the visualization we can move upward and toward the view to encounter the smaller cluster incorporating “subject indexing” and “classification systems” including *Dewey Decimal Classification (DDC)*. Next we encounter *Colon Classification* associated with “depth,” and then finally most distant from the core in the most populous cluster we find “subject headings” and various “faceted approach[es]” to “subject terminology.” This latter cluster represents the extension in this instance. The picture generated here corroborates an emphasis on library applications, subject headings and general bibliographic classifications, but all of that surrounds a theoretical core of facet analysis for information retrieval.

Also, the KWIC feature was used to discover terms including the word “faceted” in the dataset. These were predominantly “faceted navigation,” “faceted classification,” “faceted indexing,” “faceted browsing,” “faceted vocabularies,” “faceted taxonomies,” and “faceted search tool.”

3.2 “Facets” etc.

There were 1202 citations retrieved from LISTA using the keyword “facets,” which seemed also to retrieve the keyword “facet.” The chronological spread of the dataset, is shown in Figure 3.

Figure 3. “Facets” dataset in LISTA chronologically.



The shape of the activity in the 1960s is similar to Figure 1 above, but the rest of the distribution until about 2005 is essentially flat, although continuous. After 2005 there is much increased sustained activity.

There were 606 journals in the dataset; 151 contained two or more articles and 48 contained 5 or more. A small set of 18 journals contained 10 or more articles each; these are shown in Table 3.

Table 3. Journals with 10 or more articles in “Facets” dataset.

Publication Title	Frequency
<i>Knowledge Organization</i>	30
<i>Journal of Documentation</i>	25
<i>Information Processing & Management</i>	21
<i>Library Science with a Slant to Documentation & Information Studies</i>	20
<i>Cataloging & Classification Quarterly</i>	19
<i>SRELS Journal of Information Management</i>	19
<i>Journal of the Association for Information Science & Technology</i>	18
<i>Computers in Libraries</i>	16
<i>Information Today</i>	15
<i>Journal of Academic Librarianship</i>	14

<i>Library & Information Update</i>	14
<i>Nauchno Tekhnicheskaya Informatsiya Series 2</i>	13
<i>International Classification</i>	11
<i>Bulletin of the Association for Information Science & Technology</i>	10
<i>CILIP Update</i>	10
<i>DESIDOC Journal of Library & Information Technology</i>	10
<i>Library Hi Tech</i>	10
<i>Publishers Weekly</i>	10

There is a larger cluster in the top tier than was the case in Table 1, and the order of precedence is different, with *Journal of Documentation* and *Information Processing & Management* showing more productivity, and with *Journal of the Association for Information Science and Technology* (variously named) in the upper tier. Thus there is more information science emphasis than was the case above. Conference proceedings comprise a much smaller segment than usual with only 68 papers from named conferences, among which the Association for Information Science and Technology (ASIST, variously named) and its Special Interest Group (SIG)/CR Classification Workshop, the Federation for Information and Documentation (FID), and the International Federation of Library Associations and Institutions (IFLA) predominate. There are 33 papers that are reports to various academic or government institutes.

There were 1056 first authors associated with the 1202 papers. Seventy-eight authors were associated with more than one paper, and 30 authors with three or more. These are shown in Table 4.

Table 4. Most productive first authors in “Facets” dataset.

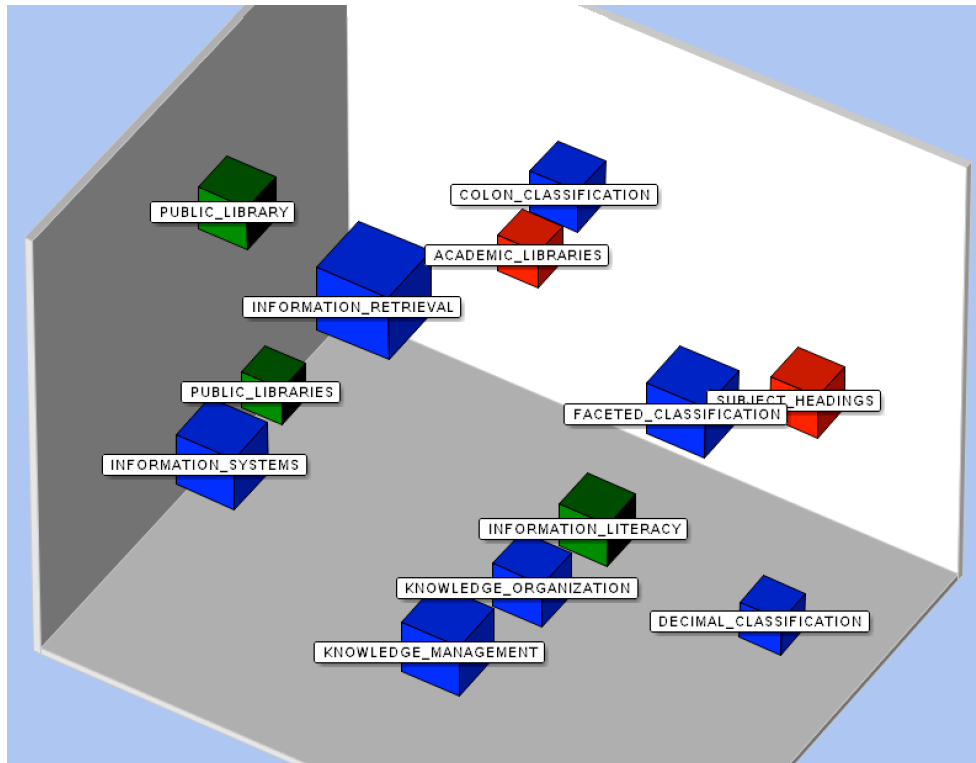
Authors	Frequency
Neelameghan, A	12
Gopinath, M A	10
Hjørland, Birger	8
Ranganathan, S R	8
Spiteri, Louise F.	7
Foskett, Douglas J	6
Broughton, Vanda	5
Atherton, Pauline	4
Gnoli, Claudio	4

Rieh, Soo Young	4
Anderson, James D	3
Arntz, Helmut	3
Buckland, Michael K.	3
Devadason, Francis	3
Dykman, Charlene A	3
Ellis, David	3
Enser, Peter	3
Gödert, Winfried	3
Mellwaine, Ia C.	3
Nobuyuki Midorika Wa	3
Notess, Greg R.	3
Panigrahi, Pijushkanti	3
Raghavan, K. S.	3
Rowley, Jennifer	3
Sano, H	3
Satija, Mohinder P	3
Scibor, Eugeniusz	3
Unesco	3
White, Martin	3
Williamson, Nancy J	3

This list is different from that in Table 2 above; although the same names appear here they are mixed with names more familiar to the KO community.

The titles and abstracts of the works in the dataset were entered into Provalis Research's ProSuite™ for co-word analysis. The WordStat™ module was used to generate frequency distributions of individual keywords and multi-word phrases. A multi-dimensionally scaled visualization of the most frequently occurring terms (both keywords and phrases) is shown in Figure 4.

Figure 4. WordStat™ MDS visualization of title terms (Stress = .24119 R^2 = .8569).



The three colors identify clustering concepts. The visualization is tilted forward for a bird's-eye view. In this case the smallest clusters involve “information literacy and public libraries” and “academic libraries and subject headings.” The largest cluster with the highest co-occurrence frequencies contains KO, information retrieval (IR) and “information systems,” *Colon Classification*, *DDC*, “faceted classification” and “knowledge management.” This larger cluster is the core representing the extension.

3.3 Combined dataset

The two datasets were merged and duplicates were removed, yielding 1682 citations in the combined dataset. This exercise also was useful for discovering problems in the dataset, such as misrepresented names (e.g., Kathryn Barre and Kathryn La Barre, or Hjørland and Hjorland) and also several unrelated citations for works from Facet Publishing. The lists of most productive authors and source journals was not affected by combining the datasets.

4.0 Discourse concerning facets in the KO domain

Informetric research has been used for decades in information science as a means of analyzing scientific communication. Since 2002 it has become a major means of conducting domain analysis for ontology extraction in KO incorporating techniques such as author cocitation analysis and co-word analysis (Smiraglia 2015). Recently, these techniques have been put to work to generate the outline of discourse that governs emerging thought in KO (see e.g, Smiraglia 2014; 2017). The usual discourse concerning facets and facet analysis runs from attribution to Ranganathan’s *Prologomena* (1937), through the promulgation of his *Colon Classification* (e.g., 1951), to the jumping off point of the Classification Research Group (1997 [1955]), and from that point a bifurcation into evolving faceted classifications (Broughton 2006) and faceted applications in information retrieval, and eventually in information architecture. But that is not the discourse that will emerge from this case study, because this research showcases only the indexed LIS literature concerning facets, and that only post 1967.

However, the discourse of facets on the research front of KO in relation to LIS begins to emerge from the data produced by the present case study. The discourse surrounding papers retrieved using the term “faceted” comes from outside the United States, or, to be more exact, comes mostly from India and the United Kingdom. It has library applications as its locus, although the term “information retrieval” appears with prominence, which suggests a sort of discourse arising between library classification and information retrieval. We must remember that in the beginning of the period tracked library classification was the more sophisticated tool and that information retrieval, such as it was, often involved intermediaries searching databases after interviews and returning retrieval sets of citations, for which the scholar then had to locate a print copy and make a photocopy. This is far removed from what we think of as information retrieval today, and even from how the paradigm was operating in the first years of the twentieth century, when we see an increase in research on faceted searching and faceted classification arise in this dataset. LIS was interested mostly in library efficiency, and secondarily in efficacy of full-text retrieval of documents. Indexing, or fact-chasing, was not the primary focus of the writings in this dataset. Thinking concerning information retrieval and the semantic web resides alongside implementation of faceted searching and the growth of faceted thesauri. Faceted classification remains important to the discourse, but the use of facet analysis is linked directly to applied aspects of information science.

The most productive authors are Neelameghan, Gopinath, Hjørland, Ranganathan, Spiteri and Broughton in the classification sphere. Titles of the works by this set of authors were analyzed comparatively and significant phrases are shown in Table 5.

Table 5. Core phrases from most productive first authors.

Hjørland	Gopinath	Broughton	Spiteri	Neelameghan
----------	----------	-----------	---------	-------------

knowledge organization	depth classification	faceted classification	essentials of faceted thesauri	depth classification version of CC
theories of knowledge	trends in research classification	knowledge organization	thesaurus construction	generation of rt links
				hierarchical associative relationship
	research in classification in india	vocabulary management	simplified model for facet analysis	
	recognition of manifestations of fundamental	organization in a digital environment	facet analysis in information retrieval	systems perspective of subject representation architecture
	depth classification version of cc	classification as a general theory	incorporating facets into social tagging	document finding
	study of causal factors	bibliographic classification as a model	analysis of current trends	model revisited
	manifestations of fundamental categories	methods of information retrieval		
	transport economics	theory for knowledge management		
		model for vocabulary management		
		dimensional knowledge structures		
		bliss bibliographic classification		

The table shows the preoccupation with traditional knowledge organization, classification, and an extension of facet analysis into thesauri. Phrases appearing above the blank line occur 4-9 times. Phrases occurring below the blank line, however, occur only twice each; these represent the evolving intension, including knowledge management, dimensional knowledge structures, social tagging, and facet analysis for information retrieval.

The list of journal sources runs from *Knowledge Organization, Information Processing & Management* and *Journal of Documentation* on the theoretical spectrum, to *Cataloging & Classification Quarterly, Library Hi-Tech* and *The Electronic Library* on the applied spectrum. Similarly, phrases from the works appearing in these journals were extracted and appear in Table 6.

Table 6. Core phrases from most productive journals.

KO IC	CCQ	JDOC	IP&M	LS SLANT
-------	-----	------	------	----------

knowledge organization	essential elements of faceted thesauri	faceted classification		depth classification
faceted classification				colon classification
faceted ontological classification systems				
faceted navigation	syntax for library of congress	digital archiving as information production using	faceted approach to conceptualizing tasks	freely faceted version of colon
domain ontology	selected guidelines for thesaurus construction	describing engineering documents with faceted	retrieval using semantically annotated product	faceted version of colon classification
internet resources	relevance of a classified catalog	archiving as information production using experts	facet product information search	study with a freely faceted
classification research	library of congress subject headings	design of subject access	semantically annotated product family ontology	system of document finding
	fbr perspective and a proposed	learners in the design	search and retrieval using semantically	generation of rt links
	facets into social tagging applications	multipurpose retrieval language tool	based opinion retrieval from blogs	
	access for music through usmarc	building a faceted classification	synthesis of class	
	analysis in information retrieval thesauri	principles and procedures	hierarchical associative relationship	
	query expansion	conceptualizing tasks in information seeking	transport economics	
	image retrieval		library classification	
			information production using experts and learners	
			faceted classification for the humanities	
			engineering documents with faceted approaches	

As before, the blank line identifies a point above which phrases occurred 4 or more times and below it only twice, thus below the blank line we see the evolving intension. Here we see especially in *IP&M* the shift from library classification, which occurs only twice, to product management, blog retrieval, information production, and conceptualizing tasks in information seeking. Here is the bifurcation apparent from the burst of activity post 2000 as the role of facets moves from structuring of general classifications into the realm of IR and the web.

This case study is part of a larger project, in which diverse snapshots of the literature representing the discourse of facets is examined; that larger study is reported in Smiraglia (2017b). In the larger study works cited are tracked in addition to the data reported here, and from that a broader narrative can emerge. However, two conclusions can emerge from this case study. The first is the utility of this research method, which is essentially a form of domain analysis, for revealing the shape of discourse. The second is the shape of the discourse of facets in LSI as represented in this case study derived from an information science dataset. Here we have one glimpse at the growth of facets in the intension of the domain, related to but not exclusively of, KO.

References

- Classification Research Group (CRG). 1997 [1955]. "The need for a Faceted Classification as the Basis of all Methods of Information Retrieval: Memorandum of the Classification Research Group: received, for Information, by the L. A. Library Research Committee, May 1955." In *From Classification to "Knowledge Organization: Dorking Revisited or 'Past is Prelude,'"* ed. A. Gilchrist. FID 714. The Hague : FID, 1-9.
- Dousa, Thomas M. 2014. "The Category of the Concrete in Early Subject Access Systems: A Historical and Theoretical Overview." In *Ontology for Knowledge Organization*, ed. Richard P. Smiraglia and Hur-li Lee. Würzburg: Ergon-Verlag, 39-86.
- Facets. Wikipedia *s.v.*. <https://en.wikipedia.org/wiki/Facet>
- Hedden, Heather. 2016. *The Accidental Taxonomist*, 2d ed. Medford, N.J.: Information Today.
- La Barre, Kathryn. 2006 "A Multi-Faceted View: Use of Facet Analysis in the Practice of Website Organization and Access." In *Knowledge Organization for a Global Learning Society: Proceedings of the Ninth International ISKO Conference, 4-7 July 2006 Vienna, Austria*. Advances in Knowledge Organization 10. Würzburg: Ergon-Verlag, p.359-66.
- Milonas, Elizabeth. 2010. "The Use of Facets in Web Search Engines." In *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the Eleventh International ISKO Conference, 23-26 February 2010, Rome, Italy*. Advances in Knowledge Organization 12. Würzburg: Ergon-Verlag, 349-55.
- Ranganathan, S. R. 1937. *Prolegomena to Library Classification*, 1st ed. Madras: The Madras Library Association.

- Rangathan, S.R. 1951. "Colon Classification and its Approach to Documentation." In *Bibliographic Organization: Papers Presented before the Fifteenth Annual Conference of the Graduate Library School July 24-29, 1950*, ed. Jesse H. Shera and Margaret E. Egan. Chicago: University of Chicago Press, 94-105.
- Rosenfeld, Louis, Peter Morville and Jorge Arango. 2015. *Information Architecture: For the Web and Beyond*, 4th ed. Sebastapol, CA: O'Reilly.
- Smiraglia, Richard P. 2014. "The Concept of Concepts: A Case Study from *American Documentation*." In *Connecting Across Borders: Globalization and Information Science: Proceedings of the Canadian Association for Information Science Annual Conference, St. Catherine's, Ontario, May 28-30, 2014*, ed. Matthew Griffis, Heidi Julien and Lisa Given.
- Smiraglia, Richard P. 2015. *Domain Analysis for Knowledge Organization: Tools for Ontology Extraction*. Chandos Information Professional Series. Amsterdam: Chandos Publishing, 2015.
- Smiraglia, Richard P. 2017a. "ISKO 14's Bookshelf: Discourse and Nomenclature—An Editorial." *Knowledge Organization* 44: 3-12.
- Smiraglia, Richard P. 2017b. "Facets as Domain Discourse. In *Dimensions of Knowledge: Facets for Knowledge Organization*, ed. Richard P. Smiraglia and Hur-Li Lee. Würzburg: Ergon Verlag.