Kathryn La Barre—University of Illinois, Urbana-Champaign

The Heritage of Facet Analysis in North America: Past Lessons as Pathways for Contemporary Exploration

Abstract
This paper will contrast the broad contours of Ranganathan's legacy in North America with a general assessment of contemporary North American facet applications. It will also offer a potential model for contemporary researchers that outlines heritage facet-analytical protocols currently in use.

1. Overview

> There are two types of 'width of knowledge.' One is knowing as much as possible of what is going on now. The other is knowing how we got to where we are – what is the heritage of ideas and practice on which we may draw (Vickery, 2004).

When the mathematician S.R. Ranganathan assumed the position of library director at the University of Madras in 1924, one of his earliest pursuits was to study the theoretical and practical aspects of librarianship with a special focus on classification. He began a course of studies with W. C. Berwick Sayers at University College of London and conducted tours of libraries throughout the United Kingdom. His planned visits to North American libraries and schools of library science were deferred until the 1950s and 1960s (Ranganathan, 1950). Ranganathan was able to maintain frequent correspondence from the mid-1920s with many in the international field that we know today as Knowledge Organization (KO), including foremost classification theorists in North America such as Melvil Dewey, Henry Evelyn Bliss, and Jesse Shera (La Barre, 2000; Ranganathan, 1957, p. 421; Ranganathan, 1992; Shera, 1949).

In 1967, Ranganathan's mentor Sayers observed, "As far as general libraries are concerned, classificatory research in the USA has taken a less spectacular form (than in Great Britain). Yet it is interesting to note that a [USA] Classification Research Group was set up … in 1959; possibly there will be a slow recognition of the value and techniques of facet and phase analysis." This paper will contrast the broad contours of Ranganathan's North American legacy, with a general assessment of contemporary facet applications. It will also offer a potential model for contemporary researchers that will outline heritage facet-analytical protocols currently in use by two North American researchers, Michèle Hudon and Kathryn La Barre.

2. Ranganathan's North American Legacy

Ranganathan was finally able to meet many of his correspondents during several trips to North American libraries and universities in 1950, 1958, and 1964. During his first extended visit in 1950 he served as a delegate to the Special Libraries Association conference in Atlantic City, New Jersey and was the chief speaker for the Golden Jubilee

celebration of the Classification and Cataloging Division at the American Library Association's annual meeting in Cleveland (Ranganathan, 1964). That same year he also served as a delegate at the Conference on Bibliographic Organization, hosted by Jesse Shera and Margaret Egan at the Graduate Library School in Chicago. (DRTC, online). Perhaps it was during one of these sessions that Ranganathan met James Perry, later affiliated with the Center for Documentation and Communications Research (CDCR), and decided to co-write an article on teamwork among disciplines, which briefly discussed Ranganathan's classification theories (Ranganathan and Perry, 1951).

Ranganathan's second extended trip to North America lasted from 1958 to 1959. During this time Jesse Shera, then dean of Western Reserve University, appointed Ranganathan as a member of the editorial board of *American Documentation.* He also gave a series of guest lectures at a number of LIS schools, including the Graduate Library School in Chicago. Ranganathan served as a panel member at the 1958 International Conference on Scientific Information (ICSI) in Washington D.C. jointly sponsored by the American Documentation Institute (now ASIST), the National Science Foundation, and the National Academy of Science. Also in attendance at this conference were two British members of the Classification Research Group, Douglas Foskett and Brian Vickery, who proposed that Ranganathan's facet analysis and faceted classification were well suited for information retrieval system design. Vickery observed here and in a subsequent paper, "The use of electronic devices for information retrieval is forcing us to think more deeply…the principles of a retrieval system must be clearly formulated…All these requirements are paralleled by the work already carried out by Ranganathan…[His] methodology is a model for all who work in this complex field" (Vickery, 1965, p. 108).

During this time Ranganathan also presented several papers at the 1959 Western Reserve University Conference on a Common Language for Machinery Search hosted by the Center for Documentation and Communications Research (CDCR) where Perry was now working with Allen Kent. So many researchers noted similarities between Kent and Perry's work with the Semantic Code beginning in the early 1950s, and Ranganathan's own Colon Classification dating to the 1930s, that the CDCR issued a report which attempted to minimize any intellectual debt to Ranganathan while admitting a high degree of complementarity between the two systems (Melton, 1960). Ranganathan's supporters in Britain, consisting mainly of CRG members, like Vickery and Foskett, continued to promote his work throughout the 1950s. North Americans decided to add their voices to the chorus. After discussions between Phyllis Richmond, of the University of Rochester, and Jesse Shera, Dean of Western Reserve University, the North American Classification Research Study Group (CRSG) formed in 1959 (La Barre, 2004a, 2004b; Richmond, 1959). The main activities of this group occurred between 1959 and 1969 and sought to capitalize on the momentum created by the ICSI conference (1958) and the Sixth Allerton Park Conference (1959) at the University of Illinois that surveyed the state of classification in American libraries (La Barre, 2004b; Richmond, 1961, p. 77). The CRSG was not affiliated with the professional societies such as the American Library Association, Special Libraries Association or the American Documentation Institute, but held informal meetings at each of

their conferences (Cochrane, 2000, 2001; La Barre, 2004a). The CRSG also created a traveling loan collection of classifications and controlled vocabularies that were housed at Western Reserve University, created an annotated bibliography for distribution, and hosted a definitions project to regularize the terminology of classification then in use in North America. Both Richmond and her co-chair Pauline Atherton Cochrane, who had met Ranganathan during his guest lecture at the University of Chicago in 1958, worked tirelessly to increase awareness of facet analysis and faceted classification through the activities of the CRSG. Each woman also found ways to extend Ranganathan's theory through research. Cochrane's own research trajectory began in 1961 at the American Institute of Physics (AIP), and was deeply grounded in Ranganathan's technique of facet analysis (Cochrane, 1965; La Barre, forthcoming). Richmond later worked intensively to bring the faceted indexing system PRECIS (PREserved Context Indexing System) into use at the Library of Congress (La Barre, 2004a; Richmond, 1976, 1981).

Ranganathan returned to North America in 1963 with a one-year appointment as visiting faculty member by Harold Lancour, Dean of the University of Pittsburgh. This facilitated his attendance at the International Seminar on the Colon Classification at Rutgers University in 1964, and his meetings with many North American researchers and classification theorists. By the mid-1960s applications based on facet analysis began to appear in systems that were familiar to some North American researchers. The American Petroleum Institute (API) used facet analysis to create a controlled vocabulary for indexing literature with nine facets: material, process, phenomenon, equipment, living organism, economic factor, place, property and operating condition (Mulverhill & Brenner, 1966). By 1967, a faceted classification was in use at the special library at the American Institute of Physics (Herschman, Inman & Lancman, 1967). *Library and Information Science Abstracts* adopted a faceted indexing scheme for library science that had been created in 1963, by members of the British Classification Research Group. Based on Ranganathan's facet theory, it used the following facets: operation, material, system, place, time and common subdivision (Library and Information Science Abstracts, 1969, v. 1).

As a result of these developments, North American awareness of Ranganathan's revolutionary facet theory increased dramatically through the 1960s. Facet theory continued to evolve in response to Ranganathan's discussions with researchers throughout the world. In the most traditional sense, facets represent basic concepts that are inherent in a given subject and are uncovered through a technique, known as facet analysis, which requires the conceptual analysis of a subject area into a set of fundamental categories. The entire process of facet analysis is governed by a canon composed of principles (specific rules), postulates (guidelines) and devices. Even though the traditional classificatory rules of logical division are central to facet theory, it diverges from tradition by strict adherence to a process of principled analysis of a given subject into homogeneous, mutually exclusive categories. The principles required:

(1) Each possible logical category to be clearly identified;
(2) Clear formulation of each new characteristic of division;
(3) Recognition of possible relations between categories.

This process of facet analysis is the basis of faceted classification, which cannot be said to exist without it (Vickery, 1960, pp. 12-13, 20).

The American Library Association recognized Ranganathan's contributions when he was awarded the 1970 Margaret Mann Citation in Cataloging and Classification "for his Colon Classification which has profoundly influenced modern classification theory and research through its faceted analysis techniques, for his works and principles and structure of the classified and dictionary catalog, and for a lifetime of signal devotion to the advancement of library science" (Atherton, 1970, p. 582). While facet theory retained a high profile for researchers in the United Kingdom from the 1970s through the mid-1980s only a handful of North Americans regularly cited Ranganathan. These included James D. Anderson, Pauline Atherton Cochrane, Timothy Craven, Eugene Garfield, Jean Perrault, and Phyllis Richmond. Their research represents a broad yet cohesive spectrum of interpretations of Ranganathan's facet theory. It would seem that awareness of facet theory bubbled quietly under the surface of scattered work in North America.

One result of the paucity of explicit discussions of facet theory in North America during the 1970s and 1980s, is that contemporary North American understanding and applications of facet analysis and faceted classification are varied and do not always reflect awareness of heritage principles. Highly visible instances of contemporary North American approaches include Rosenfeld and Morville's (2002) descriptions of faceted classification as a central construct in information architecture, and a brief nod to the utility of facet analysis in section 5.3.4 of the ANSI/NISO Z39.19-2005 *Guide to the Construction Management and Format of Controlled Vocabularies*. The faceted browsing and navigation features seen in the 2007 deployment of the Endeca software in the North Carolina State University online catalog have sparked renewed interest in facets among North American librarians. Marti Hearst's experimentations with hierarchical faceted metadata provide research-based descriptions of user interaction and interface design for faceted systems. The trajectory of these connections to the heritage of Ranganathan's facet theoretical approach is neither obvious nor easily traced. By contrasting common understandings of facets and facet analysis with contemporary approaches that are grounded in heritage principles it is hoped that more informed development will occur through continued dialogue that will enrich both.

Many researchers look to Hearst's Flamenco project for instruction in designing faceted interfaces that keep the information seeking needs of the user foremost. Underpinning Flamenco is the "faceted navigation paradigm" which seeks to provide flexible navigation through integrated search and browsing structures. Hearst recently designed a workshop for ACM SIG/IR (Association of Computing Machinery - Special Interest Group on Information Retrieval). Her definitions and approaches can serve as good indicators of current understandings of faceted approaches. Here, "facets refer to categories used to

20

characterize information items in a collection," a definition which fails to draw attention to any use of principled analysis (Hearst, 2006, 2008). Instead of the traditional facet-analytical approach, which is grounded in analysis of user interests, information seeking needs and facet analysis which draws out terms and concepts at use in a given domain; Hearst and others working in this area use facets that "are often derived by analysis of the text of an item using entity extraction techniques or from pre-existing fields in the database such as author, descriptor, language, and format. This approach permits existing web-pages, product descriptions or articles to have this extra metadata extracted and presented as a navigation facet" (Reamy, 2008). Entity extraction software can facilitate text mining processes, and in more sophisticated forms can differentiate between verbs and nouns, and identify the 'gist' of a given document (McCreary, 2009).

Traditional facet analysis depends on manual creation of high quality but increasingly expensive metadata, thus interest in automated solutions such as entity extraction for identifying facets is unbounded. Another common way to mitigate the expense of manually created metadata is the common practice of using pre-identified metadata such as that contained in database fields. Endeca's guided (or faceted) navigation in library OPACs uses pre-existing MARC metadata fields such as topic, year, or location in much the same way e-commerce implementations use the fields from the underlying inventory and sales databases used to power each website (Antelman, Lynema & Pace, p. 132).

The use of entity extraction to produce facets does not automatically violate the principles of facet analysis, unless those principles are not embedded in the algorithm and automated process itself. If a designer is working with the superficial understanding that facets are simply equivalent to categories or database fields per se, it is far less likely that the resulting facets will be as powerful or useful as those that might have been produced through application of facet theoretical approaches. In interviews with Information Architects, these kinds of approaches were characterized as "low hanging fruit" incapable of providing deep access to complex information resources and concepts (La Barre, 2006). Many practitioners noted that they look to the LIS research community to increase awareness of the power of the principles that underlie facet analysis, and to make the process itself more explicit so that it might be more readily adapted for use in automated approaches (La Barre, 2006).

3. A way forward: North American research with heritage protocols

A number of researchers in North America experimented with facets and facet analysis throughout the 1980s and 1990s. This paper seeks to highlight the current efforts of two researchers: Michèle Hudon at the Université de Montréal, and Kathryn La Barre, at the University of Illinois at Urbana-Champaign. Of special interest to those who wish to understand the application of facet analysis in creating access to information resources is Hudon's work with a collection of electronic education resources and her creation of a faceted structure to enhance discovery and retrieval. Her descriptions of the process provide contemporary designers with a useful framework that could inform more grounded use of facet theory. The facets in use by this project include: Agent (who?), Activity or process

(what?), Method or tool (how?), Space or context (where?), Time (when?), and foundations (general documents) (Hudon & Mas, 2006; Hudon, 2007).

The remainder of this paper will describe an approach that extracted facet-analytical protocols from Cochrane's work at AIP in the 1960s as part of the Documentation Retrieval Project for use as the framework for an ongoing research project, Folktales and Facets. The American Institute of Physics received funding from the National Science foundation to create a solution to information retrieval difficulties faced by physicists in 1961, through the Documentation Research Project (Cochrane, 1965). The project implemented facet analysis iteratively in two phases.

Phase I:

1. Determine the requirements of an ideal reference retrieval system by collecting search requests that would be made by a research physicist in the context of his current research work. Physicists were also asked to describe their current field in a description that was "sufficiently general to include all pertinent aspects of your work, and yet sufficiently restricted that activities in which you do not engage are not included in the description. Assume for the purposes here that you are describing your specific field to another physicist" (1965, p. 20-21).

2. Analyze (using facet analysis) the returned questionnaires.

3. Compare the findings of the analyses with the structure of existing indexing, classification, and subject heading systems.

4. Analyze extant indexing, classification and subject heading systems by inter-comparison.

Phase II:

1. Synthesize in successive approximations, a systematized description of physics research based on the physicist's description of their fields of physics research an don those procedures common to all fields of scientific research, e.g. determination of a property of some material or device by some experimental or theoretical method.

2. Construct an 'aid to indexing form' to assist authors and indexers in creating access points for articles.

First, search requests of physicists were acquired in the context of current research efforts. The second aim of the project was to create an orderly arrangement of the subject literature through the use of facet analysis in order to better incorporate multi-dimensional concepts, common research factors, and the current information needs of the research scientists. The facet-analytical protocol for this project took into account (1) common and

unique procedures, (2) literature surrogates created by research scientists, (3) the need for future revision and extension of the resulting access structures (Cochrane, 1965, p. 6). As a result of the facet analysis in Phase I, it was determined that five facets (property, object, method, emphasis, type of work: experimental, theoretical or both), were sufficient to form an implicit classification of physicists, and their research literature (Cochrane, 1965, p. 12). Phase II proceeded in an iterative fashion testing and implementing the facets as noted.

Folktales and Facets is the first phase of an extended project that seeks to enhance access to folktales through systematic and rigorous application of facet analysis and task focused models of information representation. Much in the same way that Hudon's research was able to uncover the information seeking needs of education researchers through analysis of information seeking literature of this population, the folktales project will explore extant literature on the subject, and but will also engage directly with users as part of the first phase of the project following the research protocol used by the AIP/DRP. Participants will discuss their activities and research, and elements of an ideal search and discovery system will be elicited. Researchers will also observe and analyze the interaction of the current users with the collection. In turn, the observed tasks, a sample of the collection, and extant vocabularies and indexing tools already in use for this type of material will also be subjected to facet analysis. Rather than creating a collection to study (Hudon, 2007), the folktales project will examine the folklore collection that is already part of the Center for Children's Books at the Graduate School of Library and Information Science at Urbana-Champaign. Scholars, storytellers, students, and instructors currently use this collection. In order to make the process of facet analysis more explicit, follow-on reports will discuss the procedures and findings for each step of the process. It is hoped that by making facet-analytical protocols explicit and understandable, those who seek to automate facet-analytical procedures may do so with full understanding of the nuances of the theory, greater ability to create algorithms that more full replicate traditional facet analysis, and a greater appreciation for areas of this work which may call for more detailed attention from humans.

4. Conclusion

This paper sought to establish the North American legacy of facets in order to establish the ways in which the heritage of facets and facet analysis may continue to inform contemporary research and development. Through a variety of archival sources, the following questions were explored: In what ways did North American researchers become acquainted with Ranganathan's facet theory? Which North American researchers and practitioners served as early adopters of facets as information access tools? In what ways could earlier approaches, guidelines or frameworks inform contemporary research and application development? Future work to extend knowledge of the strengths and weaknesses of facet-analytical approaches through empirical studies is urgently required. North American researchers continue to work alongside others as part of an international community that has long sought to bring Ranganathan's ideas into full implementation.

References

ANSI/NISO Z39.19-2005. (2005). *Guide to the construction management and format of controlled vocabularies*. Retrieved 24-01-2008 from:
http://www.niso.org/standards/resources/Z39-19-2005.pdf

Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a twenty-first century library catalog, *Information Technology and Libraries*, 25(3), 128–139.

Cochrane, P. (1965). *American Institute of Physics, Documentation Research Project: A review of work completed and in progress, 1961-1965*. New York, NY: American Institute of Physics.

Cochrane, P. A. (1970). Dr. S. R. Ranganathan. *Library Resources and Technical Services 16* (2), iii.

Cochrane interviews. (2001, 2002). Transcripts and tapes. Syracuse University Archives: Pauline Atherton Cochrane Papers: Box 23587.

DRTC. Documentation and Research Training Center. Life history. Positions held. *A tribute to Professor S. R. Ranganathan.* Retrieved 20-01-2009 from:
http://drtc.isibang.ac.in/DRTC/srr/index.html

Hearst, M. (2006). Design recommendations for hierarchical faceted search interfaces. ACM SIGIR Workshop on Faceted Search, August, 2006. Retrieved 21-03-09 from
http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf

Hearst, M. (2008). UIs for Faceted Navigation: Recent Advances and Remaining Open Problems, in the Workshop on Computer Interaction and Information Retrieval, HCIR 2008, October 2008, Redmond, WA. Retrieved 21-03-09 from:
http://flamenco.berkeley.edu/papers/hcir08.pdf

Herschman, A., Inman, C.L., Lancman, H. & Lerner, R. (1967). *A faceted classification scheme for Physics.* New York: American Institute of Physics. IARD 67-2.

Hudon, M. & Mas, S. (2006). Structure, logic, and semantics for Web-based collections in education. In Gerhard Budin, Christian Swertz, & Konstantin Mitgutsch. (Eds.) *Knowledge organization for a global learning society: Proceedings of the Ninth International Society for Knowledge Organization (ISKO), 4-6 July 2006, Vienna, Austria*, pp. 263-270. Würzburg, Germany: Ergon.

Hudon, M. (2007). Proposal for an alternate structure to organize and access resources in a virtual library in education. *Proceedings of the Canadian Association for Information Science (CAIS/ACSI) 35th Annual Conference*. Retrieved 20-03-09 from:
http://www.cais-acsi.ca/proceedings/2007/hudon_2007.pdf

La Barre, K. (2000). Bliss and Ranganathan: Synthesis, synchronicity or sour grapes? In Clare Beghtol, Lynne C. Howarth, & Nancy J. Williamson. (Eds.) *Dynamism and stability in knowledge organization. Proceedings of the Sixth International Conference of the International Society of Knowledge Organization (ISKO), 10-13 July 2000, Toronto, Canada*, pp. 157-163. Würzburg: Ergon Verlag.

La Barre, K. (2004a). The art and science of classification: Phyllis Allen Richmond 1921-1997. *Library Trends*, 52(4), 765-791.

La Barre, K. (2004b). Weaving Webs of Significance: The Classification Research Group in the United States and Canada. (2004). In M. Bowen & W. B. Rayward. (Eds.) *The History and Heritage of Scientific and Technological Information Systems: Proceedings of the 2002 Conference,* pp. 246-257. Medford, N. J.: Information Today for The American Society for Information Science and Technology and the Chemical Heritage Foundation.

La Barre, K. (2006). *The use of Faceted Analytico-Synthetic Theory as revealed in the practice of website construction and design*. Unpublished Dissertation. Indiana University.

La Barre, K. (forthcoming, Spring 2010). Weaving value from the past: Pauline Atherton Cochrane. Special issue: Pioneering women in information science. V. 2. *Libraries and the Cultural Record.*

Library and Information Science Abstracts. (1969). Volume 1 introduction, p. 1. New York: Alexander.

McCreary, D. (2009). How entity extraction is fueling the Semantic Web fire. O'Reilly Broadcast. Retrieved 20-03-09 from: http://broadcast.oreilly.com/2009/02/how-entity-extraction-is-fueli.html

Melton, J. (1960). *A note on the compatibility of two information systems: Colon Classification and Western Reserve University (Encoded Telegraphic Abstracts) and the feasibility of interchanging their notations*, Technical Note 13. Cleveland OH: Center for Documentation and Communications Research, Western Reserve University.

Mulverhill, J.G. & Brenner, E. H. (1966). Faceted organization of a thesaurus vocabulary. *Proceedings of the American Documentation Institute, 1966 Annual Meeting*, pp. 165-182. New York: Adrianne Press.

Ranganathan, S. R. (1950). *Library tour 1948: Europe and America, impressions and reflections.* Delhi : Indian Library Association.

Ranganathan, S. R. (1957). Unpublished letter of Melvil Dewey to S.R. Ranganathan, n.d. Prolegomena to Library Classification. 2nd ed.

Ranganathan, S. R. (1964). *Monologue on Melvil Dewey*. Retrieved 01-01-2008 Audio: http://www.miskatonic.org/library/ranganathan-audio.html
Transcript: http://www.hyperorg.com/misc/ranganathan_on_dewey_transcript.html

Ranganathan, S. R. (1992). *A librarian looks back: Autobiography of Dr. S. R. Ranganathan.* [with an evaluation of his life and work by P.N. Kaula]. New Delhi, India: ABC Publishing House.

Ranganathan, S.R. & Perry, J.W. (1951). External memory and research. *Journal of Documentation*, 7(3), 10-14.

Reamy, T. (2008). Automatic facets: Faceted navigation and entity extraction. Retrieved 21-03-09 from: http://www.kapsgroup.com/presentations/Automatic%20Facets.ppt

Richmond, P. (1959). Unpublished letter from J. Shera to P. Richmond (discussing the formation of the CRSG). Case Western University Archive: Papers of Phyllis A. Richmond: 27DD9 1:4.

Richmond, P. (1975). Richmond to University of Toronto, 26 December 1975 and 26 June 1976, unpublished document, n.d. "Exhibit in support of historical note: The development and growth of the BSC: Case Western Reserve University", pp. 13-14 [held at University of Toronto, Faculty of Information Studies, Inforum: The Integrated Library and Information Studies Laboratory].

Richmond, P. (1976). Classification from PRECIS: Some possibilities. *Journal of the American Society for Information Science*, 27(4): 240-247.

Richmond, P. (1981). *Introduction to PRECIS for North American usage.* Littleton CO: Libraries Unlimited.

Rosenfeld, L. & Morville, P. (2002). *Information architecture for the World Wide Web.* (2nd ed.). Cambridge, MA: O'Reilly.

Sayers, W.C.B. (1967). *A manual of classification for librarians.* (4th ed.). London: Deutsch. p. 375.

Shera, J. H. Unpublished letter from J. Shera to S.R. Ranganathan, (recommending North American libraries and schools Ranganathan might care to visit). November 25 1949. Case Western Reserve University Archives: Papers of Jesse Hauk Shera: 27DD5 10:3.

Vickery, B. C. (1965). Ranganathan's work on classification. In P. N. Kaula. (Ed.). *Library Science Today: Ranganathan Festschrift, Essays offered to S.R. Ranganathan on his seventy-first birthday*, p. 108. New York: Asia Publishing House.

Vickery, B. C. (2004). *A long search for* information (Occasional Papers, No. 213). Graduate School of Library and Information Science, University of Illinois at Urbana Champaign.