
Thesauri and Facets and Tags, Oh My! A Look at Three Decades in Subject Analysis

CANDY SCHWARTZ

ABSTRACT

The field of subject analysis enjoyed a flurry of interest in the 1970s, and has recently become a focus of attention again. The scholarly community doing work in this area has become more diffuse, and has grown to include new groups, such as information architects. Changes in information services and information seeking have led to reexamination of the nature and role of subject analysis tools and practices. This selective review looks at thesauri, guided navigation, and folksonomy as three activity areas in which subject analysis researchers have been attempting to address rapidly changing new environments.

INTRODUCTION

The 1970s were exciting years for those involved in subject analysis—a term used broadly here to encompass indexing, classification, thesaurus construction, and related “manual” or intellectual means of identifying topical content. Similarly, the late 1990s saw a resurgence in interest that continues to the present day. Though the study of subject analysis was hardly dormant in the intervening years, these two time periods serve as anchors for the following voyage through the scholarly literature and activities of a field in which F. W. Lancaster has played such an important role. This voyage is not a guided tour with stops at all possible points of interest, and is not intended as a state of the art review of all things to do with subject analysis (for these, see Lancaster, Elliker, & Connell, 1989; Markey & Miksa, 1987; McIlwaine & Williamson, 1999; Schwartz & Eisenmann, 1986; and Williamson, 1983, as well as Markey’s [2006] extremely thorough review of the role of library classification per se and in online

systems). This is instead a somewhat personal journey, selecting some areas of interest to illustrate how basic concepts and practices affect and are affected by new settings.

THE 1970S

In 1973, after a year of library and information science (LIS) core curriculum in cataloging and reference, Lancaster's *Vocabulary Control for Information Retrieval* (1972) offered to this budding librarian the perfect springboard to an expanded vision of subject analysis. Recommended by a mentor professor, *Vocabulary Control* was the first LIS textbook I purchased and read for pleasure rather than for a course. The sparse efficient (occasionally acerbic) writing and the assumption of reader intelligence contributed to its appeal, but the real excitement derived from the breadth of coverage and exposure to new ways of looking at subject languages and their role in online systems. In this influential text, Lancaster draws from a broad and deep understanding of machine-based as well as intellectual methods, and so it also served as an excellent invitation to explore automated information discovery more deeply. While information retrieval (IR) is outside the scope of this article, it is worth noting that the same era saw the publication of noteworthy IR texts by scholars such as Salton (1971), Spärck Jones (1971), Spärck Jones & Kay (1973), Van Rijsbergen (1975), and Vickery (1973).

In a further embarrassment of riches, many other significant publications and new editions of classics in subject analysis appeared in LIS book collections around that time, including:

- the first edition of *Thesaurus Construction and Use* (Aitchison & Gilchrist, 1972);
- the proceedings of the first Informatics conference (Aslib Coordinate Indexing Group, 1974);
- the first PRECIS manual (Austin, 1974);
- the fourth, and last, edition of *Indexes and Indexing* (Collison, 1972);
- the second edition of *The Subject Approach to Information* (Foskett, 1972);
- *The Thesaurus in Retrieval* (Gilchrist, 1971);
- *Classification in the 1970s* (Maltby, 1972);
- the magnum opus *Indexing Languages and Thesauri* (Soergel, 1974); and
- *An Introduction to Chain Indexing* (Wilson, 1971).

What accounts for this flurry of subject analysis activity in the 1970s? Obviously some of it can be attributed to the increasing availability and affordability of computers that could at last be used to create and manipulate bibliographic data. A decade earlier, most of the researchers who gathered at the seminal 1964 Symposium on Statistical Association Methods

for Mechanized Documentation (Stevens, 1966), had to deal in the hypothetical or use very small test sets; by the mid-1970s machine-readable collections large enough for reasonable exploration were on hand to support research. These technological advances also meant that large interactive online bibliographic databases were beginning to be available for application in cataloging and reference services. While the use of these was at the time restricted to staff (in the case of machine-readable catalogs) or search intermediaries (in the case of online reference searching), these systems made it possible to imagine and investigate tools and methods, which might be different from those used by card catalogs and printed indexing services, and in particular to reexamine subject analysis models, mechanisms, and techniques.

Another characteristic of this era was that the LIS world had not yet become overly specialized in its gathering places. For the researcher interested in subject analysis today, conference presentation and travel funding resources must be allocated among three digital libraries conferences, several special interest group conferences of the Association for Computing Machinery (ACM), the annual meeting of the American Society for Information Science and Technology (ASIST) and its annual information architecture summit, conferences of the International Society for Knowledge Organization (ISKO) and its newly-founded North American chapter (ISKO-NA), and any number of meetings where subject analysis is discussed in specific settings (e.g., archives or museums). With the subject analysis community scattered across many different meetings, there is no one venue where a substantial portion of that community meets face-to-face.

By contrast, in the mid-1970s, one could count on running into at least the North American contingent of the community at the annual meetings of the American Society for Information Science (ASIS) (the yearly conferences of the ACM Special Interest Group for Information Retrieval were relatively small until the 1980s). ASIS meetings were then, as they still are, big enough to be worth attending, but small enough to enable attendees to mingle easily and engage with speakers after sessions (or to thank an author for his or her inspiration). Passionate discussions of (and even songs about) indexing, searching, classification, and information retrieval used to take place in lobby bars and late-night hotel-room gatherings at ASIS conventions. It is likely that the excitement engendered by this kind of collective collegial interaction stimulated intellectual activity and scholarship, and had some influence on the spate of research and publication.

THE PRESENT

These days, many more people, affiliated with many different types of communities and agencies, are interested in subject analysis. Their unfiltered thoughts, informed and uninformed, are shared through blogs and

listserv lists, and the number of conferences of possible interest (including virtual meetings) continues to grow. The information intermediary has all but vanished from most libraries, although information research and synthesis is still a function in settings where it brings an obvious return on investment—typically corporate information centers. Large and well-established indexing services have converted back files to machine-readable form, most libraries in developed countries now have Web-accessible online public access catalogs (OPACs), and many libraries and other agencies host growing collections of digitized and born-digital resources. Skill in online searching remains part of the reference librarian job description, but most information seeking is carried out directly by end users, and not necessarily in systems created or controlled by libraries. Every individual has the potential to create content, to share that content on the Web, and to mash that content up with other content in unpredictable ways. And, more rapidly than we might think, a number of agencies are digitizing vast collections of books, to what end and with what result it is not yet clear.

These transformations have been profound but gradual, which is both a blessing and a curse. Rapid change would be difficult for libraries, publishers, and other information agencies, and would also be costly. Gradual change, on the other hand, is uneven, so that adjustments have to be made in increments, and traditional methods cannot be completely abandoned, even though they may not be optimal or necessary in the future (aspects of the concept of main entry in cataloging offer an example of this—designating one principal card where full information would be carried was a necessary space saving device in 3 x 5 card catalogs; it is not needed in machine-readable databases). Also, gradual change makes it all too easy to focus on the day-to-day and fail to notice that it might be the right time to reevaluate familiar and entrenched processes. Fortunately, while practice might be slow to change, subject analysis research has been stimulated by the problems and possibilities of networked information discovery and retrieval. Developments in thesauri, social tagging, and guided navigation illustrate some of this current interest.

Thesauri

Facet analysis is at the heart of most classification and thesaurus construction (Broughton, 2006) and is well explained in texts such as Aitchison, Gilchrist, & Bawden, (2000), although Mai (2006) adds that more thought should be given to the activity that precedes facet analysis (i.e., analyzing the domain) and recommends the application of cognitive work analysis to this task. Facet analysis leads to, among other things, the systematic discovery and assembly of the syndetic and semantic structure—the relationships intended to lead indexers and users around the vocabulary and promote match between query description and item description.

These relationships are usually divided broadly into equivalence (“use” and “used for”), associative (“related”), and hierarchical (“broader” and “narrower”), although Michel (1996) discovered well over one hundred different types and subtypes. Such relationships can be effective for information discovery in print tools. It is hard to avoid seeing “Academic Libraries, see also College Libraries” when the eye is scanning a printed index or card drawer. However, the relationship structure is less well deployed for users in online settings. Both Clarke (2001) and Milstead (2001) make the observation that increasingly large networked discovery systems are driving calls for change in relationship types and in standards for relationship development. In most indexing services these vocabulary references come into play only at the point where a user might query an online thesaurus, and Greenberg (2004) suggests that this occurs all too rarely. In OPACs, relationships are typically displayed during search, but are neither clearly nor helpfully presented in most systems.

There is some hope that the wealth of information captured in these structures might be exploited in other ways, and there has been some research along these lines (Shiri & Revie, 2005, offer a good review). Nielsen (2004) summarizes recent literature on the thesaurus and includes a section on the subtype of the searching thesaurus, where efforts are focused on the thesaurus as a tool for users rather than for indexers. Schatz, Johnson, and Cochrane (1996), for example, designed a system using a combination of thesaurus-driven and co-occurrence-based term suggestion. Soergel has explored reengineering classification schemes and thesauri into feature-rich ontologies, using a combination of automatic extraction and human editing (e.g., Soergel et al., 2004). Blocks, Cunliffe, and Tudhope (2006) present a very useful model of the interaction between user and thesaurus during information searching—they intend this as a framework for systems developers, and Shiri & Revie (2005) studied user perceptions of an existing (and typical) thesaurus interface in online searching. Harper and Tillett (2007) emphasize the potential contributions of controlled vocabularies (and authority control in general) to the Semantic Web. All of these activities augur well for controlled vocabularies, which are experiencing a revival of interest resulting from the need to find ways to improve user navigation in very large document spaces.

Thesaurus construction is usually associated with the development of postcoordinate indexing languages, which are generally regarded as the preferred choice for online bibliographic databases. Historically, however, library catalogs and many indexing services brought precoordinate vocabularies with them when they made the transition to online systems. The best known of these is the Library of Congress list of subject headings (LCSH), which is also the ancestor of most precoordinate languages found in older indexing services. In 1991, Lancaster, Connell, Bishop, and McCowan suggested that attempts to improve OPAC search based on the

existing searchable content of MARC records (including LCSH), or even on augmented records (e.g., with the addition of terms from indexes or tables of contents), were a waste of resources, since they bring about only marginal improvement, at far too great a cost. Others writing in the same era disagree, and a number of studies call instead for simplification of LCSH (these are reviewed in Franz, Powell, Jude, & Drabenstott, 1994). Marshall's later (2003) study of the value of adding generic subject headings to records underscores the flaws of subject access in OPACs. The debate continues—Thomas Mann (2006, 2007) is a staunch defender of the benefits of pre-coordinate subject headings in support of interdisciplinary scholarly research in library collections; others argue that the inflexibility of LCSH and the high costs of maintaining and using it outweigh the benefits (Calhoun, 2006). Markey (2007) makes the case that subject headings are a critical component of the next generation of OPACs.

The application of LCSH by catalogers has, in fact, not changed substantially and even some newly created digital libraries use LCSH (presumably to promote interoperability between library systems). However, the use of the MARC record for all things bibliographic is undergoing a transformation. Anticipating the use of LCSH in discovery systems based on the Dublin Core and XML rather than Anglo-American Cataloguing Rules (AACR) and MARC, and also in response to the calls for simplification, LC has been participating in OCLC's FAST (Faceted Application of Subject Terminology) project, which has parsed LC subject headings from the WorldCat database into facets (e.g., form, topical, geographic) to create a postcoordinate language, using the MARC21 authority format (Chan et al., 2001; Dean, 2004). The hope is to retain the richness of LCSH but to develop a more nimble and user-friendly tool, compatible with legacy subject headings and thus enabling conversion. The use of the term *facet* in the FAST project agrees more with its use in information architecture (IA) and guided navigation, where "topic" is one among many facets, as compared with thesaurus development, where "topic" is the primary object of facet analysis.

The belief that there are advantages to well-designed subject strings (coextensivity with item content, contextualization, and browsing) has led to several attempts outside of the FAST project to apply facet analysis to LCSH (e.g., Anderson & Hofmann, 2006). However, LCSH and its derivatives are not by any means the only alphabetical precoordinate indexing systems. PRECIS (PREserved Context Indexing System), developed by Derek Austin (Austin & Butcher, 1969; Austin, 1974) for the British Library, was born, flourished in a number of settings, and died, in the span of about twenty-five years. Based on a combination of facet analysis and linguistic theory (especially case grammar analysis), PRECIS was elegant, expressive, applicable across languages, and made the best use of the human-computer partnership in managing vocabulary control. How-

ever, it was also complex and hence costly to use, and its discontinuation by the British Library spelled its end. CIFT (Contextual Indexing and Faceted Taxonomy), a facet-based pre-coordinate system developed by James Anderson (Mutruix & Anderson, 1983) for the Modern Language Association's bibliographic databases, is simpler than PRECIS, and is still in use.

The study of the human indexer in the subject analysis process has not received as much attention recently as it did in the 1970s, when studies of indexer consistency and the effect of indexing practice on retrieval were not uncommon. In the third edition of *Indexing and Abstracting in Theory and Practice*, Lancaster (2003) devotes two chapters to indexing principles and practice, including a review of theories of "aboutness." Hjørland (2001) feels that aboutness has been poorly defined, and is closely related to relevance. Mai (2001) uses a semiotic framework to underscore the inextricable link between the social and cultural background of the indexer and the interpretation of document content. In part one of an excellent state-of-the-art review of the nature of indexing, Anderson and Pérez-Carballo (2001) survey the research on how humans process documents, interpret content, and assign index terms. It will be interesting to see whether, and how, studies of tagging by users will contribute to our understanding of content labeling.

Social Tagging

A recent development in subject analysis has seen the user move from consumer of information descriptions to participating content provider and indexer. Thomas Vander Wal coined the word *folksonomy* in 2004 to describe the "the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval" (2007), as compared with uncontrolled keyword indexing in early databases, where the intention was to promote retrieval by others. Other terms for this activity include *folk classification*, *ethnoclassification*, *distributed classification*, *social classification*, *open tagging*, *free tagging*, and *social bookmarking*. Tagging was introduced widely to the general Web-using public by bookmarking systems (most notably del.icio.us), developed to provide remote access to personal collections of bookmarked Web links (Hammond, Hannay, Lund, & Scott, 2005). Whether tagging one's own content (e.g., flickr, for uploaded images, and LibraryThing, for bibliographic records of personal library collections) or others' websites and publications (e.g., del.icio.us, for bookmarks, and Connotea, an online citation manager), the user's primary purpose is personal organization and retrieval, although Morrison (2007) suggests a number of other additional drivers. Whatever the motivation, the sharing of both tags and links can reveal other items of interest and other users of like mind, hence the "social" aspect.

The rapid growth of systems using tagging has inspired a sometimes spirited exchange as to its utility for information discovery (good over-

views of the issues can be found in Guy & Tonkin, 2006 and Kroski, 2005). From the user point of view, tagging has the advantages of a low entry barrier and low costs, and supports serendipitous discovery and social connection as well as serving its primary purpose of personal retrieval. Information retrieval specialists point to the disadvantages of user-generated terminology being overly specific or personal (e.g., “wishlist” in LibraryThing), lacking in vocabulary control (of synonymy, polysemy, or variation in person and spelling), and lacking hierarchy for browsing or generic search. Interestingly, some of the more mature tagging systems have begun to take on some of the functions of controlled vocabularies. LibraryThing and del.icio.us both use tag bundles or aliases (e.g., the tag “tbr” stands for a number of variations on “to be read”), and several systems give users the option of adding scope notes to tags. Conference chairs and event managers regularly request that attendees use a prespecified tag when blogging an event or posting pictures in flickr.

The prevailing view that has emerged is that both tagging and controlled vocabularies have a role to play in discovery and retrieval, and they are not mutually exclusive (Gordon-Murnane, 2006; Noruzi, 2007). Spiteri (2007) studied tags in three large sites for their conformance to national standards for thesaurus construction, and found some close correspondence, leading her to suggest guidelines for incorporating tagging into OPACs. Some libraries see tagging as a way of augmenting the retrieval tools in OPACs (examples include LibraryThing for Libraries and the University of Pennsylvania’s PennTags), and the ubiquitous tag cloud is being used to display everything from queries in the Ann Arbor District Library’s catalog search cloud to classification number assignments in OCLC’s Dewey browser.

Guided Navigation

When users search library-supplied systems (instead of, say, Google), they frequently do so remotely, or at least not in any way that is immediately visible to library staff. The opportunity to help users by teaching them how to do research (“teach a man to fish”), and how to leverage indexing language structures, has diminished. Most online bibliographic systems now rely on ranking algorithms to help users find the most relevant items, so any deliberate learning is bypassed, and only very frequent users may intuitively or accidentally discover and remember effective search strategies and tools.

In the absence of user learning, and with no easy way for users to exploit thesaurus relationships, attention has recently turned to what has come to be called “guided navigation”—one result of the intersection between information architecture (IA) and LIS. As designers of Web user experiences, information architects need to find ways to help users (especially online shoppers and corporate employees) navigate

through large information spaces containing objects with many potentially searchable attributes. Toys, for example, might be sought by price, age, brand, genre, material, country of manufacture, and so on, but none of these can be predicted to be the preferred starting point for a given user. Under guided navigation, a user can begin browsing by choosing any facet as the first filter, or can start with a search across all items. In successive steps, the search result can be refined by any remaining facet, and at each step item counts for topics in remaining facets are recalculated (hence the “guided”). One critical component task for the information architect developing this kind of browsing is to derive the facets and their contents. For this, the IA community draws largely on LIS, looking specifically to facet analysis and taxonomy development (thesaurus construction). Happily, this has had the effect of renewing interest in Ranganathan and bringing his work to a wider audience (e.g., Spiteri, 1998; Weinberger, 2003), even though these taxonomies are often not quite what Ranganathan might have envisioned, being neither classification nor thesaurus (Gilchrist, 2006).

Although initially positioned for (and successfully marketed to) large commercial settings, guided navigation has also been applied in bibliographic systems. North Carolina State University was the first institution to offer guided navigation across a large library OPAC, using technology developed by Endeca (Antelman, Lynema, & Pace, 2006). Not much more than two years later, guided navigation is beginning to appear in most new versions of OPAC software (e.g., WorldCat Local), as well as in indexing services (e.g., Scopus) and subject gateways (e.g., Librarian’s Index to the Internet). It seems to be intuitive that guided navigation improves retrieval, and also that it would expose the “long tail” in library collections (i.e., undercirculated items), though few studies have looked at retrieval effectiveness, and tend to focus instead on usability.

CONCLUDING REMARKS

New environments encourage new approaches, but new approaches are informed by what has gone before, and Lancaster has played an important role in that foundation. Joudrey (2002) found that *Indexing and Abstracting in Theory and Practice* was the most frequently assigned textbook for courses in indexing, and ranked sixth across all courses to do with bibliographic control, while *Vocabulary Control* has been out of print for over a decade, but is still cited in course readings, and is second in rank for courses in thesaurus construction (following Aitchison, Gilchrist, & Bawden, 2000). This means that Lancaster has influenced thousands of LIS students, and these same students have gone on to use, improve, and develop subject access systems. Until recently, improvements have been subtle and unremarkable, as the settings in which subject analysis played a role did not change dramatically. Now, however, readily and freely acces-

sible collections have reached a scale that demands new thinking about access. Study after study tells us that users turn first to Google as a source of information, and that even when they can be persuaded to use library-supplied indexing services, their search behaviors do not directly make the most of the controlled structures we labor to provide. And now users are doing their own indexing, primarily for rediscovery of known objects in personal information spaces, but with the side effect of finding how others have used the same terms, or how others have indexed the same items.

The products of indexing and organization, and some of their component activities, are being experienced by a much larger audience. The popularity of tools such as guided navigation, or the tag cloud, both of which parlay underlying subject structures into visually accessible navigation aids, speaks to the need for access systems that do not require users to do anything other than type a few words or choose a filter as a beginning step. No, we are not there yet, and I have to agree with Markey's (2006) cautionary note that we need to think deeply about the impact of mass digitization on information discovery and act swiftly to develop complementary processes. But I maintain that we do so more intelligently if we have a shared foundation and a common vocabulary, and much of that, along with a great deal of informed skepticism, has come from the work of F. W. Lancaster.

REFERENCES

- Aitchison, J., & Gilchrist, A. (1972). *Thesaurus construction: A practical manual*. London: Aslib.
- Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: A practical manual* (4th ed.). London: Aslib IMI.
- Anderson, J. D., & Hofmann, M. A. (2006). A fully faceted syntax for Library of Congress subject headings. *Cataloging & Classification Quarterly*, 43(1), 7–38.
- Anderson, J. D., & Pérez-Carballo, J. (2001). The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing & Management*, 37(2), 231–254.
- Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a twenty-first century library catalog. *Information Technology and Libraries*, 25, 128–139.
- Aslib Coordinate Indexing Group. (1974). *Informatics 1: Proceedings of a conference held by the Aslib Co-ordinate Indexing Group on 11–13 April 1973, at Durham University*. London: Aslib.
- Austin, D. (1974). *PRECIS: A manual of concept analysis and subject indexing*. London: Council of the British National Bibliography.
- Austin, D., & Butcher, P. (1969). *PRECIS: A rotated subject index system*. London: British National Bibliography.
- Blocks, D., Cunliffe, D., & Tudhope, D. (2006). A reference model for user-system interaction in thesaurus-based searching. *Journal of the American Society for Information Science & Technology*, 57(12), 1655–1665.
- Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, 58(1), 49–72.
- Calhoun, K. (2006). *The changing nature of the catalog and its integration with other discovery tools*. Washington, DC: Library of Congress. Retrieved October 1, 2007, from <http://www.loc.gov/catdir/calhoun-report-final.pdf>.
- Chan, L. M., Childress, E., Dean, R., O'Neill, E. T., & Vizine-Goetz, D. (2001). A faceted approach to subject data in the Dublin Core metadata record. *Journal of Internet Cataloging*, 4(1/2), 35–47.

- Clarke, S. G. D. (2001). Thesaural relationships. In C. A. Bean (Ed.), *Relationships in the organization of knowledge* (pp. 37–52). Boston, MA: Kluwer.
- Collison, R. L. (1972). *Indexes and indexing: Guide to the indexing of books and collections of books, periodicals, music, recordings, films and other material, with a reference section and suggestions for further reading* (4th rev. ed.). London: Ernest Benn.
- Dean, R. J. (2004). FAST: Development of simplified headings for metadata. *Cataloging & Classification Quarterly*, 39(1), 331–351.
- Foskett, A. C. (1972). *The subject approach to information* (2nd ed.). London: Clive Bingley.
- Franz, L., Powell, J., Jude, S., & Drabenstott, K. (1994). End-user understanding of subdivided subject headings. *Library Resources & Technical Services*, 38(3), 213–226.
- Gilchrist, A. (1971). *The thesaurus in retrieval*. London: Aslib.
- Gilchrist, A. (2006). Structure and function in retrieval. *Journal of Documentation*, 62(1), 21–29.
- Gordon-Murnane, L. (2006). Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher*, 14(6), 26–38.
- Greenberg, J. (2004). User comprehension and searching with information retrieval thesauri. *Cataloging & Classification Quarterly*, 37(3), 103–120.
- Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up tags. *D-Lib Magazine*, 12(1), 19–33. Retrieved October 1, 2007, from <http://www.dlib.org/dlib/january06/guy/01guy.html>.
- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social bookmarking tools (I). *D-Lib Magazine*, 11(4). Retrieved October 1, 2007, from <http://www.dlib.org/dlib/april05/hammond/04hammond.html>.
- Harper, C. A., & Tillett, B. B. (2007). Library of Congress controlled vocabularies and their application to the Semantic Web. *Cataloging & Classification Quarterly*, 43(3), 47–68.
- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content . . . and relevance. *Journal of the American Society for Information Science & Technology*, 52(9), 774–778.
- Joudrey, D. N. (2002). Textbooks used in bibliographic control education courses. *Cataloging & Classification Quarterly*, 34(1/2), 103–120.
- Kroski, E. (2005). *The hive mind: Folksonomies and user-based tagging*. Unpublished manuscript. Retrieved October 2, 2007, from <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging>.
- Lancaster, F. W. (1972). *Vocabulary control for information retrieval*. Arlington, VA: Information Resources Press.
- Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice* (3rd ed.). Champaign, IL: University of Illinois.
- Lancaster, F. W., Connell, T. H., Bishop, N., & McCowan, S. (1991). Identifying barriers to effective subject access in library catalogs. *Library Resources & Technical Services*, 35(4), 377–391.
- Lancaster, F. W., Elliker, C., & Connell, T. H. (1989). Subject analysis. *Annual Review of Information Science and Technology*, 24, 35–84.
- Mai, J. (2001). Semiotics and indexing: An analysis of the subject indexing process. *Journal of Documentation*, 57(5), 591–622.
- Mai, J. (2006). Contextual analysis for the design of controlled vocabularies. *Bulletin of the American Society for Information Science & Technology*, 33(1). Retrieved October 1, 2007, from <http://www.asis.org/Bulletin/Oct-06/mai.html>.
- Maltby, A. (1972). *Classification in the 1970's: A discussion of development and prospects for the major schemes*. Hamden, CT: Linnet Books.
- Mann, T. (2006). *The changing nature of the catalog and its integration with other discovery tools, final report, March 17, 2006: A critical review*. Washington, DC: Library of Congress Professional Guild. Retrieved October 1, 2007, from <http://guild2910.org/AFSCMECalhounReviewREV.pdf>.
- Mann, T. (2007). *The Peloponnesian War and the future of reference, cataloging, and scholarship in research libraries*. Washington, DC: Library of Congress Professional Guild. Retrieved October 1, 2007, from <http://guild2910.org/Peloponnesian%20War%20June%2013%202007.pdf>.
- Markey, K. (2006). Forty years of classification online: Final chapter or future unlimited? *Cataloging & Classification Quarterly*, 42(3), 1–63.

- Markey, K. (2007). The online library catalog: Paradise lost or paradise regained? *D-Lib Magazine*, 13(1/2). Retrieved October 1, 2007, from <http://www.dlib.org/dlib/january07/markey/01markey.html>.
- Markey, K., & Miksa, F. (1987). Subject access literature, 1986. *Library Resources & Technical Services*, 31(4), 335-354.
- Marshall, L. (2003). Specific and generic subject headings: Increasing subject access to library materials. *Cataloging & Classification Quarterly*, 36(2), 59-87.
- McIlwaine, I. C., & Williamson, N. J. (1999). International trends in subject analysis research. *Knowledge Organization*, 26(1), 23-29.
- Michel, D. (1996, June). *Taxonomy of subject relationships*. Retrieved October 6, 2007, from <http://www.ala.org/ala/altscontent/catalogingsection/catcommittees/subjectanalysis/subjectrelations/appendixbpartii.htm>.
- Milstead, J. L. (2001). Standards for relationships between subject indexing terms. In C. A. Bean (Ed.), *Relationships in the organization of knowledge* (pp. 53-66). Boston: Kluwer.
- Morrison, P. J. (2007). Why are they tagging, and why do we want them to? *Bulletin of the American Society for Information Science and Technology*, 33(7). Retrieved October 9, 2007, from http://www.asis.org/Bulletin/Oct-07/Morrison_OctNov07.pdf.
- Mutrux, R., & Anderson, J. D. (1983). Contextual indexing and faceted taxonomic access system. *Drexel Library Quarterly*, 19(3), 91-109.
- Nielsen, M. L. (2004). Thesaurus construction: Key issues and selected readings. *Cataloging & Classification Quarterly*, 37(3), 57-74.
- Noruzi, A. (2007). Folksonomies: Why do we need controlled vocabulary? *Webology*, 4(2). Retrieved October 1, 2007, from <http://www.webology.ir/2007/v4n2/editorial12.html>.
- Salton, G. (1971). *The SMART retrieval system: Experiments in automatic document processing*. New York: Prentice-Hall.
- Schatz, B., Johnson, E. H., & Cochrane, P. A. (1996). Interactive term suggestion for users of digital libraries: Using subject thesauri and co-occurrence lists for information retrieval. In E. A. Fox & G. Marchionini (Eds.), *Proceedings of the 1st ACM International Conference on Digital Libraries* (pp. 126-133). New York: Association for Computing Machinery.
- Schwartz, C., & Eisenmann, L. (1986). Subject analysis. *Annual Review of Information Science and Technology*, 21, 37-62.
- Shiri, A. A., & Revie, C. (2005). Usability and user perceptions of a thesaurus-enhanced search interface. *Journal of Documentation*, 61(5), 640-656.
- Soergel, D. (1974). *Indexing languages and thesauri: Construction and maintenance*. Los Angeles: Melville.
- Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., & Katz, S. (2004). Reengineering thesauri for new applications: The AGROVOC example. *Journal of Digital Information*, 4(4). Retrieved October 8, 2007, from <http://jodi.tamu.edu/Articles/v04/i04/Soergel>.
- Spärck Jones, K. (1971). *Automatic keyword classification for information retrieval*. Hamden, CT: Archon Books.
- Spärck Jones, K., & Kay, M. (1973). *Linguistics and information science*. New York: Academic Press.
- Spiteri, L. (1998). A simplified model for facet analysis: Ranganathan 101. *Canadian Journal of Information & Library Sciences*, 23(1/2), 1-30.
- Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, 4(2). Retrieved October 1, 2007, from <http://www.webology.ir/2007/v4n2/a41.html>.
- Stevens, M. E. (Ed.). (1966). *Statistical association methods for mechanized documentation: Symposium proceedings*. Washington, DC: U.S. Government Printing Office.
- Van Rijsbergen, C. J. (1975). *Information retrieval*. London: Butterworths.
- Vander Wal, Thomas. (2007). Folksonomy coinage and definition. Retrieved December 9, 2007, from <http://vanderwal.net/folksonomy.html>.
- Vickery, B. C. (1973). *Information systems*. London: Butterworths.
- Weinberger, D. (2003). Rediscovering Ranganathan. *Forrester Magazine*, 2003(3) 68-70.
- Williamson, N. J. (1983). Subject access in the on-line environment. *Advances in librarianship*, 13, 49-97.
- Wilson, T. D. (1971). *An introduction to chain indexing*. Hamden, CT: Linnet Books.

Candy Schwartz is a professor at the Graduate School of Library and Information Science at Simmons College, where she teaches information organization, digital libraries, and subject analysis, and coordinates the doctoral programs. With Professor Peter Herson, she is currently directing a PhD concentration in managerial leadership for the information professions, funded by IMLS (the Institute of Museum and Library Services). She is an active member of the American Society for Information Science and Technology (ASIST), served as the association's president from 1998–1999, and is the recipient of a number of ASIST awards, including Outstanding Information Science Teacher. Candy has published several books, including *Sorting Out the Web: Approaches to Subject Access*, and *Revisiting Outcomes Assessment in Higher Education* (with several colleagues), and she is coeditor of the peer-reviewed journal *Library & Information Science Research*.