Subject Indexing and Classification, 2002–2007

A close review of the articles published in Cataloging & Classification Quarterly and Library Resources & Technical Services supplemented by topical searches of other sources for the period 2002–2007 demonstrates a steady stream of research literature related to subject representation. This broad topic embraces both subject indexing (the use of controlled vocabularies for subject access) and classification (in the narrow sense of the application of hierarchical schemes—using classmarks and/or call numbers—for organizing knowledge according to subject content). Complicating matters, some authors used the term “classification” more broadly to refer to subject representation as a whole, and their essays are here addressed in the subject indexing section.

Subject Indexing

The literature on subject indexing can be grouped into four main clusters: theoretical articles; reports and studies supporting use of controlled vocabularies; critiques and revisions of the Library of Congress Subject Headings (LCSH); and discussions of approaches and challenges to achieving multilingual subject access.

Theoretical Articles

The principle of specificity in subject indexing recurred as a common concern in several of the theoretical articles, either directly or as extrapolated into discussions about controlled vocabularies versus machine indexing.

Mai (2003) contrasted general and special classification schemes (using the term “classification” to refer to all subject representation) and maintained that both types will remain relevant, despite the ever-increasing difficulty of establishing interoperability between them. Marshall (2003) addressed the principle of specificity for subject headings more directly, asserting that “assigning both specific and generic subject headings to a work would enhance the subject accessibility,” while acknowledging the diminution of subject heading consistency and collocation that would result (59). Naun (2006) identified objectivity, or impartiality, as a central principle of subject representation inherited from print culture, embodied in the “pragmatic rule of specific entry” (93).

In the most interesting of the theoretical articles, Hanson (2004) interpreted the recent de-emphasis on traditional subject analysis and classification stemming from the rise of automated full-text indexing as illustrating the shift from a modernist structured worldview to one of postmodernist indeterminacy. Far from decrying this shift, Hanson cited legal research and interdisciplinary studies as among the areas that have most prospered from the increased flexibility and creativity made possible by full-text searching. Denda (2005) similarly judged traditional controlled vocabularies to be insufficiently flexible to provide access to materials in women’s studies and other interdisciplinary fields, and advocated the creation (apparently by subject specialists or reference librarians rather than catalogers) of subject ontologies for these fields.

Reports and Studies Supporting Use of Controlled Vocabularies

Taylor and Joudrey (2002) stressed the continuing importance for all librarians of understanding controlled vocabulary use in subject access, and described their pedagogical approach to teaching subject cataloging. Miller, Olson and Layne (2005) summed up the best practices in subject reference structures gleaned from ten years of work by the ALCTS CCS Subject Analysis Committee, adding an appendix detailing the committee’s specific recommendations for providing access to and display of
reference structures in online catalogs. The contribution of new headings to LCSH was aided by the appearance of the second edition of the SACO participants’ manual (SACO 2007).

Several studies explored the results of actual subject searches in particular online catalogs. Graham (2004) examined catalog transaction logs to identify subject searches that yielded no hits and then created new authority record cross-references to improve future results for these searches, suggesting that a cross-reference structure based on “use warrant” might prove more useful than the current policy of requiring citations for all cross-references, i.e., “literary warrant.”

Transaction logs were also used by Gross and Taylor (2005) to identify actual search terms that were then re-searched to help determine that 35.9 percent of the records retrieved by keyword searches would not have been retrieved had the records lacked subject headings—substantial evidence of the importance of providing controlled subject access. Lastly, Garrett (2007) explicitly extended the work of Gross and Taylor to the issue of access for historical materials; his study of a project to add subject headings to records for materials within the Early English Book Online and Eighteenth Century Collections Online electronic resources found that some 60 percent of sample records containing the keyword search phrase “east india company” had the phrase only in subject headings.

Critiques and Revisions of LCSH

Articles on LCSH were of two main types: assessments of political and cultural bias; and discussions of proposed modifications of LCSH syntax involving the use of facets. That political and structural objections to LCSH have a long history was demonstrated in Fischer (2005), a bibliography of critical views of LCSH from 1990 to 2001. Knowlton (2005) revisited the biased headings identified by Sanford Berman in 1971 and concluded that roughly two-thirds of them have been changed, with the main persisting bias being that “many subjects headings pertaining to the Christian religion remain unglossed” (128). Strottman (2007) identified numerous flaws and gaps in LCSH headings relating to Southwestern cultures and history and ascribed them to an East Coast “hegemony” (59), omitting to mention LCSH’s dependence on literary warrant.

The most notable experimental revision of LCSH syntax was OCLC’s FAST (Faceted Application of Subject Terminology) schema, described in both O’Neill and Chan (2003) and Dean (2004), and still undergoing development and testing by OCLC and the ALCTS Subject Analysis Committee (Minutes, 2007). FAST headings are based on LCSH vocabulary but shift the emphasis from precoordination to postcoordination by constructing headings within eight distinct facets: Topical; Geographic (Place); Personal Name; Corporate Name; Form (Type, Genre); Chronological (Time, Period); Title; and Meeting Name. FAST headings can include subheadings of the same facet type, but unlike LCSH, do not combine headings of one facet type with subheadings of a different type. Mitchell and Hsieh-Yee (2007) studied the feasibility of converting headings from Ulrich’s Periodicals Directory to FAST headings and judged the conversion process quite manageable.

Another ambitious LCSH revision proposal was made by Anderson and Hofmann (2006), who proposed a “fully faceted syntax” for LCSH based on seventeen facets from Bliss’s Bibliographic Classification (BC2) scheme. The principal benefit claimed for this approach was the ability to create single heading statements coextensive with the topic of a work. In appealing for use studies of their scheme to be conducted, Anderson and Hofmann acknowledged that its complexity would be daunting for even experienced LCSH catalogers to learn (in contrast to the simplicity striven for by FAST).

Multilingual Subject Access

The challenge of providing multilingual subject access in online catalogs has been worked on extensively in Europe, most notably by the MACS (Multilingual access to subjects) project, begun in 1997 by the Swiss National Library, the Bibliothèque nationale de France, the British Library, and the Deutsche Bibliothek. Clavel-Merrin (2004) and Landry (2004) both provided overviews of the history and progress of
the MACS effort, and explained that MACS manually established links between individual headings from three subject thesauri in different languages: the German SWD/RSWK (Schlagwortnormdatei/Regeln für den Schlagwortkatalog); the French RAMEAU (Répertoire d’autorité-matière encyclopédique et alphabétique unifié); and the English LCSH.

Frommeyer (2004) considered the treatment of time in period subdivisions and chronological terms across the same three subject heading languages used in MACS (SWD/RSWK, RAMEAU, LCSH) and proposed a model for integrating time retrieval that included the creation of a chronological authority file. Providing multilingual subject access is significantly more difficult when non-European languages are considered, and Park (2007) outlined cultural and linguistic challenges to providing crosslingual name and subject access between English and Korean.

Classification

Recent classification research included considerations of the predominant Library of Congress Classification (LCC) and Dewey Decimal Classification (DDC) schemes as well as of other well-known, obsolescent, and local systems—and even the presentation of an entirely new scheme.

Subrahmanyam (2006) documented strong consistency in the application of LCC numbers, examining a sample of 200 titles as held by 52 American library systems, and finding a probability greater than 85 percent of a given title “having the same LCC-based class number across library systems” (110). Davis (2002) described a Columbia University Libraries project to create a “hierarchical interface to Library of Congress Classification” (HILCC) in order to improve subject access to electronic resources, and Chandler and LeBlanc (2006) explored a larger-scale implementation of HILCC by the Cornell University Library. Zhao (2004) discussed some common difficulties and confusion in book number assignment deriving from LCC’s reliance on the Cutter Table. Mitchell and Vizine-Goetz (2006) edited a special issue of CCQ devoted to the DDC considering aspects of its specific content (online use, the Relative Index, teaching DDC) and of its context (as shown by several international perspectives on DDC use).

Attar (2002) compared some classmarks from Bliss’s Bibliographic Classification (BC2) scheme with their counterparts in DDC, LCC, and UDC to argue that BC2 offers greater precision and brevity, despite the practical disadvantages stemming from its lack of institutional support. The influence upon BC2 and the work of Ranganathan exerted by the now overlooked Subject Classification (SC) system created in the early 20th century by James Duff Brown was outlined in Beghtol (2004). Winke (2004) researched the surviving institutional presence of another almost obsolete scheme, Charles Ammi Cutter’s Expansive Classification (EC), finding four libraries where EC remained the primary scheme and twenty-three others where EC still had some limited use.

Szunjeiko (2003) described the two special literature classification schemes used by the University of Western Australia and Murdoch University to mitigate the problem of DDC’s “inability to effectively express distinct classification for English literatures from countries other than England and the United States” (46). Lastly, Fadaie Araghi (2004) introduced his own ambitious new classification scheme—based on hierarchism and binary theory—called Universal Binary Classification (UBC).

Emerging Trends for Further Research

The ever-accelerating growth and development of the Internet has spurred much interest in Web 2.0 and Semantic Web technologies, both of which could have large ramifications for subject representation. Neal (2007) edited a special issue of the Bulletin of ASIST devoted to folksonomies, the application of “social tagging” by Internet users to provide uncontrolled but sometimes very useful metadata to Web resources. Similarly Greenberg and Méndez (2007) edited a special issue of CCQ that included an article by Harper and Tillett (2007) on the relation of LC controlled vocabularies to the Semantic Web.
Much research remains to be done comparing the relative strengths of controlled vocabulary metadata versus social tagging. Meanwhile emerging OPAC software (so called “next-gen catalogs”) has begun to implement folksonomy or tagging aspects. Usability studies of next-gen catalogs will be needed to determine whether these efforts to make catalogs align more closely to users expectations actually improve their usefulness for identifying and retrieving relevant subject content.

Works Cited


Hanson, F. Allan. 2004. From classification to indexing: How automation transforms the way we think. *Social Epistemology* 18, no. 4: 333–56.


Szunejko, Monika. 2003. Literature classification schemes at two West Australian university libraries: Murdoch University and the University of Western Australia. *Cataloging & Classification Quarterly* 36, no. 2: 45–57.


Prepared by **Alex Thurman** (at2186@columbia.edu), Catalog Librarian, Columbia University Libraries.