The Taxonomy of Maritime Information

Christopher H Hoebeke, World Maritime University
“The Taxonomy of Maritime Information”

The Maritime Librarian

4 March, 2013

I was recently cataloging a work on port state control, for which there is no term in the Library of Congress' list of Subject Headings, so I looked up the book in the Library of Congress catalogue to see how it was categorized in their system. To provide access by the subject matter, the LOC catalog record applied the following five headings:

- Harbors—Safety regulations.
- Ships—Safety regulations.
- Ships—Inspection.
- Jurisdiction over ships at sea.
- Law of the sea.

While all of these subject terms are relevant, none of them are on target, nor do they accurately hit the mark even when taken collectively. A good cross-referencing system would of course direct researchers to these broader, narrower and related terms, but students of maritime affairs researching such a topic would consistently, if not unanimously, start their search with the term "Port state control." LCSH does not even have a SEE or USE reference from this term, so searching or browsing by subject terminology in this instance is for all intents and purposes a useless endeavor.

I could provide plenty of other examples, but this single instance demonstrates the library's need to supplement LCSH with terminology actually used in the maritime sector. Today's researcher might ask why, in the era of online searching, when the entire text of a document can be searched, there is any need to apply subject terminology at all. Is it just slavish dedication to the practices of the past, with librarians transcribing and recording information according to the traditions of their monkish predecessors, simply for the sake of continuing the traditions?

I would answer that question with another: What if the document being searched nowhere uses the term that the user is searching for, but the content itself is exactly what the user is looking for? What if, for example, an author wrote an essay using the terms "ocean pollution," "pollution of the sea," "oil
spill in the Gulf of Mexico," and even made a reference to "MARPOL," but never once employed the term "marine pollution," which happened to be the term that the online researcher was using to find material on this subject?

Or consider the opposite problem, when the user searches for a term widely accepted as authoritative in his profession or field of study, but the same term is also used extensively in an entirely different context. Searching the web for documents on "piracy," for example, can be as frustrating for the user researching acts of kidnapping and robbery at sea as for the user researching violations of copyright.

Such problems are common enough when searching full-text documents. In the case of most library catalogues, where the printed text of the book is on the shelf and only a brief description of it is available for online searching, the odds of users missing out on relevant materials or having to wade extensively through irrelevant materials would be at least as great.

The librarian attempts to provide the researcher with greater "recall" (comprehensiveness of results), without sacrificing "precision" (relevance of results), by means of a "controlled vocabulary." The idea is that a single term is selected to represent a concept, regardless of whether that term is used in the document itself, and all documents using variations of the term, variations of spelling, synonyms, etc., are assigned to that term. To aid users, those variations and synonyms are often given entries in the database which point the user to the "authoritative" term.

Controlled vocabulary is likewise used to distinguish between homographs (identically spelled words with different meanings), or to distinguish between two or more concepts that use the same words (a process known as "disambiguation"), usually by appending a clarifying word or phrase. For instance, Library of Congress Subject Headings distinguishes the concept of "Piracy" on the high seas from "Piracy (Copyright)."

The cataloger or indexer is guided in the selection of controlled vocabulary by means of a thesaurus. The thesaurus can also be used within the online search system itself to provide cross references that direct users from synonymous terms to the proper "heading" or to indicate distinction of concepts. The more comprehensive thesauri also will point users (and catalogers assigning subjects) to broader, narrower and related terms.

Needless to say, maintaining a thesaurus is a great deal of work. Library of Congress has been maintaining LCSH since 1898, and its list of established headings now numbers in the millions. For general works, and many specializations, there is probably no better thesaurus, but as noted at the
outset of this article, its vocabulary is limited and imprecise in many areas having to do with maritime research.

While the WMU library is not staffed to build and maintain a thesaurus of maritime terminology from scratch, it is looking into existing work that could be used as a supplement to LCSH. The Transportation Research Board of the National Academies maintains the *Transportation Research Thesaurus*, or TRT, which obviously bears looking into. Also, Professor William Tetley's *Glossary of Maritime Law Terms* is hosted on the McGill University web site. The glossary is in its second edition, and provides excellent cross references. Unfortunately, it does not appear to have been updated since June of 2011.

But it does include the term "Port state control."