

Warrant for Concepts in Classification Schemes

Pauline Cochrane

Graduate School of Library and Information Science
University of Illinois
410 David Kinley Hall
1407 W. Gregory Drive
Urbana, IL 61801

INTRODUCTION

In 1911 E. Wyndham Hulme, in a series of articles in *Library Association Record* (reprinted in R.K. Olding, *Readings in Library Cataloging*, Shoe String Press, 1966, p. 108-140) had this to say about classification and literary warrant:

All classification is a means to an end...book classification is a mechanical time-saving operation for the discovery of knowledge in literature...the real crux of book classification [is] the nature of class headings and the principle upon which their scope or area is to be determined. Class headings are definitions of specific areas of the literary field...Can definition be based upon method and reduced to rule?...**What is to be the warrant for the areas of class headings?**...The warrant must be based either (a) upon considerations of the nature of the subject-matter to be divided, or (b) upon the physical fact of the aggregation of subject-matter in books.

According to Mill ("Logic" 6th ed. Vol. I, p. 135), subject matter is almost indefinitely divisible. For the power of the mind to frame distinctions is practically without limit. From its nature, therefore, subject-matter is singularly ill-adapted to our purpose. A classification based upon this principle (the nature of the subject-matter) would in practice lead to a universal index of minutely divided subject headings and to *the abolition of all general headings* — a scheme revived from time to time by indexing enthusiasts, but which for library purposes may be safely dismissed as an economic absurdity... Like Chemistry most all other sections of literature has a division which is determined mainly upon formal and non-philosophic lines. Books, in short, are concrete aggregates of facts selected from the common stock of knowledge, and are produced under the laws of supply and demand to meet the wants of the various bodies of the community. The result is a welter of cross classifications and of overlapping areas of definition, for the reception of which the frame-word of philosophic classification is quite insufficient. Hence we must turn to our second alternative which bases definition upon a purely *literary warrant*. According to this principle definition is merely the result of an accurate survey and measurement of classes in literature. A class heading is warranted only when a literature in book form has been shown to exist, and the test of the validity of a heading is the degree of accuracy with which it describes the area of subject matter common to the class. Definition, therefore, may be described as the plotting of areas pre-existing in literature. To this literary warrant a quantitative

value can be assigned so soon as the bibliography of a subject has been definitely compiled.... We must arm our classifier with certain limited discretionary powers: — To amalgamate under a common definition words of slightly differing areas; to register by duplicate or plural entry works containing subject matter, the association of which in books is shown, as a result of survey, to be infrequent, accidental, or purely fanciful.... the strength of the warrant varying with the number of works conforming to the type of each class definition.

More than eighty years have passed since Hulme asked for surveys which would determine quantitatively what classes would be defined in library classification systems. Depending on how you are counting we are into the third or fourth decade of automated retrieval systems and online library catalogs wherein such surveys could be done, but editors of library classification systems and of library subject heading lists like LCSH have not adopted the notion that these two systems should be integrated the way GeoRef and PsychInfo systems have. The alphabetic index to the LCC or NLM classification scheme is not identical to the descriptors/subject headings used in cataloging records produced by these two libraries and their authority files do not show all the linkages you might want to see between these two conceptual tools, a classification scheme and a descriptor/subject heading list.

Thesauri, like the *ERIC Thesaurus* have always had a categorized list of descriptors and in the printed and online form, the number of "hits" for each descriptor and category can be obtained. Many secondary services with databases online have mounted their thesaurus as a separate file which can be accessed. This is true for MeSH, but not for the NLM Classification Scheme. MeSH with its tree structures (see Figure 1) does provide a classified outline but this does not cover the content of books in the CATLINE database.

You would think by now we might have devised some kind of management information system which would collect data about concepts indexed in our databases and provide some structured analysis which resembles a library classification outline. But, alas, that is not the case. Services with a subject authority file integrated into their automated cataloging and indexing can produce matchups with a thesaurus and count number of times a descriptor has been used, but there are few if any systems which provide this information in a classified outline so that we could see Hulme's class definition exercise at a glance.

If we were to attempt such a report at the Library of Congress we would need to integrate LCSH (Library of Congress Subject Headings) and LCC (Library of Congress Classification) in a way similar to Figure 2¹. Such a display of literary warrant juxtaposed on the LCC schedule might be able to help the hierarchical structure of LCSH which has been criticized for its many weaknesses.

1. Quite often LCSH subject headings close together in the alphabet will not be placed in the "hierarchy" (BT-NT) list of each other. The class number associated with the terms, when present, does show this relationship. In this figure the range for Mineralogy is QE351-QE399.2, for Mineralogists, the class number is within that range, QE361, and for Mineralogy, Determinative, the range is within that range too: QE367-QE369. Looking at the classification outline, these concepts can be viewed in the overall conceptual framework for the field. Scanning the NTs under Mineralogy subject heading in LCSH makes one wonder how they were chosen out of all the other possible concepts shown in the classification. Is it literary warrant?

PROCEEDINGS OF THE 4th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

D24 - CHEMICALS AND DRUGS-IMMUNOLOGIC, BIOLOGIC FACTORS

IMMUNOLOGIC AND BIOLOGIC FACTORS (NON MESH)

BIOLOGICAL FACTORS (NON MESH)

VENOMS

SNAKE VENOMS

ELAPID VENOMS

ELAPID VENOMS

BUNGAROTOXINS

COBRA VENOMS

COBRA NEUROTOXINS •

DIRECT LYTIC FACTORS •

HYDROPHID VENOMS

ERABUTOXINS •

VIPER VENOMS

BIOLOGICAL PRODUCTS

ANTITOXINS

ANTIVENINS

DIPHThERIA ANTITOXIN

TETANUS ANTITOXIN

BACTERIOCINS

CLOACIN •

COLICINS

MEGACINS •

PYOCINS •

CATGUT •

COAGULASE

IMMUNE SERA

ANTILYMPHOCYTE SERUM

ANTIRETICULAR CYTOTOXIC SERUM

LECTINS

ABRIN •

CONCANAVALIN A

PHYTOHEMAGGLUTININS

POKEWEED MITOGENS

RICIN

WHEAT GERM AGGLUTINI

PICIBANIL •

PLANT EXTRACTS

DRUGS, CHINESE HERBAL

STREPTOKINASE

TOXOIDS

DIPHThERIA TOXOID

STAPHYLOCOCCAL TOXOI

TETANUS TOXOID

VACCINES

BACTERIAL VACCINES

BCG VACCINE

BRUCELLA VACCINE

CHOLERA VACCINE

PERTUSSIS VACCINE

D24.185.965.850.325

D24.185.965.850.325.139

D16.653.154

D24.185.965.850.325.220

D24.185.965.850.325.220.190

D24.185.965.850.325.220.260

D24.185.965.850.480

D24.185.965.850.480.345

D24.185.965.850.960

D24.310

D24.310.121

D24.611.125.

D24.310.121.163

D24.611.125.

D24.310.121.388

D24.611.125.

D24.310.121.824

D24.611.125.

D24.310.125

D20.85.133

D24.310.125.190

D20.85.133.

D24.310.125.230

D20.85.133.

D24.310.125.531

D20.85.133.

D24.310.125.759

D20.85.133.

D24.310.132

E7.858.784.

D24.310.200

D24.310.401

D24.611.125.

D24.310.401.203

D22.569.137 D24.611.125.

D24.310.401.371

D24.611.125.

D24.310.545

D24.185.526. D24.611.125.

D24.310.545.60

D12.776.765. D24.185.526. D24.611.125.

D24.310.545.358

D24.185.526. D24.185.526. D24.611.125.

BCG VACCINE

D24.310.894.135.128

D24.611.59.102

only /admin /adv eff /anal /class /hist /immunol /isol /pharmacol /rad eff /stand /supply /ther use /tox; BCG vaccination in prev of tuberc = BCG VACCINE (IM) + TUBERCULOSIS /prev (IM); BCG ther of various diseases = BCG VACCINE /ther use (IM) + dis /ther (IM)

79(75); was set under TUBERCULOSIS VACCINES 1974-78;

TUBERCULOSIS VACCINES was heading 1965-78; BCG VACCINATION was heading 1963-79

use BCG VACCINE to search TUBERCULOSIS VACCINES & BCG VACCINATION back thru 1966

X BACILLUS CALMETTE GUERIN VACCINE

X CALMETTE GUERIN BACILLUS VACCINE

X TUBERCULOSIS VACCINES

D24.310.894

D24.310.894.135

D24.310.894.135.128

D24.611.59.

D24.310.894.135.134

D24.310.894.135.222

1966-1996
Hitz:
8996

Figure 1a: Medical Subject Headings (MeSH)

PROCEEDINGS OF THE 4th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

) = 22 - 10015

MINERALOGY [351 - 399.2] ^{QE}

For crystallography, see QD901+

- 351 Periodicals, societies, congresses, serial collections, yearbooks
- 353 Collected works (nonserial)
- 355 Dictionaries and encyclopedias
- 357 Nomenclature, terminology, notation, abbreviations
History
- 359.A1A-Z General works
.A2-Z By region or country, A-Z
- 361.A2A-Z Biography
.A3-Z Collective
Individual, A-Z
e. g. .C5 Clarke
Dana, see QE22.D6
.H2 Haidinger
- .2 Directories
- 362 Early works through 1800
General works, treatises, and textbooks
1801-1969
1970-
- (12) 363 Pictorial works and atlases
(76) .2
(4) .8
- (13) 364 Special aspects of the subject as a whole
Mineralogical chemistry, see QE371
(17) .2
Special topics, A-Z
(5) .E4 Electronic data processing
(F47) Fluid inclusions
.F5 Fluorescent minerals
.H4 Heavy minerals
(2) .M4 Metasomatism
(9) .P3 Paragenesis
.R3 Radioactive minerals
Rock-forming minerals, see QE397
.S7 Statistical methods
- (34) 365 Popular works
- (32) 366 Study and teaching. Research
(12) .2
(16) .8 Collecting and preservation
Handbooks, tables, formulas, etc.
- Determinative mineralogy
Cf. QD87, Blowpipe
Assaying, see TN550+
- (35) 367 General works, treatises, and textbooks
(30) .2
(10) 368.9 Special aspects of the subject as a whole
(22) 369 Special topics, A-Z
.C6 Colloidal determinations
.D5 Differential thermal analysis
.F4 Fedorovskii method
.I55 Immersion method
.M5 Microscopic determinations
(O6) Optical determinations
.S6 Specific gravity determinations
.S65 Spectrum analysis
.U4 Ultraviolet
.X2 X-ray powder
- (371) Mineralogical chemistry
Descriptive mineralogy
General works, treatises, and textbooks
- 372 1801-1969
.2 1970-

Figure 2a: Library of Congress Classification (LCC)

PROCEEDINGS OF THE 4th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Note: Quite often LCSH subject headings close together in the alphabet will not be placed in the "hierarchy" (BT-NT) list of each other. The class number associated with the terms, when present, does show this relationship. In this figure the range for Mineralogy is QE351-QE399.2, for Mineralogists, the class number is within that range, QE361, and for Mineralogy, Determinative, the range is within that range too: QE367-QE369. Looking at the classification outline, these concepts can be viewed in the overall conceptual framework for the field. Scanning the NTs under Mineralogy subject heading in LCSH makes one wonder how they were chosen out of all the other possible concepts shown in the classification. Is it literary warrant?

Mineralogists (*May Subd Geog*)

- {QE361}

BT Scientists

Mineralogy (*May Subd Geog*)

- {QE351-QE399.2}

BT Physical geology

RT Crystallography

Minerals

NT Asterism (Crystallography)

- Fluid inclusions

Lunar mineralogy

Metallogeny

Metasomatism (Mineralogy)

- Mineralogical chemistry

- Paragenesis

Photography in mineralogy

Soil mineralogy

Thin sections (Geology)

MINERALOGY, *DETERMINATIVE*

Used for:

Chemical geology

DETERMINATIVE MINERALOGY

Geology, Chemical

Narrower terms:

Assaying

Optical *MINERALOGY*

Ores--Sampling and estimation

Refractive index of minerals--Immersion method

Related terms:

Chemistry, Analytic

Broader terms:

Geochemistry

Metallurgical analysis

Prospecting

- Call Number Ranges:

QE367 - QE369

Figure 2b: Library of Congress Subject Headings (LCSH)

It is not the equivalent of the tree structures in MeSH which benefit greatly from constant attention and revision.

With such a system (with links to the classification and the descriptors in the thesaurus) there could be automatic updates to captions or notes in the classification schedule and links from class number to descriptors in the thesaurus thereby providing what Hulme called "general headings." Without such general headings, our present "free text" searching systems offer no real gathering devices or displays of knowledge outlines.

If there were reports of literary warrant in our databases their analysis could be useful to the lexicographer responsible for the syndetic structure (BT-NT relationships) in the thesaurus, to the classificationist responsible for additions to the classification schedule, and to the retrieval system designer who needs new ways of displaying the structure of the searching vocabulary so that "too few hits" or "too many hits" can be mediated. Such reports would help put every term in the database's basic index "in context", showing how often it has been used and what other terms are near at hand to expand or delimit the search.

Such a report would also show what parts of the conceptual framework in the thesaurus or classification scheme do not have much coverage in the database, thereby allowing for revision of acquisition policies or of hierarchical specificity.

Adding data from actual catalogs to help monitor the concept distribution in the classification scheme could contribute toward our understanding of "warrant for concepts in classification schemes." With such understanding could come more helpful arrays of concepts in our user-oriented searching systems.

As it now stands traditional library classification schedules gather many concepts under class numbers which are quite useless because they have captions that read as follows, with no background data on specific concepts gathered by such class numbers:

**General works,
Special ... A-Z,
...(General or not indexed elsewhere),
By place, A-Z, etc.**

In every case, for online retrieval and outlines of concepts such captions would have to be revised to show the descriptors used for indexing at that class number. This would help to do what Hulme called "class definition."

The following examples illustrate this point.

Examples of Linkages between Thesaurus and Classification System

Example 1.

In the NLM Classification there is the following class number:

QW 168.5 Specific RNA groups, A-Z

After a review of the database one could construct a list of all the RNA groups classified under this number, adding those terms to the classification index and verify that each and every one is in MeSH. For example,

QW 168.5.B9 Bunyaviridae

QW 168.5.C8 Coronaauridae

These "Divide A-Z" class numbers exist in great numbers in both the NLM Classification and LCC.

Example 2.

Often a cataloger will classify an item under a number where it is not clear if that number includes the concept represented by the MeSH term also attached to that item. In such a case there should be a way to automatically link that MeSH descriptor with that class number and the index to the classification schedule. For example,

In the NLM classification schedule:

WL 355 Cerebrovascular disorders
e.g., Cerebral hemorrhage

A cataloger classifies an item on Cerebral infarction under this number. That action should trigger a new index entry in the classification schedule's index for that term to that class number and the caption's list of examples should include this term.

Example 3. Another example related to 2:

In the NLM classification schedule:

QV 350 Antibiotics (General or not indexed elsewhere)

A cataloger classifies an item on *Cephamycins* under this number. That action should automatically add an index entry in the classification schedule's index and the caption might include a note which reads: Includes Cephamycins...

MULTIPLE PLACEMENT OF CONCEPTS

It is a well known fact that concepts represented by descriptors in a thesaurus will be assigned to items which are classified under several different class numbers. (The quote from Hulme above noted that this calls for certain classifier discretion.) A formula might be worked out which could be used in reports to indicate where such terms are used *most* often. This would help with studies of concept spread.

Example 4.

Acquired Immunodeficiency Syndrome has the following MeSH tree structure:

C1.539.780.20
C2.782.815.483.23+
C20.673.483.23+

This shows concept spread, and sure enough, items on this subject have been classified in the NLM Classification under several numbers, but this is what the CATLINE database shows:

CATLINE postings (3 or more):	QW 166	—	4
	W 1	—	16
	WD 308	—	133
	ZW 1	—	1
	ZWD 308	—	13
	WY 150	—	3

Several of these class numbers are for "form", e.g., serials and would be disregarded in any concept spread studies. WD 308 is the outstanding class number for a link between the class number and the MeSH term, with an index entry in the classification schedule, unless that is, the Principal Cataloger, would determine that this concept needs a new class number to add greater specificity to the outline of information arrayed by the classification.

Example 5.

Alzheimer's Disease has 77 postings in CATLINE, 52 items were classed under WM 220, which has the caption: Organic (General or not indexed elsewhere).

After receiving such a report, the Principal Cataloger might add an index entry to this class number from Alzheimer's Disease OR better still, consider a new class number for this concept since it has so many postings (a thesaurus/classification maintenance program might track such postings and provide a report after 50 postings — or any other established threshold).

Example 6.

The Library of Congress Classification uses a range of numbers quite often as a gathering device, e.g.,

SK 295-305 [Hunting Sports] Big Game.

Bear, Buffalo, Chamois, Deer, Moose, and Elk are delineated with separate class numbers in this range. All other game is grouped in SK 305 Other, A-Z.

A printout of class numbers which occur 10 or more times in the Library of Congress catalog that only Bear and Deer have more than 10 items (15 and 18 under their respective class numbers), and 65 items were grouped under SK 305. It would appear from such a report that the separate classing of Buffalo, Chamois, Moose and Elk is unwarranted and that the SK 305 grouping should be analyzed and possibly new groups of big game should be defined. If a threshold of 20-25 items is tolerable, perhaps all these big game groups should be under SK 305 where they are each differentiated by a Cutter number (e.g., .A35 African buffalo, .B45 Bighorn sheep, etc.

READYING CLASSIFICATION SCHEMES FOR SUCH LINKAGES

All the suggestions made here relate to using the classification in online retrieval systems and not for the shelving of books. To change every item's shelf number on the basis of decisions reached in the examples above would be, to use Hulme's words, "an economic absurdity." No, what is suggested here is more like a classified catalog, where class number changes and additions are made to better group items for retrieval. Shelf numbers should not stand in the way of such improvements.

Nancy Williamson, Karen Drabenstott, Lois Chan and others have written about the new role classification schemes could have in online retrieval, but Janet Swan Hill has thrown in a note of caution because the classification schedules need editing for use in a new environment. The work on the MARC format for classification schedules did not address these issues, but the DDC Editorial Office staff are trying to edit the Dewey Decimal Classification so that it is more useful in online retrieval. Eventually we may see improvements, but first there needs to be some awareness of what is needed and what is possible. Hulme's principle of literary warrant is a good place to start.

PROCEEDINGS OF THE 4th ASIS SIG/CR CLASSIFICATION RESEARCH WORKSHOP

Further investigation into concept spread, links between concepts in the thesaurus and in the classification schedule will be useful, given the direction our retrieval systems are going. Full text searching and free-text searching need backup assistance that an organized presentation of concepts, such as a classification scheme, can do.

