FAST TRANSITIONS IN OPACS -- From Subject Heading

Lists to Faceted Strings

A talk by Pauline Atherton Cochrane
to the UIUC Library Colloquium Group, 5-12-06

Abstract:
As new software like Endeca (used at North Carolina State University) and FAST (Fast Search and Transfer) (used at Univ. of Bielefeld) becomes available to provide more flexibility with metadata in MARC records and use of authority files (such as LCSH, NA, etc.) librarians will need to consider improving the structure and display of subject terms and their displays in OPACS. Using FA (Facet analysis) or FAST (Faceted Analytical Approach), based on FAST (Faceted Analytical-Synthetic Theory) coupled with the deconstruction of subject heading strings in existing MARC records, our databases can be improved and take advantage of these new search and navigation capabilities.

---------

For me there is a personal and professional thread running through all of the developments which lead up to this talk today. It has to do with student-teacher relationships, just as Atoma Batoma and I are related. I thank him for thinking of his old teacher when he became involved in the plans for this forum.

My lifelong professional quest has been for Improved Subject Access. This was reinforced on my birthday in December, 1959 when S. R. Ranganathan came to the Graduate Library School at the University of Chicago while I was a Ph.D. student there. I still have the notes I typed up after his talk which was all about FA (Facet Analysis - a procedure for indexing) and FAST (Faceted Analytical-Synthetic) Theory, the underlying theory for doing FA correctly. Forever after, Ranganathan has been my teacher and I his student. By the next year, 1960, I was presenting his ideas at a Classification Research and Study Group (CRSG) meeting during ALA Annual Conference at Montreal (this was published in 1965 in LRTS). Now in the very twilight of my own career one of my students invites me to talk about the same subject because it has relevance in today’s context of search engines and websites.

Faceting - whether it is based on Ranganathan’s FA and FAST or on some other scheme is a hot topic for Information Architects and Website designers. Why? Because these designers recognize the value it offers for searching and navigation. Take a Wine website for example where you can search for wine by color, region where the vineyard is, price, and use with a meal. This is a website with a faceted vocabulary for searching. When we contrast this and other commercial websites we begin to notice how limited the present generation of OPACS are and how few vocabulary aids there are for the catalog searcher. How, we ask, can libraries get such a faceted search capability working in their OPACS? It will take some effort on our part to improve our databases before we can take advantage of what the software can do.

Faceting is a structured approach to metadata and to controlled vocabularies within the metadata. This feature has been recognized in the new ANSI/NISO Standard for Controlled Vocabularies (239.19-2005) and I have excerpted sections of that standard to give you official statements about faceting in the electronic information world. This is
not a far cry from our tagged and fielded bibliographic information in MARC records but it is very far beyond the MARC-tagged 650 field where the LCSH assigned string for the content object is marked up and placed. (see Figure 1, p.12 of the Standard (in 5.2.2.) and section 5.3.4)

I will be describing faceting topical metadata in more detail, but first I must report some confounding developments, also using the acronym FAST, which are not really about faceting as this standard describes it. One FAST project from OCLC purports to build a file based on the application of “faceting” to LCSH strings found in MARC bibliographic records. Another use of the FAST acronym describes Norwegian software (Fast Search and Transfer). It is being sold to libraries in Europe.

The OCLC project called FAST (meaning “Faceted Application of Subject Terminology) is really a poor example of what faceting can do for a topical vocabulary because it is based on inconsistent practices of faceting associated with LCSH over the years. I will explain why I say this in a few minutes.

Along with these confusing developments there are some positive developments of which the most important is the new ANSI/NISO 239.19.2005 standard on Controlled Vocabularies. This will impact how FA, FAA, or FAST will move into the library world.

the second development is a new CLW (Cataloger’s Learning Workshop) on Thesaurus Construction, sponsored by Library of Congress and ALCTS, the first of which will be held in a month;

the third is the new book (and an article based on one chapter of the book) by Prof. James Anderson entitled Information Retrieval Design. He explains how faceting (based on Ranganathan and Bliss principles) will improve the syntax of LCSH in OPACs and give digital libraries searching features which will be as powerful as any being used on the commercial websites.

the fourth development came to my attention through a former student and UIUC librarian, Karen Fischer, now in Iowa. Karen sent me a report which appeared on the Library Journal Academic Newswire for April 20,2006: It seems a provocative report is stirring up discussion in the ARL about whether or not to jettison LCSH (another student, Sandy Roe, brought me a copy of this report this morning). This report by Karen Calhoun of Cornell University was commissioned by LC for ARL consideration. Its title is, “The Changing Nature of the Catalog and Its Integration with Other Discovery Tools.” After interviewing 23 experts, Calhoun concludes that three potential strategies are possible: Extend, Expand, and provide Leadership.

Consider my description of faceting today as a way to extend and expand LCSH’s usefulness in the catalog thereby insuring its continued existence. Faceting will definitely extend and expand subject access to print and digital resources.

My observations today about FA rules for faceting LCSH strings in MARC bibliographic records and turning LCSH into a fully faceted vocabulary tool will be brief, but I strongly recommend that next year
your Colloquium Committee invite Kathryn La Barre and James Anderson to continue the discussion. (Kathryn is a new member of the GSLIS faculty. Her Ph.D. dissertation at Indiana U. covers the very topic of FAST on the websites I mentioned; Prof. Anderson, now retired from Rutgers, is a well known author and teacher, and I am relying on his work today to get us started thinking about faceting.

What I have always liked about our field of librarianship is that we are open to new ideas. Sometimes we even compare these new ideas with old ideas and then adapt and combine the new and old. That is what is happening now with the idea of faceting.

In the 1960s S. R. Ranganathan, Jack Mills (editor of the Bliss Classification) and others, were imbedding the FAA (Faceted Analytic Approach) into their efforts at vocabulary development for catalogs and indexes. Some took one more step and developed this vocabulary into a Faceted Classification (FC) for good record keeping. Many British librarians adopted FAA and developed faceted vocabularies and faceted classifications, but few did in the United States.

While Ranganathan and others were expounding FA and FC, the Library of Congress kept on following inconsistent procedures and even separated its classification effort from its alphabetic subject headings. This was a serious mistake in my opinion. While Ranganathan talked about FA as an approach to subject analysis that was rigorous and disciplined the Library of Congress continued without that discipline. FA provides the structure you need to be sure that all facets of the subject in the content object being analyzed are indexed. All through this period LC was being criticized for having too few subject headings on MARC bibliographic records and an unstructured vocabulary in LCSH. Ranganathan could see that FA provided an analytical record of terms which could be rotated to provide many access points. (He saw this before we had computers to do this so easily.)

The standard for thesaurus construction originated in the 1960s and immediately we saw displays of the vocabulary, beyond a merely alphabetical list like LCSH. These displays showed the structure and relationships within the vocabulary. The current standard (on page 15) suggests such a faceted structure if your vocabulary is large, and LCSH has hundreds of thousands of terms.

Anderson says that we can repair what was not done consistently by subject catalogers using an unfaceted LCSH. He is suggesting a Bottom-Up FA approach. From this effort could come an improved database of MARC bibliographic records. Possibly, this would also result in a truly faceted vocabulary, similar to what Ranganathan suggested as a Top-Down approach, using FAST principles and procedures. Maybe, just maybe, Anderson’s suggestion for improved LCSH faceted syntax will work toward that larger goal.

To get there from here we need to admit certain failings:

(1) We need to admit that the LC/Subject Cataloging Manual gives us LCSH rules which lead to inconsistencies in practice.
(2) We need to admit that the 1980s "thesaurus markup of LCSH," did not result in correct term relationships and did not really help to create a faceted thesaurus out of LCSH; and

(3) We need to admit that LCSH subdivisions and the construction of main headings has grown like topsy and no discernable faceted structure exists which could be automated the way the OCLC project would lead us to believe.

The present MARC format for the 6xx field in bibliographic records merely marks the parts of the LCSH string. It does not indicate what are topical facets and what type they are. Each type of main heading and subdivision has received a code which over time has been used in an ambiguous way. (Table I, the orange handout, shows the Types of Facets it would take to make LCSH facet labels useful.) There is now no LC/MARC practice which would result in a truly faceted string from LCSH. Codes like a & x subfields cover all topical facets in a non-delineated way; y, z, and formerly k and now v are useful but mix topical and non-topical facets. Resulting OPAC displays of LCSH strings of main headings and subdivisions are a horror. Viewing screen after screen of these displays in a seemingly confusing and illogical sequence is not very helpful (sections of Anderson’s paper I have not included in the excerpts show this and I have also shown it in many a paper).

I am mentioning all this history about how new thesaurus developments were resisted or ignored by LC and the American library world because this all rests on our backs today as we talk about Faceting Strings of LCSH in OPACs. The new standards, workshops, and new software for OPACs and websites may help us move in a new direction, but we must first understand that there is a load we must get off our backs, and this backlog is in clear view in our catalogs. As we reflect on our old ideas and re-check to see if they fit in this new world, maybe we will be ready to expand and extend them.

The ANSI/NISO standard suggests that the metadata for content objects (and the controlled vocabularies associated with these metadata) should be faceted. This is the way to take every bit of descriptive data and manipulate it during the search, display, and navigation processes. A review of FIGURE 4 on page 14 of the standard explains what Information Architects are beginning to “facet” -- various metadata for Content Objects, content objects themselves, and the corresponding auxiliary “controlled vocabularies”. Yes, times are changing.

My focus today is only on the Topic Facet of Content Objects and on the Controlled Vocabulary associated with it.

The subject headings we choose for an item we are cataloging (e.g., LCSH or terms from any thesaurus) are one kind of metadata that can be “faceted” in this new environment. Another is the vocabulary like LCSH from which those terms come. We will start, as James Anderson starts, with the assigned LCSH strings, but the project won’t be done until LCSH, as a controlled vocabulary, becomes faceted following FAST principles. Since this latter project is so difficult to envision, let’s start with revising LCSH strings on existing bibliographic records, a la Anderson. (My excerpts come from Anderson’s
forthcoming CCQ article entitled "A Fully Faceted Syntax for Library of Congress Subject Headings". With his permission I am distributing pages 17-24).

Because we have clouded the issue by placing nontopical facets (like Language and format) in our Topical Facet, we must separate them out and label them as non-topical, albeit related, facets (e.g., format, medium, audience, approach).

Now let's concentrate on how we "type" our topical facets in these LCSH/MARC records. See Table I (orange handout) and the excerpt from page section 6.3.2 of the ANSI/NISO standard which provide examples. The Anderson paper shows a MARC record with these facet types in place with each term. (page 24) (This is what I had hoped the OCLC project called FAST (Faceted Application of Subject Terminology) would have done, but alas, it has not. At best I can call their effort PFAST - pseudo-faceted application.)

As you can see on Table I, different facet labels can be applied to the same facet of topical terms. What is chosen as facet labels will depend on how a particular field structures its terminology. Of course you can start with Generic Facets, like Entities, Attributes, and Actions. No one today would suggest Ranganathan's PIMEST because it is considered too esoteric, but the Bliss/Anderson facets (such as Things, Properties, Operations, Client, and Means) are good suggestions for a start. You can revise these labels as you go along. Nothing need be set in concrete at this point. What is very important, however, is that you practice FA a la Anderson and build a faceted string with syntax that will be meaningful to your users. The computer software will always allow for updates and revisions. Just remember that almost anything you do will be better than the existing "a" or "x" codes in LCSH for creating a faceted string of terms.

The rules for this faceting process helps to decompose the LCSH string in such a way that when you reassemble these strings, during a search and retrieval process, you can get a meaningful display which will guide the searcher to more relevant items in the library's collections. THE MARC record Anderson's paper shows on page 24 is the expanded and extended faceted LCSH string used to create better OPAC displays.

On page 17 of Anderson's paper (green handout) he shows a display, with a search term, its facets, and the number of items retrieved (in parentheses), all compiled from the revised LCSH faceted strings on the MARC bibliographic records. If you were to click on one of these facets in the display, you would see more details (see page 18). This display in turn, could either be in alphabetic order as Jim has it or it could be in some more logical order such as geographic proximity. If the latter display is warranted you would see a list (or maybe even a map) with: America/ North America/USA/California/Berkeley shown together; Europe/England/London; Spain, together etc. Such displays are very doable with today's software. Endeca is probably the best software for this right now and that is why I suggested you look at the North Carolina State University OPAC. It is supported by Endeca. As good as their displays are, they would be even better if they had done the faceting Jim Anderson suggests for the LCSH strings in their MARC bibliographic records.
If we follow the rules in James Anderson’s prototype project (pages 21-23) and do this FA work in fields of interest to us, we could revise our own databases and test retrieval with new faceted display capabilities.

Different fields of interest would be chosen by various cataloger/revisers at various institutions and this work could be coordinated, edited, and revised. As we did this we would be testing Anderson’s working rules, making display suggestions, and turning up more examples similar to what Anderson brought together in his book, based on the work of his students. Eventually we could borrow records with faceted LCSH strings from each other through some automated transfer and use them to create improved OPAC displays of our resources.

If the library world adopts the new ANSI/NISO standard for controlled vocabularies to replace LC’s-SCM we could eventually construct a totally faceted LCSH thesaurus/taxonomy.

Granted this is a lot of work, but it is the only way to take full advantage of the new software for OPACs which allows for faceted strings of metadata. So, let’s start. Some prototype projects do result in vast improvements. I did a prototype project 30 years ago where we showed what was possible if we augmented MARC records with Table of Contents and Back-of-the-Book Indexing. Well, thirty years later this augmentation, in part, is now a reality. John Byrum of the Library of Congress has reported on use of Tables of Contents in MARC records in the current issue of Information Technology and Libraries! I am glad I lived long enough to see this take place.

No matter what we call it: FA, FAA, or Faceting, FAST or LCSH revision or even Thesaurus construction, we should begin. Remember Calhoun’s report tells us to Extend and Expand. We need to do this so we can do better than Google or Amazon.com, so that our cataloging metadata will make our OPACs a better discovery tool.

In the future, when we have fully adopted the concept of a faceted topical vocabulary we may return to another old idea and consider it new - I am talking about the Thesaurofacet shown in the last handout. This is a display of the hierarchical and other relationships between terms in a controlled vocabulary as well as a display of the facets for that term. Jean Aitchison developed it for engineering information years ago. It shows the structural arrangement of terms around the term “Fuels” as well as the many facets associated with the term. Long after my time on this planet, we may see general OPAC displays such as this. (I blacked out the classification notation because I do not believe it is necessary for our users to see. However, such a notation may be helpful internally for computer manipulation. If it is, Information Architects will need some structural help for they will not be able to create this notation automatically. Experts using FC will have to be called in. Will they still be called “librarians” or “catalogers”? I can’t predict that far in the future. Maybe they will be called Ontologists or Taxonomists or Knowledge Managers - I don’t know.)
Thank you for this opportunity to give my first public lecture at UIUC. I hope my old and new thoughts about Faceting will help you extend and expand LCSH into a faceted control vocabulary and that you will develop new OPAC designs and improve subject access.
Table I. TYPES OF FACETS (Topical and Non-Topical)
*See Anderson excerpt, page 21-24, "Some working rules for a fully faceted LCSH.

FOR TOPICAL CONTENT

<table>
<thead>
<tr>
<th>SRR</th>
<th>ANSI/NISO</th>
<th>LCSH/FAST</th>
<th>BLISS/Anderson*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personality</td>
<td>Entities/Things/Parts</td>
<td>Topical=a,x</td>
<td>Things/Entities/Kinds</td>
</tr>
<tr>
<td>Matter</td>
<td>Attributes/Constituent Matls. Properties/States</td>
<td>Topical=a.x</td>
<td>Matls/Properties/Part</td>
</tr>
<tr>
<td>Energy</td>
<td>Actions/Activities/Operations/ Processes/Events</td>
<td>Topical=a,x</td>
<td>Process/Operations/ Client/Product/Means</td>
</tr>
<tr>
<td>Space</td>
<td>Places</td>
<td>Geog=xy</td>
<td>Space</td>
</tr>
<tr>
<td>Time</td>
<td>Times</td>
<td>Chrono=xy</td>
<td>Time</td>
</tr>
</tbody>
</table>

NON-TOPICAL FACETS:

<table>
<thead>
<tr>
<th>ANSI/NISO</th>
<th>LCSH/FAST</th>
<th>BLISS/Anderson*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form(Type,Genre) (prev.=k,x,now v)</td>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>008 MARC field</td>
<td>Audience</td>
<td>Approach (Method, Point of view, etc.)</td>
</tr>
</tbody>
</table>

Table II. ACRONYMS SCORECARD

FAST - Faceted Analytico-Synthetic Theory
FAST - Fast Search and Transfer (a Norwegian OPAC software)
FAST - Faceted Appln of Subject Terminology (OCLC/LCSH)

Processes leading to faceted displays of vocabularies:
FA - Facet Analysis
FAA - Facet Analytical Approach
FC - Facet Classification and Thesaurofacet