Most library students will have studied Ranganathan and faceted classification, but unless they live in India, they are unlikely to have read his works in detail or used his Colon Classification scheme. Nonetheless, his groundbreaking work has been influential in traditional library classification schemes. Facets are fundamental to the Bliss Bibliographic Classification, and are important in the Dewey Decimal Classification.

Faceted classification depends on separating subjects into their component parts, and allowing access through one or more of those parts, according to user needs. It is considered to be an ideal approach for combining the best of browsing and searching online. It works particularly well for online retrieval, as facets can be combined post-coordinately (i.e. while searching), rather than having to be combined in a set citation order for shelf arrangement. Faceting is easiest to implement with uniform collections (e.g. wines, recipes) but may have the most impact in complex multidisciplinary environments.

**History of Faceted Classifications**

Ranganathan, an Indian mathematician and librarian, systematically developed the idea of facets in his Colon Classification, named after the punctuation used between facets. Ranganathan’s facets were placed in the order PMEST: Personality (the main focus), Matter (things), Energy (actions), Space (place) and Time.

An example on the SLAIS site shows the notation used to express the concept:

```
L,45;421:6;253:f.44’N5
```

```
Medicine,Lungs;Tuberculosis:Treatment;X-ray:Research.India’1950
```

(http://www.slais.ubc.ca/courses/libr517/winter2000/Group7/facet.htm)

The Bliss Bibliographic Classification uses the following facets:

*Purpose of the subject*
It's defining system, end-product, etc), then its Types, Parts, Processes, Actions and Agents. Bliss is a fully faceted classification scheme, whereas more commonly used classifications, such as Dewey Classification and Library of Congress Classification, apply aspects of facet theory but are not fully faceted.

Library classification schemes determine notations for specific facets, and then decide on the order of the facets within the notation (i.e. decide on the citation order). For example, the notation for 'nursing children with cancer' is HXO QEM Y, derived from HXO for Paediatrics, HQE for Cancer, and HMY for Nursing. (The initial letter for the class (H) is dropped when combining the subclasses; this example is from http://www.sid.cam.ac.uk/bca/bchist.htm).

Notational issues are mainly relevant to schemes used for shelving books, where notations determine the location of items on the shelf. On the Web, where searches can be used to combine selected facets in any order, faceted classification avoids these clerical difficulties.

Thesauri using faceted classification have also been used for many years. One of the best known is the Art & Architecture Thesaurus (http://www.getty.edu/research/tools/vocabulary/aat), now in its second edition. It is divided into seven facets, which are further subdivided into 33 subfacets or hierarchies.

Work on view-based searching and related topics has been done at the Polytechnic (now University) of Huddersfield since 1973 (http://www.view-based-systems.com/history.asp).

An early example was CANSEARCH, an expert system which included ‘a touch screen display formed by selected hierarchies for facets expected in queries for cancer therapy literature. A rule-based program controlled the interaction and formulated legal MEDLINE search statements.’ Other work was done with EMBASE (another medical literature database) and student records.
This general approach to faceted searching on the Web has been informed by the Flamenco Search Interface Project (FLexible information Access using MEtadata in Novel COmbinations; \url{http://bailando.sims.berkeley.edu/flamenco.html}) at the University of California, Berkeley. The project investigates the use of large faceted category hierarchies in user interfaces, and the site links to examples of fine arts images, architectural images and a tobacco documents archive. In ‘Finding the flow in Web site search’ (\url{http://www.sims.berkeley.edu/~hearst/papers/cacm02.pdf}), Marti Hearst and others discuss a project implementing search for architectural images, using combined browsing and searching. In usability tests, they found that users were satisfied with the system, despite the presence of unfamiliar features, which often deter users. Ideas from this research are now found in the commercial products Endeca, Siderean Seamark Server (previously bpallen Teapot), and i411 (see below).

The potential for the use of Facet Analytical Theory (FAT) to develop a ‘multidimensional network of subject terms for use with digital collections’ is being evaluated using a prototype developed by Arts and Humanities Data Service (AHDS) and the Humbal Humanities Hub (\url{http://www.ucl.ac.uk/fatks}). FAT is considered to be potentially useful for cross-disciplinary access to topics such as ‘the influence of philosophy on literature’ (\url{http://www.ahds.ac.uk/autumn_2002_newsletter.pdf}).

**What is Faceted Classification?**

Faceted classification means breaking subjects into standard component parts, or facets. For some topics, this is relatively easy – books have authors, illustrators, publishers, dates of publication, copyright dates, and so on. For other topics it is more complex, but it is usually possible to consider high level facets such as materials, processes, equipment and so on, and then to create a hierarchy with narrower terms. The groups are facets, and the terms used to describe individual items are values (or attributes) within these facets. So in the facet *author*, one value may be ‘Patrick White’, and in the facet *materials*, one value may be ‘steel’.

A faceted classification is based on a controlled vocabulary, and implementation requires indexing of documents with metadata from the controlled vocabulary. According to the *Search Tools Report*
Faceted classification often works well for e-commerce applications, where specific attributes are applicable to all products. For example, someone might want a toy for less than $10, or for a newborn baby, or with Harry Potter on it. They can find these by searching for the required values in the facets cost, age, and character. Faceting can also be applied to general subject access, although working out appropriate facets can be time consuming for more complex or less structured data. Nonetheless, it is in these complex areas that faceting can bring important benefits by allowing users to search on separate aspects of topics. In the example given above from the AHDS Newsletter, ‘the influence of philosophy on literature’ could be distinguished from ‘the influence of literature on philosophy’ and other topics described using the terms ‘philosophy’ and ‘literature’.

The best faceted classifications on the Web allow users to combine searching and browsing. At all stages breadcrumbs are displayed showing the path they have taken, and allowing them to backtrack if needed. When browsing, the user can refine searches by drilling down a hierarchy to more specific terms, or by adding values from different hierarchies. For example, you can narrow a search by drilling down from ‘vertebrates’ to ‘mammals’ to ‘whales’ in the animals facet, or by selecting ‘vertebrates’ from the animals facet and then selecting ‘Australia’ from the place facet. Because available options are presented to the user at each step, they only have to recognise the term of interest, not decide what search term to search on.

As the user refines his/her search, the display shows the number of hits at each point, letting the user know whether it is necessary to further refine the search, or whether the number of hits is small enough for the user to decide to look at the lot.

A straightforward Search option is also required for simple searches. For example, in a recipe database I might use the search facility to find recipes using kaffir lime leaves (of
which I don’t expect many), but use faceted browsing to combine different requirements (e.g. ‘roast chicken dishes suitable for children’). Some systems also allow you to search for all resources that do not have a certain facet (e.g. foods with no peanuts).

**Generally Applicable Facets**

Faceted classification could also be used to refine searches according to generally applicable facets such as format, user appropriateness, type of material (e.g. overview), genre, time and place. Automatic classification studies have found that computers can identify subjects much better than they can identify genres or predicted users (e.g. ‘for children’), while other research has found that computer searches are not good for identifying overview material. It is very difficult when searching on terms such as ‘indexes’ to separate documents about indexes from documents that are indexes. This is highlighted by the Kosmoi page on classification. It is a Web page about classification, but a column on the left headed ‘Amazon.com’ lists supposedly related books such as ‘Paterson First Guide to Caterpillars of North America’, ‘DSM-IV made easy: the clinician’s guide to diagnosis’ and ‘Crime classification manual’ (http://kosmoi.com/Technology/Web/Architecture/Classification).

Examples of Faceted Classifications on the Web

A number of Web sites, particularly e-commerce ones, now use faceted organisation. The first two examples below use Siderean Seamark Server, the third uses Endeca Guided Navigation, the fourth and fifth use the i411 Discovery Engine, and the last one uses individually developed software.
Faceted Classification

Written by Glenda Browne
Sunday, 02 September 2007 17:16 - Last Updated Sunday, 04 November 2007 11:28

Annotated Wordnet

At Annotated Wordnet (http://www.siderean.com/wordnet17.jsp) you can refine searches by ‘type’ (e.g. noun), ‘is a kind of’ (e.g. city; bird genus, writer), ‘is a part of’ (e.g. France, Africa, Texas) and so on. If you select ‘is a kind of city’, the next page lists cities, and then lets you refine by ‘is a kind of’, ‘is a member of’ (e.g. ‘Hanseatic League’, ‘Twin Cities’) and ‘is a part of’. The refinement options that are offered are those that are useful for this results set. There are 536 cities in the site. The refinement list on the second page shows that 109 of these are ports, 1 is a watering place, and 22 are in Mexico. This information lets the user know whether to select an item of interest or whether to continue refining the search.

DC- 2002 Dublin Core Conference

The online proceedings of the DC- 2002 Dublin Core conference (http://www.siderean.com/dc2002.jsp) can be searched by:

- **Subject by category**, e.g. ‘activities’, ‘organizations’
- **Subject**, e.g. ‘Dublin Core’, ‘RDF’
- **Creator**, e.g. ‘Jane Greenberg’
- **Time of event**, e.g. ‘14 October’
- **Type of event**, e.g. ‘Plenary Session’

As with the Wordnet site you can see how many hits will be retrieved for each selection. For example, there are 16 hits for ‘organizations’, 5 for ‘RDF’, 2 for ‘Jane Greenberg’ and so on.

Tower Records

At Tower Records (http://www.towerrecords.com) users first search by typing in a keyword or filling in a form with various categories (artist, guest artist, album title, song, label, and so on). Later pages allow for refinement of search according to facets which are displayed at the left of the page. These facets include genre, feature (such as ‘boxed sets’ or ‘in stock’), price and artist. Results are grouped into useful categories; for example, a keyword search for ‘Mozart’ retrieves hits with Mozart in the composer name, the Ensemble name, and so on. These
Faceted Classification

Written by Glenda Browne
Sunday, 02 September 2007 17:16 - Last Updated Sunday, 04 November 2007 11:28

are grouped by category, and within the category by specific name (e.g. 'Leopold Mozart') so the user can hone in on the specific name they want.

Dun & Bradstreet

Nineteen million businesses can be searched at the Dun & Bradstreet site (http://www.dnbbiz.com). One category is ‘Nonclassifiable Establish...’ filing in the N’s.

Genome Analyzer

The Genome Analyzer (http://genome.i411.com/leto/ok.asp) categorises 11,431 genes by biological process including ‘behaviour’, ‘cell communication’, and ‘sensory perception’. The last category, containing more hits than the rest combined, is ‘z-unclassified biolog...’.

Te Kete Ipurangi: the Online Learning Centre

Te Kete Ipurangi provides English and Maori sites for curriculum resources: http://www.tki.org.nz/e/search. This is a nicely presented site, which offers grouped options for selection. You first choose the ‘filter’ category (e.g. language, subject, education topic). You are then presented with keyword options applicable to that category. Keywords you select from the list, and keywords you add yourself, are then shown in the third column, where they can be edited if needed. The only problem I had with this site is that it doesn’t show the number of hits at each stage. So you might select ‘Fijian language’ and then further limit the search, only to retrieve nothing because there are only 3 documents altogether in Fijian.

Other Sites Using Faceted Classification


or go to http://eat.epicurious.com/recipes/

and select
**Faceted Classification**

Written by Glenda Browne
Sunday, 02 September 2007 17:16 - Last Updated Sunday, 04 November 2007 11:28

*Enhanced Search*

Children’s books online: [www.icdlbooks.org/library](http://www.icdlbooks.org/library). This site is quite slow – there is an excellent demo version at [www.icdlbooks.org/library/help/search_area.htm](http://www.icdlbooks.org/library/help/search_area.htm).

Posters: [http://robotwent.com/posters](http://robotwent.com/posters) uses FacetMap (see below).

**Software**

You can implement a facet map on the Web using free software, your own software, or a commercial program (probably needed for large-scale projects).

Commercially available software for faceted search and browsing includes Endeca ([http://www.endeca.com](http://www.endeca.com)), i411 ([http://www.i411.com](http://www.i411.com)) and Siderean Seamark Server (previously codenamed bpallen Teapot; [http://www.siderean.com](http://www.siderean.com)), which uses RDF and Semantic Web concepts. The research done by the University of California, Berkeley, uses software they built using Python, MySQL and the WebWare toolkit ([http://www.sims.berkeley.edu/~hearst/papers/cacm02.pdf](http://www.sims.berkeley.edu/~hearst/papers/cacm02.pdf)).

**Free Software: FacetMap**

FacetMap ([http://facetmap.com/index.jsp](http://facetmap.com/index.jsp)) is a software package that allows you to create and test your own faceted classification scheme on the Web. The site offers ‘3-minute concept info’, which provides information while leading you through a demonstration of FacetMap. As you select a facet from a list, that list is replaced by a list of subfacets.
In the FacetMap demonstration, the first screen allows you to Browse Varietal, Browse Region, and Browse Price. A new feature allows you to set the specific price range you are interested in, rather than just selecting a range from the list.

Figure 1: Initial Browse Screen With Choice to Search by Varietal, Region or Price

If you select ‘French’ in the Region facet, you are shown a new screen in which Browse Region now offers selections of regions within France, and the number of hits shown in the Varietal and Price facets has been adjusted to reflect the more limited area being considered.

Figure 2: Second Browse Screen, After Selecting France (Subcategories Under Region Replaced by Regions of France)

You can continue to drill down within the Region hierarchy, or can move to another facet of relevance. If you select ‘White Wines’ under Browse Varietal, the screen adjusts to show the specific white wines available in France, and changes the numbers again. For example, there are still 4 wines available from Alsace, but the 20 from Champagne have disappeared, probably because they are ‘bubbly’.

Figure 3: Third Browse Screen, After Selecting France and White Wines (Subcategories Under Varietal Replaced by Specific White Wines)

The Web page shows breadcrumbs (the history of the search, in this case ‘>The World>French’ and ‘>Any Varietal>White Wines’) and the top 10 of the hits the search has narrowed down to. At any stage you can limit the search using any facet.

You can create and test your own FacetMap at http://facetmap.com/demosetup/index.jsp.
Start small and test as you go to ensure that problems are ironed out in the early stages. FacetMap files can be created in plaintext (use Notepad, not Word), in FacetMap markup (simple XML; offers a few more options) or in XFML (a language for faceted classification of Web pages.)

**XFML (XML Format for Hierarchical Faceted Metadata)**

XFML ([http://xfml.org](http://xfml.org)) is a subset of the topicmap standard and is optimised for the sharing of faceted metadata between Web sites. XFML will be compatible with topic maps in that any XFML file can be easily transformed into valid XTM. Work is also being done to make XFML compatible with RDF. (See *Online Currents*, April 2002 issue, for an introduction to topic maps).

**Advantages and Disadvantages of Faceted Classification**

Faceted classification is effective because it splits subjects into their component parts and allows retrieval on whichever attributes of the subject are important to the person who is searching. Special features include the combination of hierarchical browsing and searching, and the ability to switch between these two approaches as needed. In sites which show the number of hits for each option, it is clear to the user whether they should refine their search further or go straight to the item itself.

Disadvantages include the lack of a clear presentation of the scope of the site, and lack of special access to popular topics ([http://www.kmconnection.com/DOC100100.htm](http://www.kmconnection.com/DOC100100.htm)). Faceted classification requires a commitment to the creation and maintenance of the classification, and application of metadata to Web pages. It is therefore easiest to implement with structured and tagged data. Automation can be used to extract metadata from databases. Some areas are more easily structured in this way than others. Both the Genome Analyzer and the Dun & Bradstreet site used categories called ‘nonclassifiable’, indicating that it is not always easy to allocate a category for each item being classified or that there has been a delay in tagging those sites.

Many of the sites described above are successful implementations of faceted classifications, using their existing metadata to a fuller extent than is normally possible with search engines, and allowing users more control and better feedback when they search.
Further Information

You can subscribe to the Faceted Classification Discussion (FCD) mailing list at http://groups.yahoo.com/group/faceted_classication. There is more useful information at http://www.iawiki.net/FacetedClassification.

All Web sites were accessed on 5 August 2003.