# Using Faceted Classification To Provide Structure For Information Architecture

Program for Educational Transformation Through Technology
UNIVERSITY OF WASHINGTON

## Challenges

### *Imperfect knowledge representation*

Real human knowledge is, by nature, inconsistent, ill-defined, and unstructured. It is indescribable, impossible to represent faithfully without some loss of meaning. As a result, any method of knowledge representation is flawed. The products of human knowledge lie in documents, and all attempts to describe or organize those documents are flawed as well (Berman, 1984; Wheatley & Armstrong, 1997).

### *Multiple access points*

Digital libraries or web sites often rely on subject hierarchies (such as Yahoo) to arrange documents for browsing. However, many documents are "about" more than one thing. This often leads to either duplication of documents or the limiting of access to one list of subjects. The former leads to wasted time and resources. The latter removes possibilities for meeting users' needs (Wheatley & Armstrong, 1997).

### *Multiple seeking strategies*

Users employ several information seeking and retrieval strategies to find what they need, including browsing, known site, empirical, analytical, and similarity. Unfortunately, many information architectures only allow for a few of these (Fidel & Efthimiadis, 1999; Fidel, et al., 1999).

### *Complexity*

Real knowledge is multi-layered, where meaning is synthesized from several different perspectives. It is infinitely multi-dimensional. The intersection of more perspectives on knowledge leads to higher relevance, but few existing information systems include this functionality (Broughton, 2001; Ellis & Vasconcelos, 2000).

### *Change over time*

Real knowledge is chaotic, disorganized, and constantly changing. Documents change as well, and their meaning changes as society changes. As the shape of the knowledge changes, so must the structure of the systems used to access that knowledge (Pilsk, et al., 2002).

*Knowledge, a rude unprofitable mass,*
*The mere materials with which wisdom builds,*
*Till smoothed and squared and fitted to its place,*
*Does but encumber whom it seems to enrich.*
*Knowledge is proud that he has learned so much;*
*Wisdom is humble that he knows no more.*

*- William Cowper*

## Lessons from Libraries

### *Knowledge Organization*

In libraries, the shape of knowledge is often inferred from the way documents are arranged. The systems used to arrange documents for access are called **classification schemes**. The different classification schemes apply varying philosophies, assumptions and approaches to elucidate a listing of all the concepts addressed by documents in a collection. This list of concepts is then used to both organize and access those documents. On the web, classification schemes have often been replaced by **taxonomic structures**, such as subject hierarchies (eg. Yahoo!, Open Directory) and AI-generated ontologies.

### *Knowledge Representation*

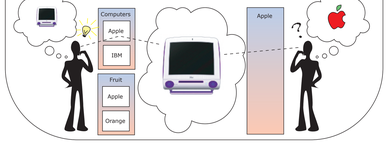
Librarians have traditionally employed a variety of techniques, such as indexing, abstracting, and cataloging, to create **surrogates** of documents. These surrogates are better known these days as **metadata**. Since the meaning of the documents themselves is not always readily apparent from the words contained in them, the surrogates can provide semantic access to those documents and thus aid users in finding what they need.

### *Fixed vs. relative location*

When library collections contained mostly books, the issue of shelf order and location was of far more consequence than in the digital world. Two systems were developed for managing the locations of books. Organizing documents by **fixed location** places documents, often in areas closed to the users, in a single place, affording greater bibliographic control. Arranging a library by **relative location** puts documents in the place where they are accessed, and the physical layout of the collection mimics the structure of the classification scheme. The former, while unrivalled in manageability, hinders user access. The latter allows user access, but requires much maintenance. In the digital world, documents may be placed in both fixed (as in a database) and relative locations (on a web page). With the marriage of relational databases and web scripting seen in most digital libraries and large-scale web sites, the hindrances to using novel classification schemes for document description and organization have been removed.

### *Bibliographic Control & Access*

Bibliographic control refers to the power over how documents are accessed and used. Ideally, the way users access and use information would be congruent with reality and with their own needs and desires. Most information systems fall short of this goal, since each user has their own way of seeing the world. Classification systems that more closely match human knowledge and the user's mental models put more control into the hands of the user.




### *Epistemology & Classification*

There are several ways to go about creating a classification system:

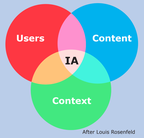
1. **Rational**: use logic and everything you know
2. **Empirical**: use observations from the "real" world
3. **Cultural**: use what everyone else agrees upon
4. **Contextual**: use what works best for each situation

As the fourth approach implies, each approach has different strengths and weaknesses for different contexts. Each of these approaches is linked to a philosophical school of thought, and certain types of classification schemes are more hospitable to certain approaches (Hjørland & Albrechtsen, 1999). The type of classification scheme based on the contextual approach is **faceted classification**.

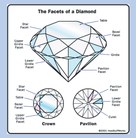
## Methods

### *IA & User-Centered Design*

**User-centered design** is a is a holistic approach to information systems design that takes into account the organizational, personal, and technical systems that surround information. By pairing extensive user analysis with learning outcomes research, we drafted a set of profiles and strategies to guide us in our design. We used these ideas, along with usability testing, in designing a prototype **content management system** that would utilize the power and flexibility of faceted classification to manage, describe, organize, and provide access to documents.




After Louis Rosenfeld




### *Facet Analysis*

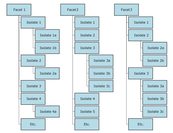
We determined that, given the requirements of the context, content, and users of the systems we were building, it was necessary to use a **faceted classification** scheme to describe, organize, and access documents. Embedded in the process of **facet analysis** is an elucidation of the core concepts and practices involved in using, managing, and sharing information.

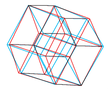
### *Faceted Classification*

S.R. Ranganathan pioneered a method of classifying documents using a **multi-dimensional** notation called Colon Classification. By organizing documents according to several subjects simultaneously, the shape of knowledge can more faithfully be rendered in the shape of the collection.

**Facet analysis** is the method by which a faceted classification system is constructed. It involves

1. Study the context, the content, and the users
2. Identify the major types of concepts (**facets**) covered by the above
3. Use a taxonomic structure to fill each facet with concepts (**isolates**)
4. Categorize each document according to each facet



A tesseract: a 4-dimensional hypercube

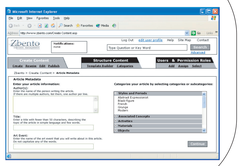
Ranganathan's faceted scheme is 5-dimensional:

1. Who: Personality
2. What: Matter
3. How: Energy
4. Where: Space
5. When: Time

### *Content Management*

Our prototype content management system, ZBento, utilizes templates for standardized **creation** and **categorization** of original content. The system also allows for the **collection** and **description** of existing documents.

All this is integrated with a faceted classification system, which provides both cataloging and retrieval interfaces.



### *From Facets to Wireframes*

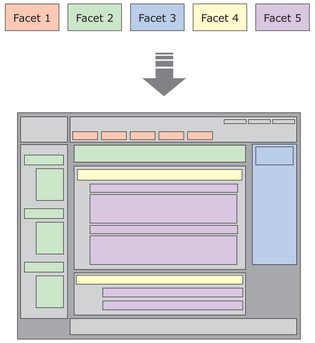
Mapping the structure of the faceted classification scheme into the navigational elements of a web site involves the following steps:

1. Establish an order for the facets.

2. Connect each facet to a type of navigation.
   - ☐ a. Facet 1 -> Global navigation
   - ☐ b. Facet 2 -> Local navigation
   - ☐ c. Facet 3 -> Associational navigation
   - ☐ d. Facet 4 -> Clustering
   - ☐ e. Facet 5 -> Sorting

3. Fill that navigational element with the isolates of the facet.

4. The documents displayed on each page will be classified at the intersection of each facet above the current context.

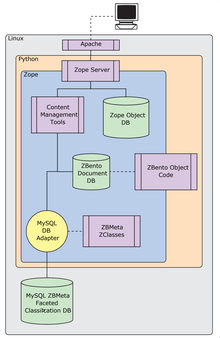



### *System Architecture*

We used a combination of open source tools for low-cost, rapid prototyping of the ZBento system.

The facets, isolates and the classification for all documents resides in the MySQL database, which runs externally to Zope. The objects within Zope and Python contain methods that provide an interface between the Zope content management tools and the data.

The documents are stored in the ZBento Document DB.




## Implementation

### *Context & Content*

We designed the prototype for the Legal Gateway project in conjunction with the Washington State Bar Association's Council for Public Legal Education.

The content will be culled from a variety of partner web sites and organizations, and original content will be generated by legal experts around the state.

In this context, the facet analysis resulted in the following facets:

1. **Purpose**: how document will be used
2. **Topic**: "aboutness", subject addressed
3. **Type**: legal format of document
4. **Question**: concern addressed

### *Technology*

The prototype was built with scalability in mind. To accommodate the high degree of instability inherent in the documents the system was to contain, the classification system was stored in a relational database to allow for dynamic querying and full-text indexing.

## What's next

### *Thesaurus & Query Processing*

We have developed tools to demonstrate how the faceted classification system can be used as a thesaurus, a glossary, and a query processor. Such functionality would meet the requirements of a variety of information seeking strategies.

### Sub-Facets

The Topic facet for the Legal Gateway project is somewhat vague and contains much duplication. Thus, in the future, the "Topic" facet may be broken into 3 sub-facets:

a. **Problem**: legal issue addressed
b. **Strategy**: solution proposed
c. **Context**: area of society discussed

### *Other Contexts*

We hope to apply these same principles in other areas of study, including Architecture and Orthopaedics. Since the users, content, and context will be different for each project, the facets and processes involved are expected to be quite different. We will also be producing a design specification for an enterprise-wide content management system based on our prototype.

**More about us:**

- http://lawforwa.org (ZBento development site)
- http://depts.washington.edu/pettt (About PETTT)

**Interesting Links:**

- http://www.zope.org (about Zope)
- http://flamenco.sims.berkeley.edu (similar project)
- http://epicurious.com (faceted commercial site)

**Aaron J. Louie, William Washington, & Eric L. Maddox**

## References:


Berman, Sanford, ed. 1984. Subject Cataloging: Critiques and Innovations. NY: Haworth Press.

Broughton, V. (2001). Faceted classification as a basis for knowledge organization in a digital environment: the Bliss Bibliographic Classification and the creation of multi-dimensional knowledge structures. *New Review of Hypermedia and Multimedia*, 7, 67-102.

Ellis, D.; Vasconcelos, A. (2000). The relevance of facet analysis for World Wide Web subject organization and searching. *Journal of Internet cataloging*, 2:3-4, 97-114.

Fidel, R.; Davies, R.K.; Douglass, M.H.; Holder, J.K.; Hopkins, C.J.; Kushner, E.J.; Miyagishima, B.K.; Toney, C.D. (1999). A visit to the information mall: web searching behaviour of high school students. *Journal of the American Society for Information Science*, 50:1, 24-37.

Fidel, R.; Efthimiadis, E.N. (1999). Web searching behavior of aerospace engineers. *Proceedings of SIGIR '99*, 319-20.

Hjørland, B., and Albrechtsen, H. (1999). An analysis of some trends in classification research. *Knowledge Organization*, 26:3, 131-39.

Pilsk, S.; McIntyre-Colby, S.; Andrew, P.G.; Wilson, A. (2002). Organizing corporate knowledge: the ever-changing role of cataloging. *Information Outlook*, 6:4, 30-4, 37-40.

Wheatley, A.; Armstrong, C.J. (1997). Metadata, recall, and abstracts: can abstracts ever be reliable indicators of document value? *ASLIB Proceedings*, 49:8, 206-13.