A multi-faceted view: Use of facet analysis in the practice of website organization and access.

Abstract: In 2001, information architects and knowledge management specialists charged with designing websites and access to corporate knowledge bases seemingly re-discovered a legacy form of information organization and access: faceted analytico-synthetic theory (FAST). Instrumental in creating new and different ways for people to engage with the digital content of the Web, the members of this group have clearly recognized that faceted approaches have the potential to improve access to information on the web. Some of these practitioners explicitly use the forms and language of FAST, while others seem to mimic the forms implicitly (Adkisson, 2003). The focus of this ongoing research study is two-fold. First, access and organizational structures in a stratified random sample of 200 DMOZ websites were examined for evidence of the use of FAST. Second, in the context of unstructured interviews, the understanding and use of FAST among a group of eighteen practitioners is uncovered. This is a preliminary report of the website component capture and interview phases of this research study. Future work will involve formalizing a set of feature guidelines drawn from the initial phases of this research study. Preliminary observations will be drawn from the first phase of this study.

In any sphere of life, practice precedes theory. Lifeforce [sic] stimulates man to improvise, to design and to develop various aids both at the physical and at the mental levels…After a long experience is gained with an improvised aid, a theory is developed in order to understand the aid deeply and to systematize, improve, refine, and develop it. So has it been with classification too. … (Ranganathan, 1971).

I. Overview

The global world of information is characterized by fluctuating borders that present challenges which impel continual, creative and proactive adaptations in strategies and solutions. Working with the postulate that information organization and access is best supported by mutable systems, the current research study examines reflections of the fluid associative and formal structured interrelationships among information packages that are best accessed through dynamic, responsive systems first described by Bliss, 1929, 1933; Bush, 1945; Otlet, 1934; Ranganathan, 1957, 1967. While grounded in the sociological and cultural milieu of the present information realities and current international practices of web design, this study is also anchored in the intellectual and theoretical foundations of knowledge organization.

In 2001, a group of information architects and knowledge management specialists charged with designing websites and access to corporate knowledge bases seemingly re-discovered a legacy form of information organization and access: faceted analytico-synthetic theory (FAST). This group of users has been instrumental in creating new and different ways for people to engage with the digital content of the Web and they have clearly recognized that faceted approaches have the potential to improve access to information on the web. Some of these practitioners explicitly use the forms and language of FAST, while others seem to mimic the forms implicitly (Adkisson, 2003).

It must not be forgotten that the historical growth and development of FAST would not have been possible without the work of a number of seminal thinkers throughout the world, among them Paul Otlet (Belgium, Universal Decimal Classification), Henry Evelyn Bliss (United States, Bliss Classification) as well as S. R. Ranganathan (India, Colon Classification). The fact that FAST retained its currency and continued to develop
throughout the years is primarily due to the efforts of three groups who invested heavily in this new paradigm of information organization and access after WWII. These three international groups were: the Library Research Circle (LRC) in India, the Classification Research Group (CRG), in England and the Classification Research Study Group (CRSG) in North America.

II. Research study

Current applications utilizing faceted approaches are varied and include website search and browsing systems. By utilizing the methods of content analysis of websites and semi-structured interviews with Information Architects and Knowledge Management specialists, this study will attempt to uncover whether or not evidence of faceted approaches in website information access and organization structures exists. In so doing, it is hoped that connections between the intellectual and theoretical foundations of FAST in information organization and current practices in information architecture will be made explicit.

The focus of this research study was two-fold, capturing access and organizational structures used in websites, and examining whether or not there is evidence of the use of FAST in website construction and the design of website search tools. Using the Alexander and Tate (A & T) (1999) typology of websites, a stratified random sample of 200 websites were selected from four DMOZ categories of interest, Shopping, Business and Reference and Society. These categories were chosen for their marked differences in content and rationale for creation and existence in order to provide the possibility for marked differences in practice. In addition, eighteen practitioners who explicitly invoke both the language and the forms of FAST were interviewed using an open, but guided (semi-structured) schedule of questions and prompts in order to elicit current knowledge as well as to determine the ways in which their practices conform to or depart from traditional applications of facet analysis.

II. a. The sample websites

Alexander and Tate (A & T) (1999) created a five part typology of websites in order to assist in assessment of website content and quality. This typology was used to guide the choice of sample websites. This typology was chosen because it presents a conceptual understanding of the web by breaking a large universe of websites into 5 rough types. This study will compare Alexander and Tate Informational type websites with Business/Marketing type websites in an attempt to repeat and assess Adkisson’s (2003) preliminary study of the use of faceted classification on e-commerce websites. These two types of sites have been selected as being conceptually related, one providing access to information, the other to goods; but are driven by different motives, one being primarily service-oriented and the other profit-oriented. Sites were then selected from those in the Open Directory Project: DMOZ: http://dmoz.org which gives access to a cross-section of the web chosen by volunteer editors under strict editorial guidelines. It is hoped that due to these policies, that the randomly drawn sample of 200 DMOZ websites has resulted in a collection of websites that may well represent the best, most reliable, most frequently consulted and carefully constructed sites on the Web. The DMOZ categories further provide an operationalized way of considering the Alexander and Tate types, as many of the DMOZ categories map to them, thus providing a way to draw the Alexander and Tate types down to the level of the individual page.
Site sample (n=200) by A&T type [IN or BM] and DMOZ category

Figure 1: Outlines the composition of the 200 sample websites by DMOZ category: Reference (REF), Society (SOC), Shopping (SHOP), and Business (BUS) and by Alexander and Tate (A&T) type – Informational (IN) or Business/Marketing (BM).

Rosenfeld and Morville (2002, p. 4) provided guidance in the selection of the components of interest for this study in their description of the four major components of a website: the system of organization, the search system, the navigation system and the labeling system. This study examines two of these major components, the website search and navigation systems. Search systems on websites are easy to identify and are well known to web users who typically look for “the box” when beginning a search. The current research uses the following definition of navigation elements: “any elements that help a user locate information on a website and allow the user to easily move from page to page within the site. Navigation aids may be text, graphics or a combination of these” (Alexander and Tate, 1999, p. 139). Though Rosenfeld and Morville (2002, p. 170) distinguish between search and navigation systems on websites, they urge designers to integrate these components in ways that allow users to jump easily between the two. In the language of the practitioners who explicitly invoke FAST, faceted navigation and faceted searching are central to the applications they create. “Instead of forcing one way to view the items, Faceted Navigation allows users to view the items in any way they want. At the same time they learn how the items are structured so that they may consider other search strategies in the future” (van Welie, 2004). Thus browsing or ‘integrated’ search and browse systems as described by Rosenfeld and Morville (p. 131) were also examined. Integrated systems allow users to search and browse in one central location and often integrate both actions into one interface seamlessly for users.
Table 1: A ranked list of website components across the 200 sample websites.

<table>
<thead>
<tr>
<th>Search Features:</th>
<th>(REF) 28 Reference</th>
<th>(SOC) 50 Social</th>
<th>(SHOP) 57 Shopping</th>
<th>(BUS) 65 Business</th>
</tr>
</thead>
<tbody>
<tr>
<td>basic search</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>advanced search</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Browse Features:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>browse</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>integrated search/browse</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Navigation Features</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>navigation only</td>
<td>2</td>
<td>7</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>embedded navigation</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>sitemap</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

It is clear from this table that the most common feature across all types of sites is the basic search mechanism. It is somewhat surprising that the sitemap, a navigational feature is far less common than might be expected. Another highly ranked feature is the advanced search mechanism, with the exception of Business, which seems to prefer the use of a navigational sitemap. Also surprising is the heavy secondary reliance among reference sites on the use of navigational features rather than an advanced search feature to assist visitors in information seeking. One feature, integrated search and browse, is remarkable, as much for the complexity of execution as for the fact that it seems to serve as an advance indicator for a site in which there is a high likelihood for the use of FAST. Sites with this feature tend to have a set of co-occurring features, such as embedded navigation, typically a drop down menu attached to a global or site level navigation bar, which provides access to sub site, or local areas. It would seem that this integration of search and navigational features provides a marker of technological sophistication. Both features require the use of advanced information architecture in a way that sitemaps, or the simple use of buttons to help a site visitor navigate through a site do not. Perhaps this accounts for the fact that a site which uses these kinds of advanced features is also one which is more likely to use FAST. Might an IA who designs such a site be more skilled and aware of the discussions about FAST within the community?

II. b. The practitioners

Participants were recruited by posting an invitation to participate on several listservs with IA, KM or information organization topics among them, SIGIA and the Faceted Classification discussion list. Others who explicitly invoke both the language and the forms of FAST whether on blogs, lists, or at conferences in the past few years were also invited to participate. Respondents were interviewed using an open, but guided (semi-structured) schedule of questions and prompts. It was hoped that this method would provide access to a broad reach of practitioners. The table below describes the sample members.
Seven of the interviewees are members of the Information Architect / User Experience community (IA/UX). UX is a much broader conception of IA. Four are members of the Knowledge Management (KM) community. Seven are involved in software design, engineering or product management. Two are also members of the academic community, and the rest are either self-employed consultants or work at for-profit corporations. Within the sample is a good mix of consultants and managers. The two academic members are also software engineers who spoke mainly about their creations and the practice of working within a programming and applications framework, and not primarily about their research orientations.

One thrust of the interview process was to discover whether there was a common knowledge base, both in terms of people and resources that were commonly consulted and this is covered in the following tables.

<table>
<thead>
<tr>
<th>Person consulted</th>
<th>Affiliation / creation</th>
<th>Times named</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hearst, Marti</td>
<td>Associate Professor, SIMS, U. C. Berkeley/ Flamenco search interface, various publications</td>
<td>8</td>
</tr>
<tr>
<td>no one</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Ranganathan, S.R.</td>
<td>Professor University of Madras and Hindu University/ Prolegomena (1967)</td>
<td>4</td>
</tr>
<tr>
<td>Schwartz, Candy</td>
<td>Professor, Simmons College</td>
<td>2</td>
</tr>
<tr>
<td>Svenonius, Elaine</td>
<td>Professor Emerita, Department of Information Studies UCLA/ Intellectual Foundation of Information Organization. (2000)</td>
<td>3</td>
</tr>
<tr>
<td>Pollitt, Steve</td>
<td>View-Based Systems, Ltd.</td>
<td>2</td>
</tr>
<tr>
<td>Rosenfeld, Lou</td>
<td>Lou Rosenfeld, LLC/ Information Architecture for the World Wide Web. 2nd ed. (2002)</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3: Most likely to be consulted for information about facet analysis or faceted classification. A total of 24 names were listed, 9 were academics. 2 are no longer living.

The list of people cited was relatively co-extensive with the list of sources. Other resources of note were the Faceted Classification discussion list, Peterme, a blog maintained by Peter Merholz, and Boxes and Arrows, an online publication with topics of interest to interaction designers, information architects and graphic designers.
III. Practice and understanding

I’m fond of saying half the people in the world think they invented faceted classification. I know I did for awhile.

Some form of this statement was uttered by a third of the interviewees. Though there are members of the group who have studied canonical FAST literature, most found out about it when someone else identified their practice, software or design as an instantiation of faceted classification. Many of the members of the group were directed to, or found out about Marti Hearst’s work with the Flamenco browser as their first exposure to faceted classification. Others read *Information Architecture for the World Wide Web*, which contained a discussion of faceted classification in relation to information architecture. The often cited examples of faceted classification heavily favor applications, with a total of 15 websites cited. All websites and all but one of the six applications cited were created by for-profit organizations. Marti Hearst’s Flamenco application is the notable exception, and the most highly cited exemplar.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Example Description</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Flamenco demo search application [Academic]</td>
<td><a href="http://bailando.sims.berkeley.edu/flamenco.html">http://bailando.sims.berkeley.edu/flamenco.html</a></td>
</tr>
<tr>
<td>4</td>
<td>Endeca (application)</td>
<td><a href="http://endeca.com">http://endeca.com</a></td>
</tr>
<tr>
<td>3</td>
<td>Epicurious (website)</td>
<td><a href="http://www.epicurious.com/">http://www.epicurious.com/</a></td>
</tr>
<tr>
<td>2</td>
<td>Barnes and Noble (Endeca site)</td>
<td><a href="http://www.barnesandNoble.com">http://www.barnesandNoble.com</a></td>
</tr>
<tr>
<td>2</td>
<td>Siderean (application)</td>
<td><a href="http://www.siderean.com">http://www.siderean.com</a></td>
</tr>
<tr>
<td>2</td>
<td>Facetmap (application)</td>
<td><a href="http://facetmap.com/">http://facetmap.com/</a></td>
</tr>
</tbody>
</table>

Table 4: Most often cited examples of faceted classification in action.

But what do the interviewees mean when they talk about faceted classification? Overall, the interviewees demonstrated a sophisticated understanding about fundamental concepts such as facet. Most indicated that a facet is a dimension, attribute, characteristic, category or property. Many indicated an appreciation for the associative or hierarchical constructions that FAST makes possible and explicit. Within this community, the most common understanding of a faceted classification is a system with the following components: facets are displayed as part of a search or navigation system and are visible at all times, the results are displayed using facets as an organizational device, and a null or empty set is never returned as a result display. Many indicated that it was difficult to think of faceted classification outside of a navigational or search system. As one interviewee noted, “One of the things we are still trying to understand here is how to move rapidly to take faceted interfaces and begin to use them not just for retrieval purposes, but for managing an entire class of information.” The observation of one interviewee speaks for many, “I think we use faceted classification, and everybody understands it more or less, but nobody has really formulated it for us in a way we can understand. The practice we have needs to be theorized a bit and formalized.”

As for usability, some expressed an understanding of the importance of domain knowledge when approaching facet analysis. “Do a lot of user research to determine what kinds of content people are looking for and the ways in which they search.” This is important because, “Facets and their display are completely dependent on the target population, their needs and interests.” While there is an understanding that available technology should make faceted interfaces easy to implement, “to many people it is still just an interface”, and not yet “a complete data model” though this is something clearly
desirable for many. The interface design issues seem to be among the most exciting, and
difficult challenge for many in the sample and is often addressed as “a wicked interface
issue”. Questions range from how complex the a faceted interface can be before it
overwhelms, how to relate attributes to a subset of entities, and whether or not to allow
multiple picks within a facet. Almost all of the interviewees expressed a strong interest in
usability testing of faceted search and navigation interfaces to provide support for their own
anecdotal evidence that these interfaces are powerful, useful and intuitive.

As to the relation of practice to FAST theory, there is no agreement as to the necessity of
one of the fundamental canons, that of mutual exclusivity and what that means in practice.
While some interviewees feel that orthogonal facets are unnecessary in today’s digital
environments, others adhere to it rigorously. It is interesting to note that Travis Wilson will
be presenting his Facetmap software, at the March 2006 Information Architecture Summit.
Designed according to a “strict faceted classification model”, Facetmap forbids a
combination of attributes at any given facet level (in his example: ice cream with a flavor of
raspberry-chocolate is not permitted according to the strict faceted classification model). He
goes on to state, “This is counterintuitive, controversial, and if you subscribe to S.R.
Ranganathan's original facet theory, heretical. Yet a faceted classification is most effective
when built upon that restriction.”

Another disputed boundary area is the desirability or efficacy of universality in facet
creation. “I can’t imagine the classical facets being useful in identifying your own facets
and then applying them to your own system. That’s just not the way things are.” Another
interviewee said, “I think it is just silly to have a universal way of organizing the world.”
Yet other interviewees noted a common observation, that of the frequent synchronous
occurrence of fundamental categories in other applications and implementations which
closely mirror Ranganathan’s five fundamental categories – whether these facets were
automatically extracted from data in some way, or created by human indexers. Other
perceived limitations of faceted structures relate to the difficulty of choosing or selecting
facets when the data that is being organized is unstructured or lacks metadata or occurs
outside a well-defined domain. Clearly, this is a terrain in which much experimentation with
and discussion of FAST (whether fully understood or not) is occurring. It is interesting to
note that many of these same discussions occurred in earlier work completed by the CRG,
CRSG and LRC. The solutions this group will eventually reach may be a fruitful source of
theory-building.

III. Future research direction

Faceted classification is one way of giving the IA community the tools they need to effectively shape,
craft, and channel information repositories in ways that enable designers to get the benefit of human
insight and which can be immediately useful to people in the broader community.

A preliminary assessment of the sample website components indicates that
approximately ten percent show evidence of some degree of facet analysis. More rigorous
analysis of these sites is being conducted. Facets that are evident in these website designs
are always used in interfaces which attempt to combine or integrate browsing and searching
in order to enable site visitors to find what they are seeking. This remains true regardless of
site intent or category. The continuing interest among KM and IA practitioners in FAST
endures along with a sense that though the ideas of faceted classification may well be
embedded in practice – more guidance and practical theoretical help is needed in order to
create the kinds of compact scalable systems notable for their conceptual clarity that these
communities know are possible with FAST. It appears that robust application of FAST is
still elusive. Guidelines are currently being drawn from content analysis of the sample websites and the observations elicited in the interviews of designers. The principles and postulates of FAST will be combined with the observations drawn from practice in the final phase of this study in order to create a checklist that may enable web designers interested in applying FAST as part of their own website design and construction practices to do so in a consistent and robust manner. One potential contribution of the proposed study is the possibility to apply what is uncovered in the course of this study especially as relates to novel web design practices in order to enrich the current state of FAST theory, and in turn for FAST theory to enhance the design and construction of websites.

References