

Effective Display of Browsable Classification on the WWW and other Hypertext Media

James D. Anderson

School of Communication, Information, and Library Studies
Rutgers The State University of New Jersey, New Brunswick
jda@scils.rutgers.edu

"In the networked world, browsing has supplanted direct searching as the primary means to locate information" — *Christine Borgman (2001)*.

"Librarians invented a better kind of classification decades ago, that is called faceted classification ... so why are hierarchical classifications of information content still being used?" — *Marcia Bates (2002) in "After the Dot-Bomb: Getting Web Information Retrieval Right This Time."*

1. INTRODUCTION

The problem I would like to address is the effective display of browsable classification arrays for information discovery in hypertext media such as the World-Wide Web (WWW). For the first time, we have the possibility to move beyond the traditional display of classification as a one-dimensional hierarchical sequence of classes, each class consisting of one or more related concepts combined according to a consistent, or idiosyncratic, citation order of facets. In this new context, it would seem that faceted classification would be ideal. Facets can be displayed separately, giving users the option of choosing the facets that interest them (Anderson 1990). No longer would users be forced to accept the citation order of facets determined in advance by a classificationist.

But facet displays are mostly absent from the WWW. I force my students to use facet displays in designs of information retrieval databases, but they resist, preferring the more traditional display of multi-facet relationships. Why is this? Is it because facet displays simply haven't been used enough to demonstrate their power? Are they too new and unfamiliar? Or is the reason more basic. Do users not "think" in terms of facets, but rather in terms of relationships among concepts from a variety of facets? If the latter is the case, are users more comfortable with classification displays of relationship clusters, rather than the display of separate facets?

2. BROWSABLE CLASSIFICATION

Before the advent of the WWW and other hypertext media, classified displays played a minor role in browsing for information resource discovery.

Shallow enumerative hierarchical classifications were used for browsing abstracts in printed indexing and abstracting services, but these classifications rarely extended beyond three or four levels.

In libraries, especially in North America, users typically consulted an alphabetical catalog, found the address (call number) of the item that interested them, and then searched the shelves for that or similar items. The role of classification was merely to place similar or related items next to, or near to, the sought item on the shelves, so that users could browse the shelves once they located a particular classification notation. However, the meaning of classification notation was rarely presented at the shelves, other than the broadest of class names.

Users of libraries rarely began their explorations by browsing classification. It was just too difficult to display detailed and complex classification schemes in a linear, fixed, static medium. Take the most widely used library classification, for example. The Dewey Decimal Classification (DDC) provides browsable summaries for the 10 main classes, for the 100 primary subdivisions, and for the 1000 sub-subdivisions, but after that, the user of the printed classification schedules is thrust into all the detail of very narrow and specialized classes, most of which do not interest them. It is very difficult to browse through this detail to find particular classes of interest.

Hypertext has changed all that. Hypertext is the ideal medium for the display of classification because users can select the particular path that interests them at every level.

3. ROLE OF FACETS IN CLASSIFICATION DISPLAY

In the traditional one-dimensional, linear display of classification, it made little difference whether the underlying classification scheme was a faceted classification, an analytico-synthetic classification such as the DDC, or a more enumerative classification such as the Library of Congress Classification (LCC). In this context, the only real difference among these styles of classification is that the citation order of facets in complex classes tends to be more regular and consistent in faceted classifications than in analytico-synthetic or enumerative classifications.

With the advent of hypertext media, such as the WWW, faceted classifications no longer have to be displayed as if they were enumerative classifications. The facets can be displayed separately, so that users can choose and combine facets as they wish. They do NOT have this option in the traditional one-dimensional display of even faceted classification. A. Steven Pollitt has been one of the principal investigators and proponents of facet displays rather than the one-dimensional display of precoordinated concept clusters (Pollitt 1998).

4. AN EXAMPLE: THE MLA CLASSIFICATION

An example may help to clarify the options that face us in the hypertext environment. The Modern Language Association has a fully faceted classification which it uses for both indexing and classification (Anderson 1979, 1980). Yet, of necessity, in its printed version, the MLA International Bibliography must display this faceted classification with a predetermined citation order of facets, so that the end result is quite similar to the one-dimensional displays based on the analytico-synthetic DDC or the enumerative LCC.

Here are the MLA facets:

- Specific literatures: e.g., English literature, American literature, Chicano literature, Puerto Rican literature.
- Performance media: e.g., theater, story-telling, recitation.
- Languages (if different from language of national literature): e.g., English, Spanish, Swahili.
- Periods: e.g., 20th century, 19th century.
- Individuals (real): e.g., Thomas Hardy, Emily Dickinson, Abraham Lincoln, James Baldwin.
- Anonymous works: e.g., Beowulf, Mother Goose, The Bible.
- Groups/movements: e.g., Avant Garde, Beat Generation, hippies, lesbian poets, African American writers, children.
- Genres: e.g., poetry, drama, non-fiction novels.
- Works: e.g., The wind in the willows, Alice in wonderland, Giovanni's room.
- Features: e.g., dialogue, poetic realism.
- Literary techniques: e.g., visual metaphor, imagery, symbolism.
- Themes/motifs/figures/characters: e.g., [treatment of] love, hate, war, Manifest Destiny, salvation, Huck Finn, Cinderella.
- Influences (recipients): e.g., [influence on] Harlem Renaissance, Generation of 1898.
- Sources: e.g., [influence of] Harlem Renaissance, Generation of 1898, The Bible.
- Processes: e.g., characterization, translation.
- Types of scholarship: e.g., criticism.
- Methodological approaches: e.g., sociological approach, psychological approach, Marxist approach.
- Theories: e.g., Freudian theory, evolution (as theory).
- Devices/tools: e.g., computers, concordances.

In the printed MLA Bibliography, the display of this classification looks something like this (headings excerpted from the classified section of the *MLA international bibliography* 1997, 1999):

```
British and Irish literatures
  English literature
    [periods]
    1800-1899
    1900-1999
      [persons]
      Ford, Ford Madox (1873-1939)
      Forster, E. M. (1879-1970)
        [genres]
        Novel
          [works]
          Howards End (1910)
          A Passage to India ((1924)
            [citation and notes]

  Irish literature
  Scottish literature
  Welsh literature

Breton literature

Commonwealth literature
  Australian literature
  Canadian literature
  New Zealand literature

English Caribbean literature

American literature

European literature
  Central European literature
    French literature
      [periods]
      1800-1899
      1900-1999
        [persons]
        Gide, André (1869-1951)
          [genres]
          Novel
            [works]
            L'Immoraliste (1902)
              [citation and notes]

  German literature
  Netherlandic literature
  Czech literature
  Slovak literature
  Hungarian literature
  Polish literature
```

Northern European literature
Southern European literature
Latin American literature

And here is a hypothetical display of facets laid out so users can select the particular facets that interest them:

[Help] [Keyword search] [Browsable subject index]
[Author, Title, Journal indexes] [Background]

Welcome to the Modern Language Association of America's

MLA INTERNATIONAL BIBLIOGRAPHY

Access to 1,200,000 records for critical scholarship in journal articles, monographs, dissertations, and proceedings, 1981-1999. Choose Keyword search, Browsable subject index, Author, Title, Journal indexes, or select more detailed displays of up to three of the primary topics below:

Literature - or select: specific literatures, performance media, languages, periods, individuals, groups/movements, genres, works, literary techniques, themes/motifs

Folklore - or select: types, performance media, genres, places, periods, groups/movements, literary techniques, themes/motifs

Language / linguistics - or select: languages, places, periods, linguistic aspects, groups/movements

And/Or Select general aspects: special features, influences (recipients & sources), processes, types of scholarship, theories, devices/tools, disciplines, scholars, and special types of documents (e. g. , bibliography)

5. AN EXAMPLE FROM A STUDENT DESIGN PROJECT

As noted previously, I force my students to design IR databases with faceted structures, so that users can get an overview of the entire conceptual content of the IR database from an initial display of the primary facets. Users may then choose two or three facets of primary interest, and create their own browsable display based on these facets. I find it difficult to get my students to think in facets. Associations among concepts that cut across facets, that is, clusters of relationships, seem to be much more natural to them.

I point out to my students that the number of potential relationship clusters is practically unlimited, making it difficult to use them for an overview of the content of an IR database or digital library (or physical library, for that matter!). But the number of key facets is generally much smaller, often fewer than 20. By displaying the facets of a domain, one can get a good overview of the whole. Then, in accordance with Ben Shneiderman's (1998) mantra ("Overview first, zoom and filter, then details on demand"), users can pick the facets of greatest interest and browse them in greater detail. By selecting the classification headings of greatest interest in one, two or three facets simultaneously, they can be taken to the details about matching documents provided in a full surrogate record, or even full text.

Thus my students provide a list of facets on their opening screens, giving users the option of picking up to three facets to browse simultaneously. Here is an example from one recent student project by Melissa Hoffman (text only, without the graphics and typographic features used in the actual design; the number of postings in each category is included):

Browse Subject Categories:

Items in "Lesbian Studies" are indexed under ten subject categories. Browse them one at a time by clicking the link to find terms you can add to a search, or browse up to three categories simultaneously by checking the boxes and clicking Go!

☐ People (18,532) e.g. Sappho, Anne Lister, Gertrude Stein, Eleanor Roosevelt...

☐ Time Periods (15,027) e.g., 1970s, Renaissance, 19th century...

☐ Nationality/Ethnic Groups (14,700) e.g. American, English, Asian, Chinese, Latinas...

☐ Attributes (9,043) e.g. anger, fear, hate, infidelity, love, monogamy, pride, promiscuity...

☐ Actions (19,500) e.g. coming out, cross-dressing, dating, voting...

☐ Groups/Organizations (16,500) e.g. activists, ACT UP, bisexuals, Daughters of Bilitis, mothers, poets...

☐ Events/Movements/Symbols (12,568) e.g. Stonewall, sex wars, labrys, Clinton inauguration...

☐ Places (14,523) e.g. Asia, bars, China, prison...

☐ Disciplines/Theories (17,932) e.g. queer theory, literary criticism, feminism...

☐ Literature/Works (9706) e.g. biography, criticism, "Written on the Body"...

6. SO WHICH IS THE MORE EFFECTIVE DISPLAY STRATEGY?

So, even though I am convinced personally that facet displays are the way to go in the hypertext world of the world-wide web, the evidence of actual use seems to counter this.

It seems to me that this question of effective classification display is a ripe issue for research — usability testing of these two fundamentally different approaches (facet displays versus hierarchical displays).

Does the scope of the classification make a significant difference? Are classifications dealing with a single discipline or domain more amenable to facet displays, compared to classifications covering broader areas?

Here's my proposal: Take an existing, widely used classification such as DDC or Yahoo's directory and recast it as a facet display, so that users have a choice — they can browse facets, or they can browse hierarchical enumerative relationship clusters. I fear they will choose the latter, in part because that's what they are used to. But why? Just because that's been the customary approach for so long? Or is it closer to the way humans naturally organize information — not by facets but in clusters of relationships that cut across facets?

Let us conduct similar studies in narrower domains as well.

Studies such as these may help us be better able to recommend displays based on facets, or the more traditional approach of enumerated hierarchical relationship clusters.

REFERENCES

Note: Reducing author forenames to initials, as commanded by APA, violates all LIS standards for the identification and naming of persons. LIS standards call for using the name that is best known, usually the name used by the person in her/his publications. Why does SIG/CR advocate the violation of basic LIS standards? APA is NOT an IR or LIS standard — it is a PUBLISHER's standard designed to save the maximum amount of space in print-on-paper publications. "Anderson, J.D." confuses me with hundreds of unrelated persons with the same family name and initials! It seems to me that SIG/CR should have a higher regard for LIS and IR standards! — JDA.

- Anderson, James D. (1979). Contextual indexing and faceted classification for databases in the humanities. In *Proceedings of the 42nd annual meeting of the American Society for Information Science*, volume 16; 1979 October 14-18; Minneapolis, MN (pp. 194-201). White Plains, NY: Knowledge Industry Publications.
- Anderson, James D. (1980). Prototype designs for subject access to the Modern Language Association's bibliographic database. In Joseph Raben & Gregory Marks (Eds.), *Databases in the humanities and social sciences: proceedings of the IFIP Working Conference*; 1979 August 23-24; Dartmouth College, Hanover, NH (pp. 291-295). Amsterdam; New York: North-Holland.
- Anderson, James D. (1990). Ad hoc, user-determined classified displays based on faceted indexing. In Susanne M. Humphrey & Barbara H. Kwasnik (Eds.), *Advances in classification research* (pp. 1-7). Medford, NJ: Learned Information, for the American Society for Information Science.
- Bates, Marcia (2002). After the dot-bomb: getting web information retrieval right this time. *First Monday*, 7(7), http://firstmonday.org/issues/issue7_7/bates/index.html, accessed 7/30/2002.
- Borgman, Christine (2001). Quote on the book jacket of Rice, Ronald E.; McCreadie, Maureen; & Chang, Shan-ju L. *Accessing and browsing information and communication*. Cambridge, MA: MIT Press.

- Modern Language Association of America (1999). *MLA international bibliography of books and articles on the modern languages and literatures 1997*. New York: MLA.
- Pollitt, A. Steven (1998). The key role of classification and indexing in view-based searching. *International cataloguing and bibliographic control*, 27(2), 37-40.
- Shneiderman, Ben (1998). *Designing the user interface: strategies for effective human-computer interaction*. 3rd ed. Reading, MA: Addison Wesley Longman.