

Syracuse University

SURFACE

School of Information Studies - Faculty
Scholarship

School of Information Studies (iSchool)

7-13-2002

Commercial Websites and the Use of Classification Schemes: The Case of Amazon.com. In Lopez-Huertas, Maria J. Challenges in Knowledge Representation an Organization for the 21st Century: Intergration of Knowledge Across Boundaries.

Barbara H. Kwasnik
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/istpub>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Kwasnik, Barbara H., "Commercial Websites and the Use of Classification Schemes: The Case of Amazon.com. In Lopez-Huertas, Maria J. Challenges in Knowledge Representation an Organization for the 21st Century: Intergration of Knowledge Across Boundaries." (2002). *School of Information Studies - Faculty Scholarship*. 144.

<https://surface.syr.edu/istpub/144>

This Article is brought to you for free and open access by the School of Information Studies (iSchool) at SURFACE. It has been accepted for inclusion in School of Information Studies - Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

Barbara H. Kwasnik

School of Information Studies, Syracuse University, Syracuse, NY, USA

Commercial Websites and the Use of Classification Schemes: The Case of Amazon.com

Abstract: The structure and use of the classification for books on the amazon.com website are described and analyzed. The contents of this very large website are changing constantly and the access mechanisms have the main purpose of enabling searchers to find books for purchase. This includes finding books the searcher knows about at the start of the search, as well as those that might present themselves in the course of searching and that are related in some way. Underlying the many access paths to books is a classification scheme comprising a rich network of terms in an enumerative and multihierarchical structure.

1. Introduction.

In a previous paper (Kwanik & Liu, 2000) we reported on a classification scheme used on the commercial auction website, eBay.com. The aim was to explore the nature and use of classifications in web environments where:

- the contents of the sites are in constant flux;
- the user population is unknown, or if it is known, we can assume it is diverse;
- it is desirable that the classification be very simple and straightforward so that all levels of users can learn it; and where
- the web environment provides multiple access routes and easy, flexible, and complex representations, so a classification system with requisite richness is desirable so that these functions can be supported.

In our previous work we chose eBay.com because it was an example of a dynamic, large, messy, and inconsistent, but surprisingly robust classification. The aim of the classification in eBay is not so much to provide access to specific items, but rather to maximize the chances of a person coming across an item (and thus, bidding on it). In this paper, we will extend our inquiry of the use of classifications in commercial websites to the scheme implemented for books in amazon.com. This large, commercial site is similar to eBay.com in its dynamism, and also in its provision of many access points and multiple, flexible routes for finding items – indeed for items being “found” even without searching. It is also similar in that the aim of the site and its classification is to encourage and enable people to purchase books now, in the present, and not necessarily to provide long-term, predictable and enduring future access routes to them, as in a library.

The amazon.com classification is different from eBay’s, however, in that it is not the users who classify the items from a list of possible headings. Rather, the classification terms as well as the “thesaurus” from which terms are drawn are designed by amazon.com. This presents some additional opportunities for extending

our knowledge of how such classifications are applied, and what role they play in the access to items.

As in the previous study, we approach the analysis of amazon.com's classification scheme using the following criteria:

- scope of the classification – what does it cover?
- vocabulary – level, consistency
- structure -- the explicit and implicit relationships among entities
- granularity (scale) – the level of specificity
- expressiveness – how well does it reflect the domain it classifies?
- hospitality – how well does it accommodate new concepts?
- browsability – how well does it support and facilitate exploration of the domain?
- usability – how easy is it to use?
- coherence – how well does it “hang together”?
- consistency – is it predictable?
- exhaustivity – how completely does it cover the domain?

2. The amazon.com classification.

The classification scheme of “subjects” on the amazon.com website is a richly connected network of interrelated terms representing not only the topics of books, but also many other aspects. Thus the scope of the scheme encompasses both content and form. Here are the top categories as presented at the first level of the “Browse Subjects” feature offered when a book search is initiated.

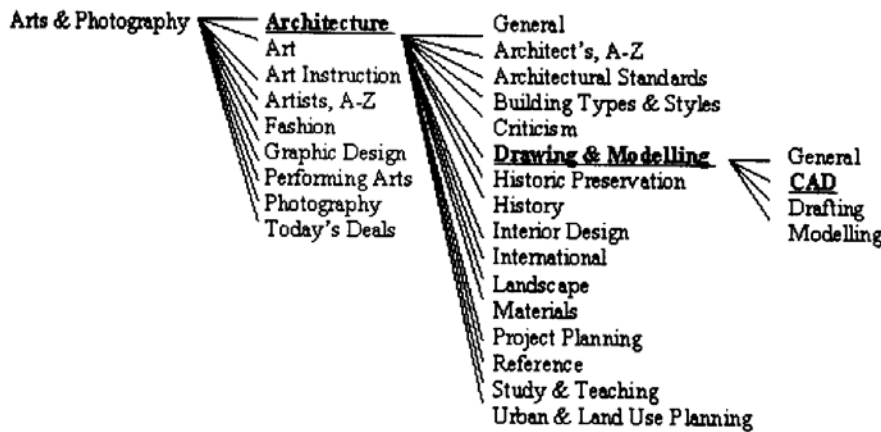
Each of these categories can be further expanded. Categories at the top level comprise quite a diverse collection of entities. Some are topics (*History, Sports*); some are genres (*Science Fiction, Romance*); some reflect a perspective (*Christian Books, Gay & Lesbian Books*); some refer to format (*Audiobooks, e-Books, Large Print*). The *Teens* category could be about or for teens, while the *Today's Deals* category stands apart as unique in emphasizing price over topic, form, or audience.

Arts & Photography	Cooking, Food & Wine	Large Print	Religion & Spirituality
Audiobooks	e-Books	Literature & Fiction	Romance
Audio Downloads	Entertainment	Mystery & Thrillers	Science
Biographies & Memoirs	Espanol	Nonfiction	Science Fiction & Fantasy
Business & Investing	Gay & Lesbian	Outdoors & Nature	Sports
Children's Books	Health, Mind & Body	Parenting & Families	Teens
Christian Books	History	Professional & Technical	Today's Deals in Books
Computers & Internet	Home & Garden	Reference	Travel

Further, the top classes are not at the same level of granularity. For instance *Literature & Fiction*, when expanded, includes many other categories, including some of the top categories. Nor is the classification mutually exclusive. A given

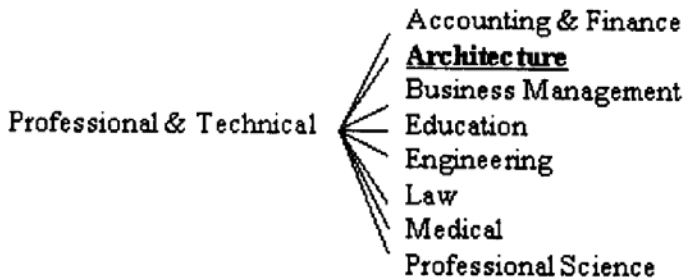
category can be found under several of the top categories, as well as under several subcategories, at different levels of specificity.

It is not possible to initiate a search in amazon.com through the subject categories. Instead a user starts by entering a term in a search box and is then presented with some top picks, and at that time is also given some related broad categories for additional searching. Let's say one types in "CAD" (Computer Assisted Design). Several suggested books are presented, and when one of them is clicked, for instance, *AutoCAD 2000: No Experience Required*, the details of information pertaining to that book are presented. Attached to the description, at the bottom of the page, are the various subjects associated with the book as assigned by amazon.com. At the same time, at the top of the page, one of the clickable tabs is "Browse Subjects." Expanding the class *Arts & Photography*, the subject path to this book is as follows:



A given book, such as the one in this example, may appear under several different categories, and even under different branches of the tree. The following classes are also suggested. Each of these takes the searcher on a different but often overlapping path.

- Engineering>Mechanical>Drafting & Mechanical Drawing
- Computers & Internet> Graphics & Illustration> General
- Computers & Internet> Web Development> HTML, Graphics, & Design> Web Graphics
- Computers & Internet>Graphics & Illustration> CAD> AutoCAD



Terms are repeated, sometimes in a slightly different form, in a variety of positions in the classification. For instance, the term Architecture appears under Arts & Photography, but also as follows under the category Professional & Technical

3. Expressiveness

Expressiveness refers to how well the terms in a classification describe and reflect the domain being classified. In the case of amazon.com's classification, the sheer number of terms assigned to any one book ensures a high level of descriptive power. Thus a book on CAD can be found whether the user construes it as software, architectural technique, or graphics. Put another way, the structure of the amazon.com scheme allows for a faceted approach, without necessarily building a faceted structure. That is, the classification achieves a multi-perspective view by redundancy of paths rather than by attention to the expressiveness of any given part of the scheme. In case any one category does not work, or is not found, another is readily available. This also ensures that *something* is found, regardless of the search specifications – a feature that, of course, enhances the possibility of making a sale.

Another measure of expressiveness is the ability of the classification to accommodate every entity within in. Normally, the presence of a "General" or "Miscellaneous" category might be a clue that the classificatory scheme is insufficiently developed. In the case of amazon.com, however, the use of *General* is merely additional insurance that a book will be found. For example, the book, *Heaven Help Us: The Worrier's Guide to the Patron Saints*, by Alice and Clare La Plante, is assigned the following categories:

Religion & Spirituality>Christianity>Catholicism>Inspirational
 Religion & Spirituality>Christianity>Catholicism>Saints
 Religion & Spirituality>General
 Religion & Spirituality>Spirituality>Prayer
 Biographies & Memoirs> General

The use of *General* does not imply that there is not a specific enough category for this book, but rather that in case a user does not think of it, or if he or she is merely browsing in a more general category, the book will be encountered.

4. Consistency, Coherence, and Usability

As already mentioned, the amazon.com scheme is not consistent in terms of scale, terminology, or structure. Terms are nested, arranged, and presented without much attention to overall coherence. Furthermore, there is no overall view of the categories and relationships among them. The schematic representation of the classification is not available to users, nor is there any one place where a user can learn about all the terms available (except for the list of top terms on the opening page of the *Browse Subjects* tab). The categories are accessed piecemeal as one goes along. Each term is clickable, and leads the searcher either to other terms or a selection of books.

This messiness and lack of mapping and predictability, however, serves the user well in one respect: it removes the necessity for the user to follow a rational

and orderly path. At every click, a new set of options in terms of books, subjects, related objects, and so forth are presented. The user merely has to choose. The system is designed for maximum connectivity between all the parts.

The negative aspect of this is that the user never has a sense of “place.” Multiple paths seem to lead to the same spot, and there is a feeling of “having been there before.” Nor does the user know whether a search is exhaustively covering the topic. He or she must rely entirely on the site’s assignments of categories. The confusion is mitigated somewhat by the availability for any given search of a list of subject categories that have been visited. The user can thus keep track of where he or she has been.

5. Subjects for browsing and subjects for searching.

Without explicitly saying so, the amazon.com site provides subject terms for browsing, and subject terms for direct searching. Here’s an example: The book *Justice Hall* by Laurie R. King is a mystery that features a fictitious young woman named Mary Russell who becomes involved with another fictitious character (from a different series of mysteries), Sherlock Holmes. Here are the subject categories amazon.com suggests for browsing (presumably to find books like this one):

Mystery & Thrillers>Mystery>British Detectives
 Mystery & Thrillers>Mystery>Historical
 Mystery & Thrillers>Mystery>Women Sleuths
 Mystery & Thrillers>General
 Mystery & Thrillers>Authors, A-Z>(K)>King, Laurie

In addition to these, and directly below this list one finds the following:

Search for books by subject:

Russell, Mary (Fictitious character)
 Fiction
 Holmes, Sherlock (Fictitious character)
 Women detectives
 England
 Fiction – Mystery/Detective
 Mystery & Detective – Traditional British
 Mystery & Detective – Women Sleuths

It is not clear, nor is it explained anywhere on the site, what these subjects represent and how they’re different from the browsing terms. Some seem to be from the classification scheme described above. Some seem to be more general (e.g., England, or Fiction), and some very specific (e.g., Holmes, Sherlock). The terminology is sometimes isomorphic with that of the classification scheme, and sometimes different. What’s the difference between Mystery & Detective and Mystery & Thriller? It’s difficult to tell. When clicked, these terms lead to lists of books, rather than to an expanded list of subjects. In other words, this is the mechanism for direct searching.

6. Comparison of amazon.com's classification with traditional bibliographic classifications

Dewey Decimal Classification (DDC) and the Library of Congress Classification (LCC) are traditional classification schemes for books used in the United States and elsewhere. These two classification schemes provide interesting comparisons with amazon.com because each emerged from somewhat different paradigms of classification design. DDC is a classification based on a model of knowledge, as reflected in academic disciplines and traditional disciplinary boundaries. LCC, on the other hand, emerged as a pragmatic way of organizing an existing collection, the actual books in the library that served the U.S. Congress. Over time, DDC and LCC have blurred the line between being "top-down" or "bottom-up" classifications, but the distinction that defined the core of these two schemes continues to inform many crucial decisions in classification design and knowledge representation: do we classify based on enduring consensual (and often academic) models of how things "go," or do we classify based on popular and perhaps ephemeral models that may change, but which have the potential for being truer in some ways to the users' perspective on the world?

The classification in amazon.com is an interesting blend of traditional approaches to classification and a creative application of the multiple access points enabled by web technology. For example, traditional classifications provide a conceptual ordering that has as one of its goals the presentation of any given concept in the context of its superordinate, subordinate, and related concepts. In amazon.com's scheme we find this context as well in the expandable subject terms (as in DDC) and the subject strings presented to the user. At the same time, there is no special care taken with ensuring that this context is conceptually coherent, stable, or consistent.

In general, amazon.com's scheme can be viewed as more pragmatic and enumerative than as based on a model of knowledge. In this way it is more like LCC. There is a difference, though, because in LCC the main classes are mutually exclusive, while in amazon.com it seems that the main classes are designed to reflect some equal distribution of postings (rather than conceptual congruence). So, for instance, there is a top category for *Mystery & Thrillers*. This is on the same level as *Literature & Fiction*, a category that might include *M & T*. One could argue that for conceptual consistency, *Mystery & Thrillers* might better belong in the subcategory, *Genre Fiction* under *Literature & Fiction*. The number of postings to the *M & T* category, however, makes this a very popular category, and thus it appears as a class unto itself.

7. Conclusion.

The study of classification in dynamic commercial websites has implications for the rethinking of classification work in general. While amazon.com has a core mission of selling things, we are coming to understand that such sites have generated many parallel non-commercial uses as well. It is not uncommon for amazon.com to be used for bibliographic work and collection development, for instance.

The classification scheme in amazon.com has been implemented to maximize the chance of finding a given book, but also of finding related books. In this the goals are similar to traditional library-classification goals, except that the

design and implementation are different. The classification in amazon.com uses multiple access routes, a simple but redundant vocabulary, and not much attention to structural integrity in order to provide a robust and rich network of subjects. This comes at the price of sometimes being lost, and of a lack of an overall view. If we could emulate the flexibility of the commercial site and preserve the structural integrity of the traditional library approaches we would have classifications that not only reflect the collection, but also better reflect the multiple perspectives of users.

References.

Kwasnik, Barbara and Liu, Xiaoyong. Classification Structures in the Changing Environment of Active Commercial Websites: The Case of eBay.com. ISKO Sixth International ISKO Conference, Toronto, 2000.