

Syracuse University

SURFACE

School of Information Studies - Faculty
Scholarship

School of Information Studies (iSchool)

7-13-2000

"Classification Structures in the Changing Environment of Active Commercial Websites: the Case of eBay.com". In Beeghtol, Clare, Howarth Lynne C., and Williamson, Nancy J. (eds.) Dynamism and Stability in Knowledge Organization.

Barbara H. Kwasnik
Syracuse University

Xiaoyong Liu
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. and Liu, Xiaoyong, ""Classification Structures in the Changing Environment of Active Commercial Websites: the Case of eBay.com". In Beeghtol, Clare, Howarth Lynne C., and Williamson, Nancy J. (eds.) Dynamism and Stability in Knowledge Organization." (2000). *School of Information Studies - Faculty Scholarship*. 140.

<https://surface.syr.edu/istpub/140>

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

Classification Structures in the Changing Environment of Active Commercial Websites: The Case of eBay.com

Abstract: This paper reports on a portion of a larger ongoing project. We address the issues of information organization and retrieval in large, active commercial websites. More specifically, we address the use of classification for providing access to the contents of such sites. We approach this analysis by describing the functionality and structure of the classification scheme of one such representative, large, active, commercial websites: eBay.com, a web-based auction site for millions of users and items. We compare eBay's classification scheme with the *Art & Architecture Thesaurus*, which is a tool for describing and providing access to material culture.

1. Introduction.

We come to this project with the assumption that the use of classification is useful in information organization. In traditional settings, indexers may use the classification inherent in a controlled vocabulary to intellectually organize the contents of a collection with the aim of enabling future retrieval of this content. Searchers, in turn, use classifications to find information. A classified collection also supports browsing -- the exploration and discovery of information that is not specified or known in advance. A classification can also serve as an overall representation of a collection's content, thereby enabling the user to learn about the domain being searched, to view options, and to establish the scope and limits of the resources. There are many existing classification schemes for documents as well as for artefacts of all sorts, but it remains an open question whether the existing schemes can be adapted for use in the virtual environments of large, dynamic commercial websites such as eBay.com and amazon.com. There are special considerations for creating classifications in these environments:

- 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;
- Since many large sites are built in a cooperative manner, it is often desirable that the classification be very simple and straightforward so that easy additions by both site administrators and contributing users are possible;
- At the same time, on many of these sites, the 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.

While these considerations are not unique to virtual environments, the growth and pace of change on the Web, coupled with the potential for rapid, flexible interfaces, create a special case for study. We report here on a portion of a larger study, presently underway, that aims to formally analyze the structures of the classification schemes of several popular commercial websites, and to compare the classification structures with the structures of their traditional bibliographic counterparts.

In this paper we focus on the structure and use of the categories of eBay.com, an online auction site. We chose eBay because it comprises an environment where classification plays an important part in representing the collection and in aiding users in retrieving the information they need. In addition, eBay.com is successful and heavily used, with a reputation for responsiveness to user needs and preferences. Thus, we hope we have chosen an example of best practices.

The classification scheme of eBay.com was analyzed describing the following characteristics:

- 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?

In our evaluation we paid particular attention to the special considerations of the Internet environment. At the same time we drew some comparisons between eBay’s classification and that of the *Art and Architecture Thesaurus (A&AT)* whose scope is also the body of material culture (A&AT, 1994). Our comparison was guided by the following questions:

- In what ways is eBay’s classification similar and different from the A&AT?
- How are these differences/similarities a reflection of the dynamic and diverse Internet environment?

2. eBay.com

Since its launch on Labor Day in September 1995, eBay.com has become the largest person-to-person auction site. Individuals – not big business – use eBay to buy and sell items. Over 3.5 million new auctions and 400,000 new items are added every day. Sellers can post items, while potential bidders can explore the descriptions of items, compare items, exchange email, and place bids. To facilitate selling and bidding, eBay has developed its own classification structure and these categories are searchable. Sellers are responsible for choosing the category under which the item will be listed from a classification scheme of about 2,900 categories. Bidders and sellers can also search for items using many other strategies, including the following:

- *keywords* in the titles and descriptions of posted items, supported by *Boolean operators*
- items up for auction *by a particular seller*
- items bid on *by a particular bidder*
- *recently completed* auctions
- by *eBay item number*
- by items *most recently posted*
- by items *closest to close of auction*
- by a “watch” list of *bookmarked* items.

Of these multiple strategies the use of the classified category structure is but one. This is an important consideration in the assessment of the classification, because it is clear that users may successfully use eBay without ever having to understand the categories or how to consistently apply them. There are many other ways, besides using the categories, for locating specific items or groups of related items.

2.1 The Structure of the eBay.com Classification

The eBay.com classification is a contextual, pragmatic, and inductively created classification – built up from the actual collection of items for sale rather than from some conceptual or theoretical framework. When we first started this project in November, 1999, there were eleven top categories. Soon after, the classification underwent a major revision based on feedback from users and the exigencies of more evenly distributing the vast numbers

of posting in some popular categories. We have no doubt that this structure will change again, and continue to change as the distribution of auction items changes. As of February, 2000, there are fourteen top categories:

Automotive	Great Collections
Antiques & Art	Jewelry, Gemstones
Books, Movies, Music	Photo & Electronic
Coins & Stamps	Pottery & Glass
Collectibles	Sports
Computers	Dolls, Doll Houses
Toys, Bean Bag Plush	Everything Else

These fourteen categories are further subdivided into a maximum of four sublevels in a tree-like structure, sometimes hierarchical, but not always. The logic of the top categories is purely pragmatic. Each top category represents from approximately 2 percent to approximately 25 percent of the collection. The largest category in November, 1999 was "Collectibles," representing about a third of the collection. This was obviously too large (and vague), so in the revision, additional top categories were formed and these were used to reduce some of the density in the old categories. The "Collectible" category seems to be in constant flux – the nature of what is collectible being so contextually determined. Other categories underwent similar pragmatic shifts, additions, divisions, and changes. For instance the "Sports Memorabilia" category changed to "Sports," and the "Sports Memorabilia" category became a subdivision of it.

We analyzed all the categories for frequency of posting in the categories and the criteria by which subdivisions are made. Table 1 shows a small portion of the classification along with the number of postings (items available for auction) in the various categories as of February, 2000. The "Criterion for Division" column is our interpretation of the logic used to differentiate this category from its parent category. So, for instance, "Ancient World [Antiques]" are differentiated from all "Antiques" by time and place; "Ceramics," by material; "Books & Manuscripts" by form, and so on.

Top Category	Level 2	Level 3	No. of items on auction	Criterion for Division	% of All eBay
Antiques&Art	General		24,279		0.707
		Ancient World	10,263		0.299
		Architectural	2,962	Time/Place	0.086
		Asian Antiques	2,894	Use	0.084
		Books, Manuscripts	1,569	Place	0.046
			5,605	Form	0.163
			3,128		0.091
			13	Place	0.000
			1,684	Place	0.049
			63	Place	0.002
		432	Place	0.013	
		266	Place	0.008	
		19	Place	0.001	
Ceramics	Latin American	986	Material	0.029	
Etc.					

Table 1. Example of the eBay Classification, with Number of Postings and Criteria for Division of Categories.

This excerpt from the classification is typical of the whole. The rules for division are neither exhaustive nor consistent. For instance, "Antique Books & Manuscripts" are further

divided by place – but sometimes by continent, sometimes by country. There are no guide terms (facet indicators) as there are in the *Art & Architecture Thesaurus*. So while the *A&AT* specifically indicates that a set of subcategories is divided <by location> or <by function>, the eBay classification does so only implicitly (and rather inconsistently). There are also some persistent anomalies. One can find antique books in either the “Antiques” category or the “Books, Movies, Videos” category.

Throughout the classification the use of the “General” category accounts for a large percentage of all postings. In the excerpt shown in Table 1, there are 10,263 “General” Antiques and 3,128 “General Books & Manuscripts.” Of the postings in the example in Table 1 specific categories are used by sellers only about half of the time. This implies that the eBay classification is not adequately expressive, but at the same time it is not clear that the purpose of the classification is to provide close, precise, descriptors. From the seller’s point of view, the category that gives the most exposure (i.e., is most likely to be chosen by bidders) is the most desirable, rather than the one that provides the most “correct” label.

There are no rules for how to choose a category, no scope notes or instructions, and it is at the discretion of the seller which aspect of the object to emphasize by the category that is chosen to represent it. For instance a book titled *Depression Era Dime Store Kitchen Glass* was classified by its seller under:

“Books, Movies Music: Books: Non-Fiction: Collectibles”,

while a similar book, *The Depression Kitchen Glass Book* was classified by its seller under:

“Pottery&Glass: Glass: Kitchen Glassware: General.”

The first book would find itself with other books on collectibles; the second with kitchen glassware. The classification structure itself is not consistent enough to provide guidance by example. By contrast, the *Art & Architecture Thesaurus* requires category names to be placed in their most generic location if there is a logical choice of locations. Thus “columns” would go under “architectural elements” rather than under “temples.” (*A&AT*, 1994, p.33). If this rule were applied in the eBay example, the second book would have to be with other books (its most generic “home”). As it stands, the seller has some discretion in assigning the category. Whether this is useful to browsers is an open question. In other words, in viewing the results of a search for the antique kitchen glass being auctioned, does a browser also want to see *books* on the topic? Put another way, the marketing aspects of selling on eBay can influence the use of the category structure in a way that is perhaps inconsistent with high-precision searches, but might be useful in terms of creative juxtaposition of items from different branches of the classification structure. The purpose of all item representation in eBay is, after all, to present opportunities, rather than to catalog for posterity.

There are no syndetics in eBay’s classification scheme: no cross references to *use* and *used/for* terms or to related subjects as there are in the *A&AT*. This does not, however, preclude eBay users from producing creative groupings of items using other strategies available to them, similar to citation analysis. For instance, bidders can view the seller’s other auctions (the implication being that the seller will be selling similar items). One can also view the bids of a particular bidder. These are indirect ways of creating connections between items. Sometimes these yield richer relationships than might be provided by the category labels because they might include other similarities such as price range or style that are more salient to the user.

2.2. The Scope of eBay’s Classification

The eBay classification covers millions of items presently up for bid. The range of these items is extremely broad, but it is contemporary – that is, the items exist in the present. Consequently the classification is responsive to the distribution of items as they are submitted by sellers over the course of months. While Beanie Babies continue to be popular, for instance, eBay maintains a whole separate category for them. Once they fade from popularity,

that category may merge with another, such as “Toys.” The shifting scope of the eBay collection, its rapid turnover, as well as the fact that the classification is used by amateurs as well as experts, all conspire to make the eBay classification rather *ad hoc* in its structure. By contrast, while the *A&AT* is also designed to cover an extremely broad scope of material culture, that thesaurus is built to accommodate not only contemporary artefacts, but also to anticipate the description of artefacts in the future. Towards this end the *A&AT* is systematically structured, using a theoretically based faceted approach that will endure changes in trends and popularity. eBay’s classification is built as a *response* to the scope of the collection, while the *A&AT*’s classification is built *in anticipation* of use.

2.3. The eBay Classification Vocabulary

There are three sources of “vocabulary” in eBay: 1. the titles and 2. the descriptions, written by sellers in natural language, and 3. the categories, which are supplied by eBay and chosen from a controlled vocabulary by sellers when they put their items up for bid. One of the problems that eBay shares with the *A&AT* (p.40) is the difficulty of describing both generic as well as unique and one-of-a-kind objects. So an item may be generically a “Toy,” but specifically it is a particular issue of a Beanie Baby, with a particular date, name, etc., readily recognizable by aficionados. Similarly, in the world of material culture, some items are generic (such as a suit), while some are unique and form a category of one (such as the suit worn by Elvis Presley in *Viva Las Vegas*). This tension is particularly obvious in eBay’s classification and is reflected in the many category terms that are the names of specific brands, issues, and so on rather than generic terms.

Since the terms that comprise the *titles* of items on eBay are devised by the sellers, it would seem logical that the controlled terminology of the *classification* should balance out the possible confusion of homographs, spelling variants and errors, such as misspellings. This does not seem to be the case. Consider the example in Table 2. The term “light” is used in two senses, but the various categories under which the items were classified do not consistently help disambiguate these senses.

Title of Item (Provided by Seller)	eBay Category (Provided by eBay but Chosen by Seller)
Diet Dr. Pepper Light	Collectibles: Advertising: Soda: Dr.Pepper
Coors Light Beer Lamp	Collectibles: Breweriana: General
Red Crayola Night Light	Collectibles: Decorative: General
Genessee Light	Collectibles: Breweriana: Signs,Tins

Table 2. An Example of Vocabulary and Category Inconsistency on eBay

One of the interesting aspects of the eBay world is the development of a sublanguage that seems to crop up in the titles of items supplied by sellers but not in the controlled vocabulary. One of these terms, for instance, is “shabby chic” – a descriptive term meaning, in essence, “the paint is peeling but it’s charming.” This term is used as a sort of facet to describe everything from lamps to chipped garden gates. “Shabby chic” does not appear as an eBay category, however, even though a search on this term brings up several hundred items. On the other hand “Made in Japan” is a category for ceramics (as distinct from “Asian”). So it seems that some insider/expert terminology is used in the classification following the principle of literary warrant, while other insider terminology is reserved for the auction lingo of the titles and descriptions. eBay polls its users and advertises upcoming changes, but is not clear how colloquial terms might eventually become part of the official vocabulary.

3. Evaluation of eBay’s Classification Scheme

In formal classification terms, the eBay scheme has many shortcomings. It is admittedly opportunistic in its structure and vocabulary. It is neither totally coherent nor always

predictable. There is very little guidance and few rules for users to consult in choosing a category. In terms of granularity, the vocabulary is simultaneously generic and microscopically specific, sometimes within the same category. The pervasive use of the "General" or "Everything Else" category suggests that the classification is not sufficiently expressive. While facets are implicit in the criteria for class division, these facets are not specified and are inconsistently applied. The overall structure of the classification is unstable and shifts periodically with rather sweeping changes (such as changes to the top categories), major migrations and no syndetics to guide the user through the change. Certainly when compared to the *A&AT*, the eBay classification is not systematic, nor is it based on a theory of classification, and it does not conform to thesaurus standards.

Finally, the classification scheme is not a very good way of exploring the various domains of items included in eBay auctions. For instance, the scheme does not support exploration of the contributing streams of expression in the Arts & Crafts Movement. One can find a variety of diverse items described as "Arts & Crafts," but the connections between them are accidental rather than systematic, or at any rate they are not explicit. This is because the eBay classification is not meant to be a stable representation of knowledge in the formal sense (Kwasnik, 1999), but rather a handy enumeration of items and a partitioning of these items into manageable clumps.

On the other hand, the eBay classification is surprisingly robust and hospitable. When you consider that it reflects the terminology used by several million people for an amazingly wide and constantly shifting array of items, it is really quite impressive. In addition, it is flexible and responsive to the changing scope of the collection it describes. In terms of accessibility, there are very few terms in the main categories that are difficult to understand.

One of the strong points is the way in which the classification is meshed with all the other access strategies, providing the user with many avenues of pursuit. Thus, if the classification fails or falters, there is always some other way to find "good stuff." Moreover, there are several creative ways of clustering items *besides* the categories. In this way the classification itself becomes a tool in a suite of tools for description and retrieval. The redundancy makes eBay easy, intuitive, and fun to use, even if it is not used optimally.

4. Future Research

As mentioned earlier, the discussion in this paper comprises a portion of a larger ongoing project in which we are analyzing the classification structures of commercial, active, large websites and comparing them to their traditional counterparts. We are presently analyzing the classification scheme of amazon.com and will next compare it to traditional bibliographic schemes, such as the *Dewey Decimal Classification* and the *Library of Congress Classification*. We would also like to compare the classification scheme of a medical website with the *MeSH* classification, but we have yet to identify a site that uses its own categories as extensively as do amazon.com and eBay.com.

From this exploratory study we would like to learn how we can enhance new classification work with what we know from our existing tools, and also, how the changing virtual environment is creating new opportunities for dynamic, responsive classifications.

5. References

- eBay category overview (2000). <http://listings.ebay.com/aw/listings/overview.html>.
- Kwasnik, B.H. (1999). The role of classification in knowledge representation and discovery. *Library Trends*, 48 (1): 22-47.
- Petersen, T. (1994). *The Art & Architecture thesaurus*, 2nd edition. New York: Oxford University Press.