My Summer at Amazon.com

Kate C. McDill
km7372@u.washington.edu

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Recommended Citation
DOI: https://doi.org/10.7771/2380-176X.3173

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
My Summer at Amazon.com

by Kate C. McDill  <km7372@u.washington.edu>

Traditional library and information science (LIS) theory and practice are well developed and have come to be used by e-commerce companies to sell goods and services. The axiom of the marketplace is that sellers need buyers to be able to find what they want — effortlessly. In the information world, people with information want people needing it to be able to find and access it easily. Last summer I worked as an intern at Amazon.com, assigning browse nodes to nodeless ASINs in the toy & baby stores. What this really means is I applied subject indexing to merchandise, utilizing the in-house taxonomy, with the goal of improving the information retrieval.

Amazon.com’s goal, besides the obvious one of profitability and making Wall Street happy, is for a customer to locate needed information or to purchase goods easily. This requires robust taxonomies, called browse trees in the language of Amazonia. The items in the Amazon.com catalog are all assigned an Amazon Special Identification Number (ASIN) and every browse leaf node has its own ID number. I spent the summer connecting ASINs to all applicable browse nodes using MS Excel spreadsheets.

Amazon.com has two types of stores. For the Books, Music and Video (BMV) stores, they use cataloging and subject terms direct from the specific industry. All other merchandise is sold through “hardline” stores, which requires a person (me) to assign browse nodes (index) and subject terms to each item individually. For the toy store, in addition to the “category” browse tree, there are “age range” and “brand” browse trees, which makes indexing toys a greater challenge. The browse structures are not faceted but are hierarchical. The browse trees have main branches and leaf nodes. Each leaf node is assigned a specific number in the database and when a customer clicks on “category” or “age range,” the different branches are shown. For example, the board game Monopoly would be assigned the leaf node: Toys > Categories > Board Games > Classic Games. A customer could also find it through the age-range categories. Unless a searcher knows the age-range suggested by the manufacturer there is a bit more guess-work involved. Monopoly is recommended for 8 and up and so the game would surface in three age ranges: 8-11, 12-14 and grownups. It also would be listed under two leaf nodes in each age range: games > board games and games > classics. This game is also searchable as Monopoly or by the brand name. Parker Brothers and Hasbro both hold brands to Monopoly. The brand branch gives an additional two places to find the game: Parker Brothers > Family Game Night and Hasbro > Family Game Night. The total number of leaf nodes assigned to Monopoly would be nine. Can you imagine assigning 9 LC codes to a single book?

A consumer searching Amazon.com would think that the items are co-located in many different places. Each item obviously only has one home in the warehouse, such as in a library. The necessity of cataloging items in a broad array of places in an e-commerce site validated my LIS education: items must be organized in many different ways to be understandable to the users. Additionally, the browse structure must be flexible enough to handle fluctuating inventory and consumer interest. (The toy industry is notorious for items blossoming in popularity and dying very quickly.)

The application of LIS theory and practice is alive and well on the Amazon.com site.

Poking around to find ATG back in 1992 (can you believe that was ten years ago?) to see who we were talking about, here are a few of the Rumors from the first issue of the year (v.4/i1) (ATG had five issues back then, was 48 pages, cost $25 a year, and Rumors was less than two pages long!) — Jolanda von Hagen had left Springer Verlag, Ameritech had bought NOTIS, Christine Lamb was named Director of the Faxon Institute, Audrey Melkin had left Wiley for Henry Holt, Rozicka Bindery had changed its name to Southeast Library Bindery, Inc., and Gary Shirk had just been made President / COO of Yankee Book Peddler. And that’s not all and it’s just the first issue.