"Next Generation" Library Catalogs in 15 Minutes

A "next generation" library catalog starts with the idea of traditional library catalog and expands it meet the changing expectations of library patrons.

Collections

A "next generation" library catalog includes *content beyond the content a library owns*. It includes the content needed by the students, instructors, and scholars necessary to do their learning, teaching, and research. While it will never be possible to acquire all the necessary content, some of it will include bibliographic data from journals, theses & dissertations, government documents, images, movies, sounds, data sets, etc. By combining the metadata as well as the full text it will be better able to perform relevancy ranking and identify obscure data and information. Moreover, the full text will enable the system to provide the enhanced services outlined below. In the meantime federated/broadcast search will fill a gap but its promise will never be fulfilled for all the reasons we already know. Speed. Network latency. Dumbing down metadata. Screen scraping. Long term maintenance.

The library catalog lies at the intersection of collections and services.

Indexes and databases

A "next generation" library *catalog is an index*, not a database, in order to facilitate search in a manner expected by the patron. An index allows for easy keyword searching, phrase searching, fielded searching, sorting, and most importantly relevancy ranking of output. Such processes are very difficult to implement against a database because when searching a database it is necessary to know the structure of the database before hand. Databases are a component of a "next generation" library catalog, but the index is the tool facilitating search.

The most significant difference between traditional library catalogs and the "next generation" library catalog are the *services applied against the content* of the system. These services go beyond enhancing the search process. They go beyond intelligent find, relevancy ranking, spell check, synonym facilitations, faceted browse, and cover art. They even go beyond putting search results in the context of the user's environment -- a topic for another presentation.

Services

Instead, a "next generation" library catalog will provide services against the things discovered. These services can be enumerated and described with *action statements* including but not limited to: get it, add it to my personal collection, tag and classify it, review it, buy it, delete it, edit it, share it, link it, compare & contrast it, search it, summarize it, extract all the images from it, cite it, trace it, delete it. Each of these tasks supplement the learning, teaching, and research process. They are tools and processes our students, instructors, and researchers use to accomplish their individual goals. All of these processes are things libraries already provide in our physical environment, but with the advent of our globally networked environment libraries need to figure out how to provide these services on people's computer desktops.

I summarize the three points outlined above in the following way. "Collections without services are useless. Services without collections are empty. The library catalog lies at the intersection of collections and services."

Implementation

The computing model for such a system is not too difficult to illustrate. More or less, this is the is the same model employed by the eXtensible

Catalog (XC), Primo, and probably other applications. The keys to implementing this model in software lie in putting into practice three *engineering principles* and a few *community/leadership activities*.

The engineering principles begin with making small applications that do one thing and do one thing well; do not go about creating one huge system. Second, make sure your small parts work well with each other. This usually means supporting a concepts called standard input, standard output, and standard error. Finally, use plain text as much as possible, not binary data, since plain text is a universal interface.

When it comes to community/leadership principles, employ community-driven standards as much as possible. This will enable modularity. Support experimentation, innovation, and play. This will enable your institution to be nimble, flexible, and more responsive to user needs. These activities allow institutions to be leaders as opposed to

followers. They empower you to be more proactive. Understand that *everybody has something to offer*. A "next generation" library catalog is primarily not a computer problem as much as it is a library/university problem. By getting as many people involved as possible it will be easier to create a system that meets most people's needs. The creation of the Linux operating system is a good example of what can be done using the engineering and community principles outlined above.

Eric Lease Morgan
University Libraries of Notre Dame

November 13, 2007