What is a Collections Search ?

This term describes any application for digitally searching a museum's collection. This can be made available to the public or used by only the internal staff for curation, and can be made available to a selected 'professional' or research audience. There are many issues that such a search can address. When setting up a search, ease of use or understanding and responsiveness are addressed, also issues of ownership or privacy all combine to determine the search technology chosen and how it is applied. The search realm and results can be tailored based on the user.

Why have a Collections Search ?



Scalable Web and Search Solutions™

Collections Search for your Museum:

A white paper by Rich Roth, TNR Global and We Love Museums, April 2012 (v1)

Enhanced your museum's digital experience can have many paths. Offering a collections search can open up your museum to a whole new audience or enhance the experience of your existing audience. By allowing the public to explore your collection, they can see there is more to the collection then expected. This applies even if the search is open only to a select audience, This can draw visitors back to the museum, or interest those who wouldn't normally get to see your collection.

Why not just use Google search?

The difference between Google search and a proper collection search is the difference between external and internal data views. Google can only index what is visible on a web page. Google has no understanding of what is on a web page, beyond the emphasis of certain words. An collection search creates an index using full access to all the details of each item, as well as the ability to include relationships between objects, such as location or provenance.

What is the process of creating a Collection search:

Functionally the system for searching a collection has there are two overall processing pipeline steps in a search of a collection, no matter what the format of the information or the underlying search technology used:

Index and query

As shown in this diagram:



What makes a search system different from a more common database is the preprocessing of the data that takes place. This process allows the query requests from the index to produce more complex and faster results then a database, without requiring the exact structuring required by a database. This preprocessing also provides the information to offer spelling and related suggestions.

The indexing system includes: source data inputs, then connection to a document processor, which feeds into an indexer that loads the actual index. This process is repeated as often as the data is changed - and includes both unstructured text and minimally structured textual parameters to provide relationship information. The exact format of the index is a function of the special search system and varies widely.

The raw information about the collection needs to be processed into a format that can be indexed. This might include data entry, or more likely in today's computer based world, converting the existing curatorial tracking database into a format that can be loaded into the indexer of the search system. For many that have used a program like Past Perfect or ContentDM, there are standard connectors to extract the needed information in the right format for directly loading into the indexer.

It is important that this is done so that each item can be updated in the index as information for specific items are changed or as items are added or removed from the collection. The information also has to be structured to allow queries to include important information, cross-references and relationship details. An example of why this is important is because it allows the person to drill into the results of a query to further explore the item or related items — a typical found item will be part of a set of items by category or item characteristic and someone exploring might want to see similar items. This is done at query time by using the relationships loaded at indexing time.

For collections indexed using custom databases, or simple text techniques, a custom index connect must be created. For most cases, this is based on existing generic tools for indexing database records, or format text recognizers. Once these are created, they can be reused as the collections change.

For most museum collections, updating is an occasional process, as acquisition are made or after periodic collection inventories.

Querying happens whenever a search request occurs. This is more commonly provided on a web site, either publicly facing, or a protected area of a public web site ,or even internal to the museum, such as a exhibit kiosk. Each of these situations calls for a different presentation, and might include pre-defined limits on what part of the collection is searched or what information is presented. This might also include special processing, such as showing only images with watermarks or cross-references to upcoming events.

The query process is integrated with the museum web site. For web frameworks or CMS, this is done using a standard search plugin. For custom coded web sites, via standard form/search results API's.

Where: where does this all happen ?

No matter what the technology or techniques used, there is still a computer somewhere that runs the various steps. This could be at the museum, or the museum's ISP, with the web site, or today, very often both are 'in the cloud' - which is really nothing more then a computer 'out there' run by a company like Amazon being responsible for keeping it all running. The techie term is 'infrastructure' and there are various options with various costs and responsiveness appropriate to the size and nature of the collection and search activity.

The systems are selected to match the reliability and performance needs of the museum. This might be as simple as the size as measured by cpu speed, memory and disk size or a configuration of multiple servers, possibly at different locations, connected into a common network with load balancers. There are a number of open source search options, many build around the Lucene search core.

Ownership and Intellectual property:

This is an issue that concerns most museums which we discuss collection search, especially where the positive benefits are not very clear. There are very few issues that have more reactions and less substance than concerns and fear over how search of collections will damage the value of the museum property or ownership rights will be effected . First off, this issues can always be addressed in how the search is structured, what information is made available and, esp with images, how the images are displayed: resolution and use of water marks. More important, the best way to make a museum collection more valuable is to make the public more aware of it -- this increases the value of the collection, not decreases it. The concerns regarding intellectual property are largely misplaced and those that are valid can be addressed through proper application of technology.

Public search vs protected search

Search does not presume public access to either the search screen or the total possible results, especially images. Many museums, specifically curators, have concerns about publishing public information. The collections database should be considered as raw material in determining how best to present the collection to a broader audience. The simplest approach is to allow searching the whole collection and present all the results in a meaningful display. The next step is the extreme one of limiting the visibility of the results to only logged on visitors.

A more effective approach is to better define the audience (or audiences) for the collection and the search. Then use this information to structure the presentation of the results. The public view can be a more restricted display, while protected view and can be more expansive and detailed.

Why TNR for your muse - in two words - Technology and Passion

Technology:

TNR Global is one of the top small search technology firms, we have over 10 years experience with the latest technology in creating powerful custom search systems.

Passion:

We have passion - passion for our work and passion for museum's goals — to share the unique experience that you have created and developed through your collection. The creation and continuation of every museum comes from a passion for their area of interest. We have both the passion for the technology and for the nature of the museum experience.

Experience and Knowledge:

The founder of TNR Global and principal of We Love Museums, Richard Roth, is a life-long museum lover and 4th generation techie. His development team is a key member of the development team for one of the largest online industrial search engines Thomasnet.com which indexes and searches 10's of millions of parts, documents and web pages and updated on a daily basis. He has been involved in data conversion projects for over 40 years, including founding Dataviz, a primer and founding company in the field of data conversions. Personally, he has visited museums worldwide, is a member of dozens and a 15 plus year member of the Shaker Seminar. Mr Roth is involved in a number of archival and curatorial indexing projects.

About We Love Museums

WLM is a web site dedicated to the museum goer and you need to be listed, and listed correctly. We have found some 15,000 museums based on web searches but only you can provide the right story to our visitors.

We are working on a number of options, such as the personal membership pages, and ties to other reviews, calendars or reciprocal memberships Just consider a special home display for your members only.

welovemuseums.com

- over 15,000 US museum listings
- reviews
- wlm@home an image database of 10,000 art images for home display.
- Personal pages for museum memberships

Contact:

TNR Global: museums@tnrglobal.com http://www.tnrglobal.com We Love Museums: info@welovemuseums.com http://www.welovemuseums.com