IR of Non-Text Materials

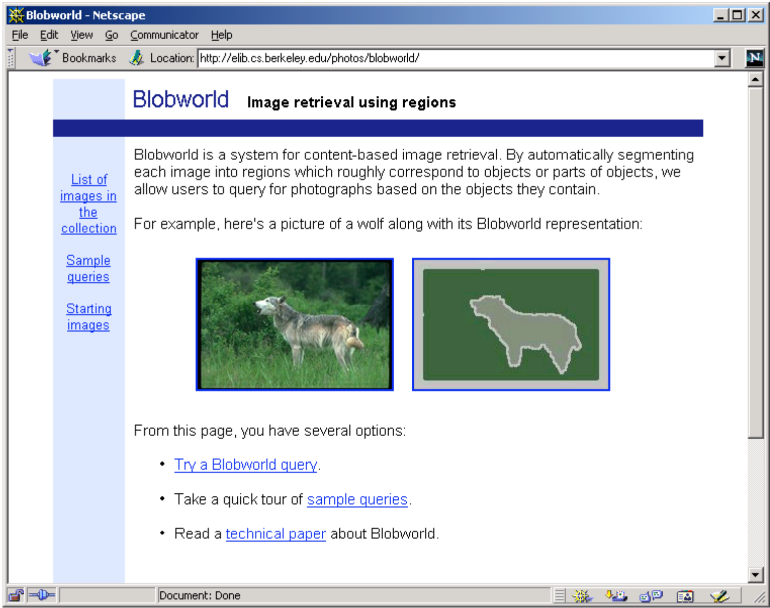
IR applied to non-textual materials isn't that much different from IR for textual materials. The primary differences are the attributes of the individual types of documents that are stored and searched. For example, people using maps locate items on the map by latitude and longitude. So it follows that when extracting data from a map to be the document representation, one should store data intrinsic to that object: latitude and longitude. Here's a list of examples:

Content	*Attribute*
Maps	Latitude, longitude; continent; country; …
Photographs	Subject of the photo; date and place, etc.
Bird songs and images	Field mark, bird song
Software	Task, algorithm, class
Data set	Survey characteristics
Video	Subject, date, record speed, etc.

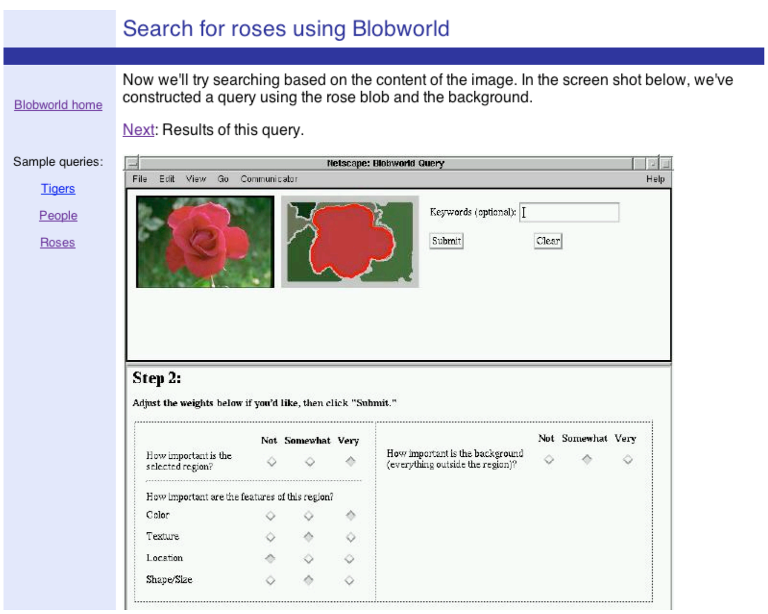
How might we locate non-text materials? One way, naturally, is through human indexing, through manually created metadata records. There are also automated techniques such as context (associated text and links, similar to web IR) and especially *image recognition*. Finally, of course, there are end-users who can browse, through the interface, thumbnails and descriptors of the material. Experiment by searching Harvard's Visual Information Access Collection online, (http://via.harvard.edu:9080/via/deliver/advancedsearch?_collection=via).

Automatic image recognition

The specifics of automatic image recognition are inappropriate for this class, but the use of image retrieval IR systems are useful. One example is UC Berkeley's "Blobworld" (http://elib.cs.berkeley.edu/photos/blobworld/)



The next image suggests how one could search for an image, specifically a rose, using Blobworld:



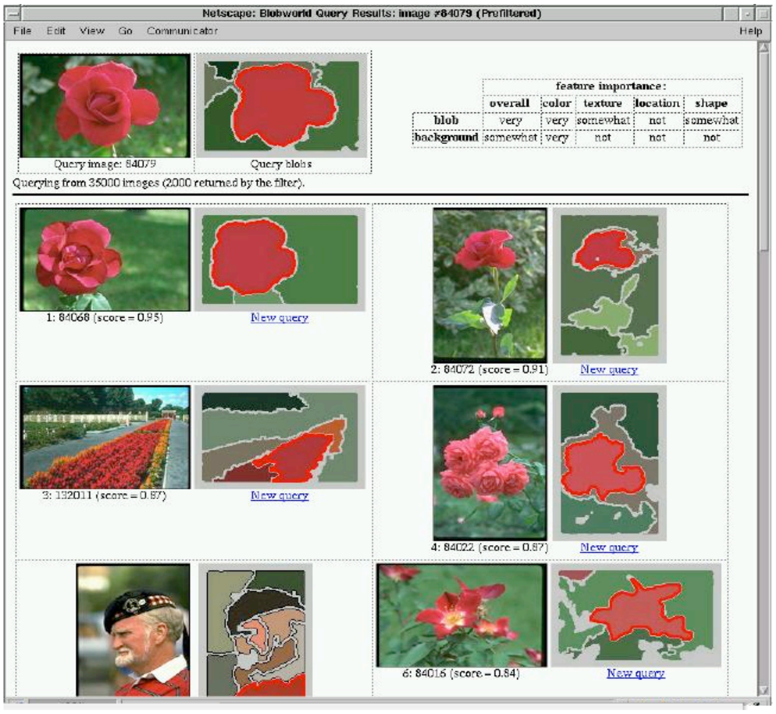


Image Understanding:

Surrogates

Surrogates play an important role in retrieving images. Depending on the source of the surrogate, the searching technique will vary. In a library environment, there is usually a catalogue record for the item. In art and architecture libraries, there may well be some kind of specific classification schema that are useful for identifying theme unique to the arts. In archives, museums, and some libraries, there are "finding aids." In the fields of Archives, the finding aid is the lower unit of access, which points to the *container*, but not to the specific *items* within a container.

As the Harvard visual collection demonstrates, after searching for a query term, visual collections require end-users to browse through the summaries (thumbnails, titles, some other unique identifier, etc.).

Catalogue records for non-text materials Any cataloguer knows that anything can be catalogued, using the GMD (general materials designation) field and good subject tracings. Metadata standards, therefore, also apply, such as Dublin Core. Visual materials tend also to have specific subject-based metadata standards that apply to specific categories of materials. For example, for geospatial materials, there is FGDC. Regardless of how the record is created by the indexer or cataloguer and overlooking the specific *human* meaning extracted from using metadata standards, the creation of surrogates for visual items means text-based searching methods can be applied.

Automatic creation of metadata records

Earlier you read about automating means to extract features from text documents. There are also techniques for automatic extraction from non-text materials. However, the effectiveness varies considerably by the technique used and the field of use. While some simple, black-and-white images, such as fingerprints, are reliably processed, images are by far the most difficult to process automatically. Many techniques consider the characteristics of color, texture, shape, and probabilistic methods of guessing an object's identity by comparing the rough outline to similar, already-identified objects.

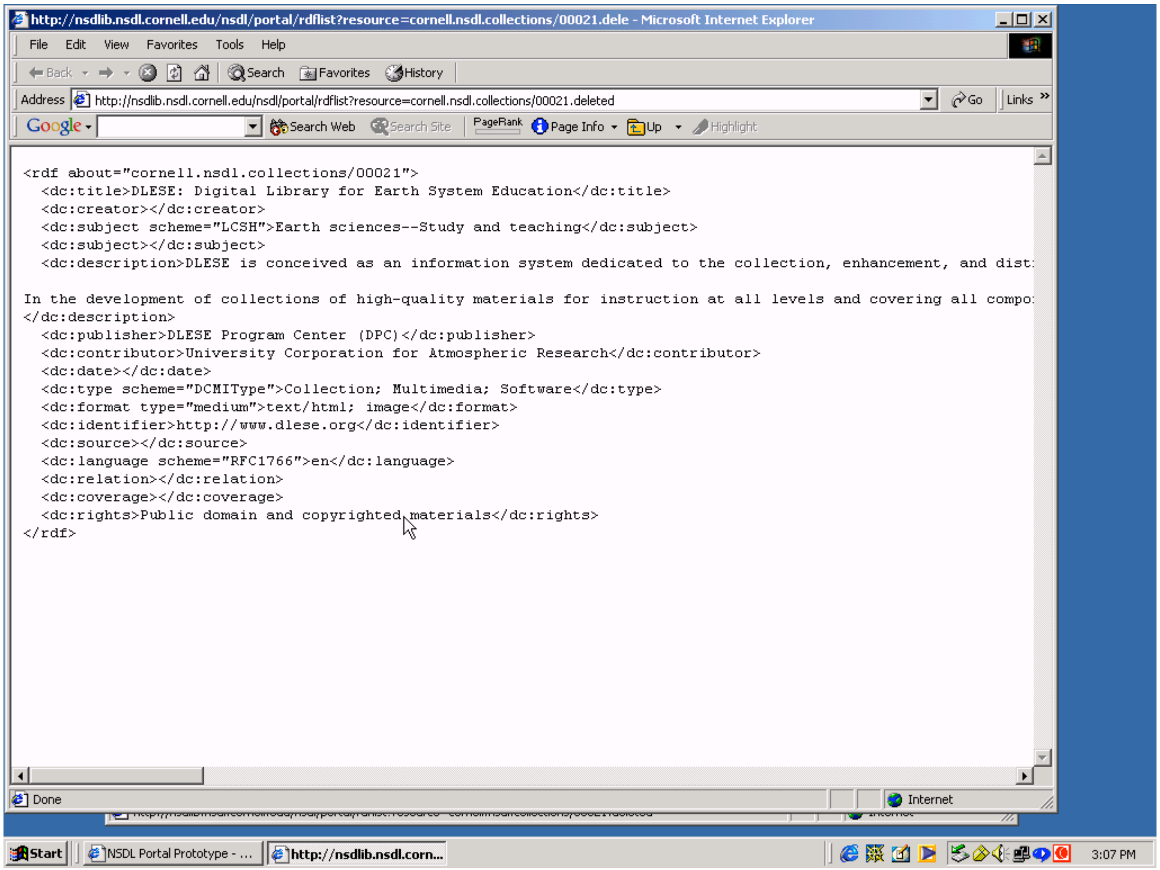
Interestingly, music retrieval is rather successful. For printed scores, optical recognition techniques are quite mature. For sound, e.g., bird sings, spectral analysis is reliable.

Human created metadata: EAD:

Archives students are already familiar with the usual arrangement of archival material and finding aids. The finding aid is a list (or inventory, or index) or other text document created to describe the holdings. The finding aid provides fuller information than is usually contained in a catalogue record … or it could be even less specific. As mentioned above, archives' finding aids do not process the lowest level object in a collection, the "item." This means the item cannot be found through automatic means.

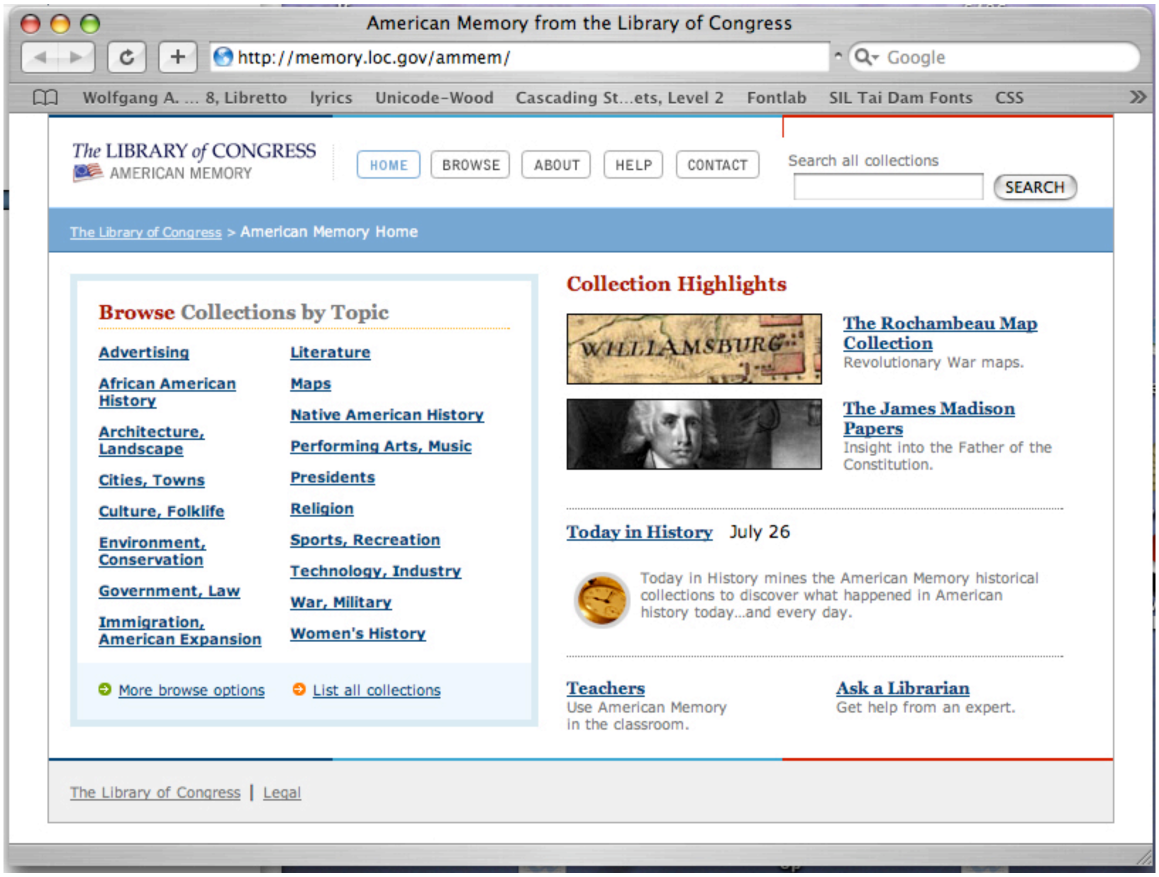
The most commonly used SGML implementation today for archives is the *Encoded Archival Description* (EAD). All XML are heavily structured; much of the information is derived from hierarchical relationships, although this is being challenged in some research. The challenge here is just what to apply to documents (the same metadata used to describe the entire collection or something new?) For the moment, though, we must consider Archives and emphasizing *collection-level metadata*. For example, one record might describe all the images in a photographic collection, instead of a real surrogate to identify individual items.

Here's a sample of a Dublin Core metadata record for digital library of non-textual material:

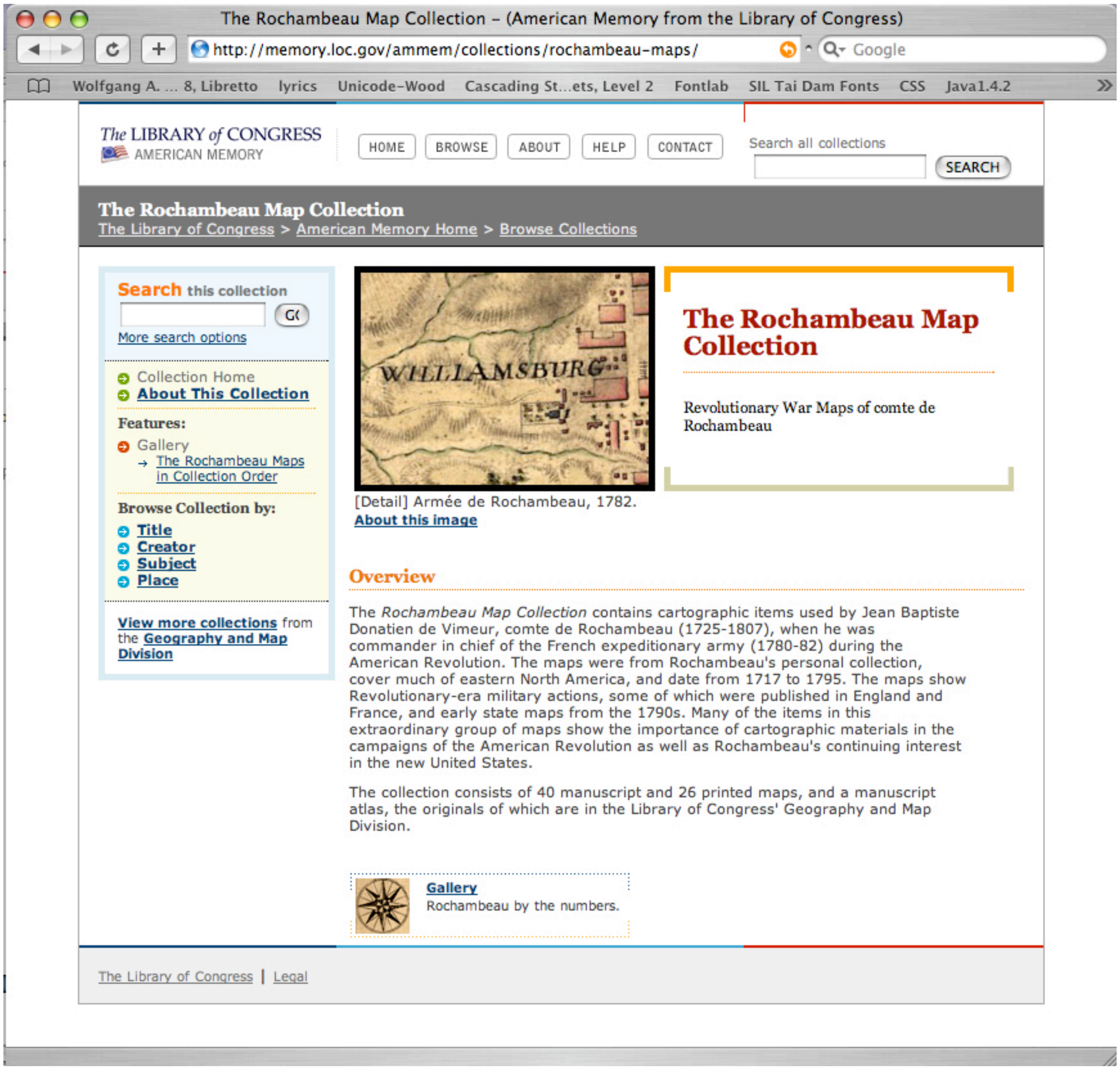


Photographs

Probably the most popular and useful (and free) photographic collection is in the United States Library of Congress' American Memory collection [http://memory.loc.gov/ammem/mgpqery.html]. Each photograph in that collection is described by a MARC record. This means searching for text and non-text materials are processed exactly the same way.



Individual items are located either universally ("search all collections") or by domain (here the Rochambeau Map Collection). Below is a copy of the homepage for the Rochambeau collection, followed by the "search descriptive information" screen.


The Rochambeau Map Collection – (American Memory from the Library of Congress)
http://memory.loc.gov/ammem/collections/rochambeau-maps/
Wolfgang A. ... 8, Libretto lyrics Unicode-Wood Cascading St...ets, Level 2 Fontlab SIL Tai Dam Fonts CSS Java1.4.2

The LIBRARY *of* CONGRESS
AMERICAN MEMORY

HOME BROWSE ABOUT HELP CONTACT

Search all collections SEARCH

The Rochambeau Map Collection
The Library of Congress > American Memory Home > Browse Collections

Search this collection GO
More search options

Collection Home
About This Collection

Features:
Gallery
→ The Rochambeau Maps in Collection Order

Browse Collection by:
Title
Creator
Subject
Place

View more collections from the **Geography and Map Division**

WILLIAMSBURG

[Detail] Armée de Rochambeau, 1782.
About this image

The Rochambeau Map Collection

Revolutionary War Maps of comte de Rochambeau

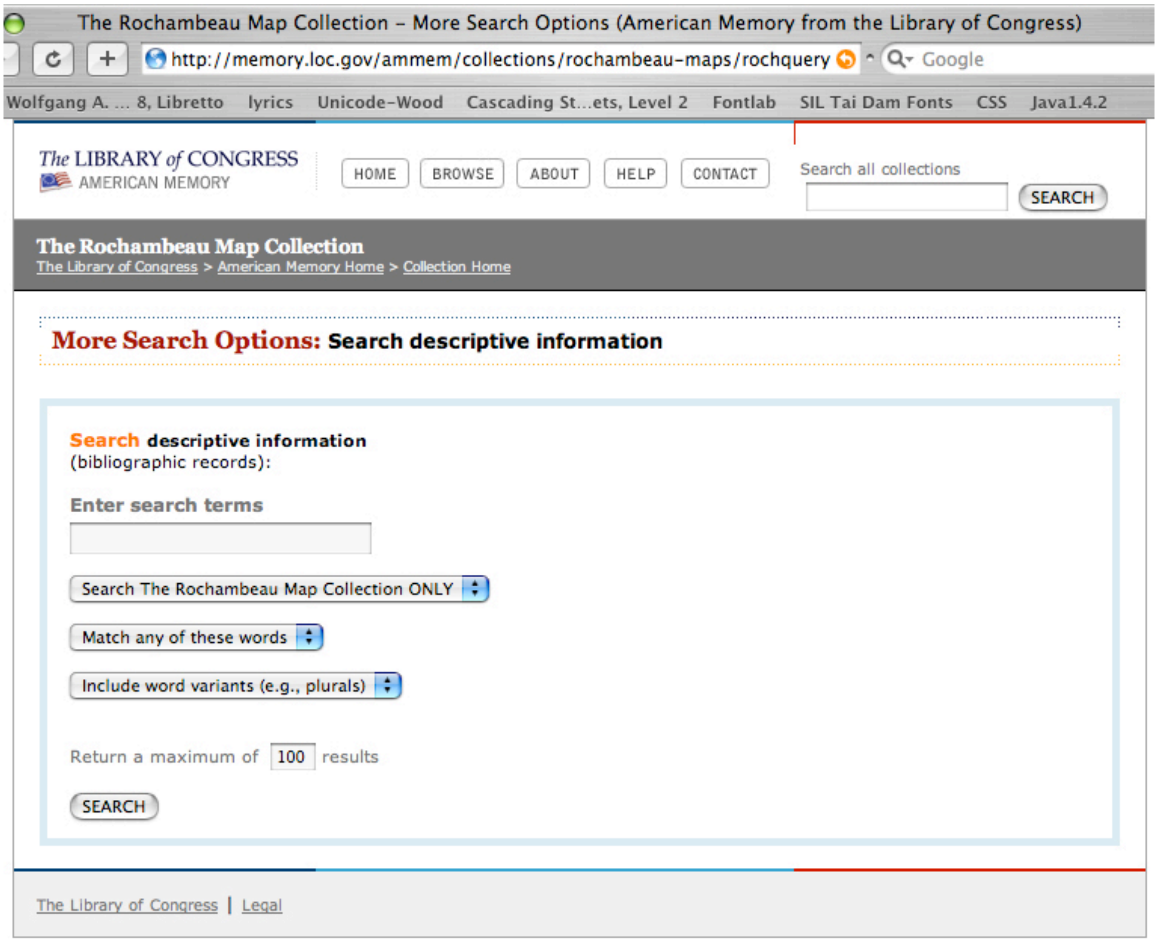
Overview

The *Rochambeau Map Collection* contains cartographic items used by Jean Baptiste Donatien de Vimeur, comte de Rochambeau (1725-1807), when he was commander in chief of the French expeditionary army (1780-82) during the American Revolution. The maps were from Rochambeau's personal collection, cover much of eastern North America, and date from 1717 to 1795. The maps show Revolutionary-era military actions, some of which were published in England and France, and early state maps from the 1790s. Many of the items in this extraordinary group of maps show the importance of cartographic materials in the campaigns of the American Revolution as well as Rochambeau's continuing interest in the new United States.

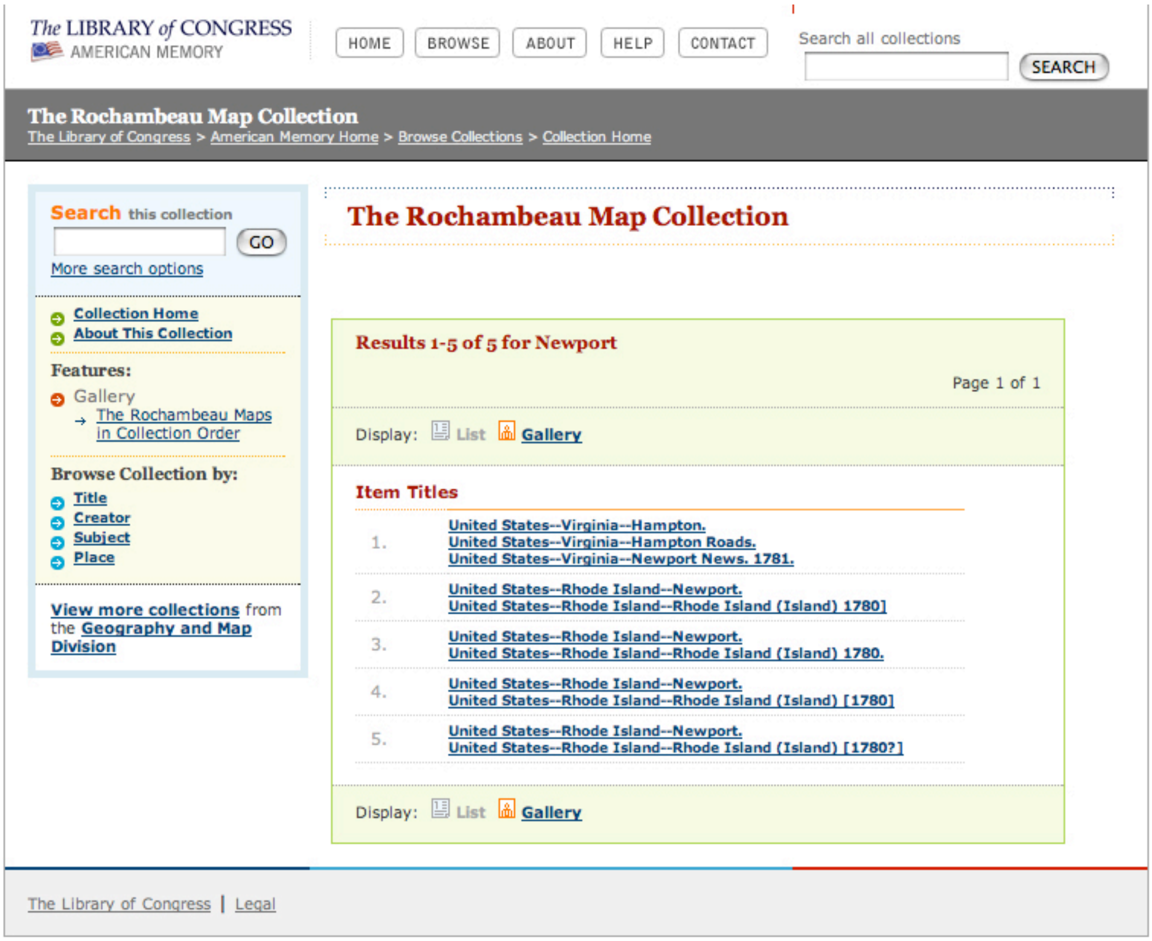
The collection consists of 40 manuscript and 26 printed maps, and a manuscript atlas, the originals of which are in the Library of Congress' Geography and Map Division.

Gallery
Rochambeau by the numbers.

The Library of Congress | Legal



In response to the query term "Newport", the retrieval set offers a "gallery" (for browsing) or ranked items (searching):



Next is a screen shot of the actual item (selected item #4 above):

The Rochambeau Map Collection
The Library of Congress > American Memory Home > Browse Collections > Collection Home

Result 4 of 5 for Newport

Back to Results list Previous Item | Next Item

The Rochambeau Map Collection

Click on picture for larger image, full item, or more versions

Rights and Reproductions

Plan de Rhodes-Island, et position de l'armée françoise a **Newport**.

Created/Published

[1780]

Notes

Scale ca. 1:26,000.
Manuscript, pen-and-ink and watercolor.
Has watermarks.
Relief shown by shading.
Shows fortifications, fields of fire, and position of French naval vessels.
Oriented with north to the right.
Includes index to defenses.
Note on manuscript label originally mounted on verso: A joindrê à lettre de M. de Rochambeau au Prince de Montbarey datée 19 juillet 1780.
Reference: LC Maps of North America, 1750-1789, 1017

Subjects

Newport (R.I.)--History--Revolution, 1775-1783--Maps, Manuscript--Early works to 1800.
Rhode Island (R.I. : Island)--Defenses--Maps, Manuscript--Early works to 1800.
United States--Rhode Island--Newport.
United States--Rhode Island--Rhode Island (Island)

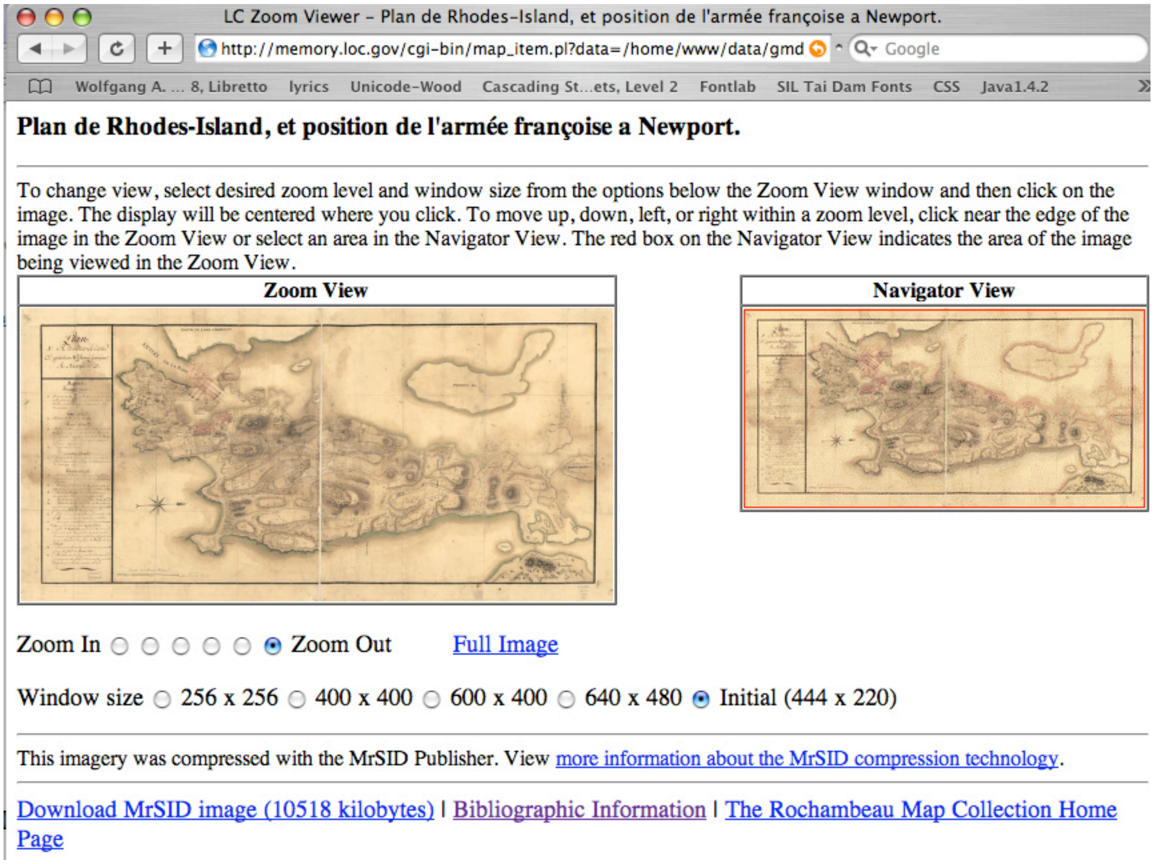
Related Titles

[Rochambeau collection ; 38]

Medium

col. map 53 x 113 cm.

Finally, clicking on the image returns the image with multiple display options and a navigator view window:



Cataloguing Difficulties Automatic recognition methods for photographs are still rather primitive. Unfortunately, manual processing faces its own challenges. One major consideration is the size of the collection. It is believed that there are over 1,000,000 photographs at Harvard that are not catalogued. Some photographs, of course, are duplicates of others. Another liability, although it can be fun to resolve, is that no one may know anything about the photograph: who are the subjects in the photo? Where was it taken? When was it taken? By whom? Why? What other "information" is there in the photo that the cataloguer should identify? Are there clues in the photo that answer the other questions?

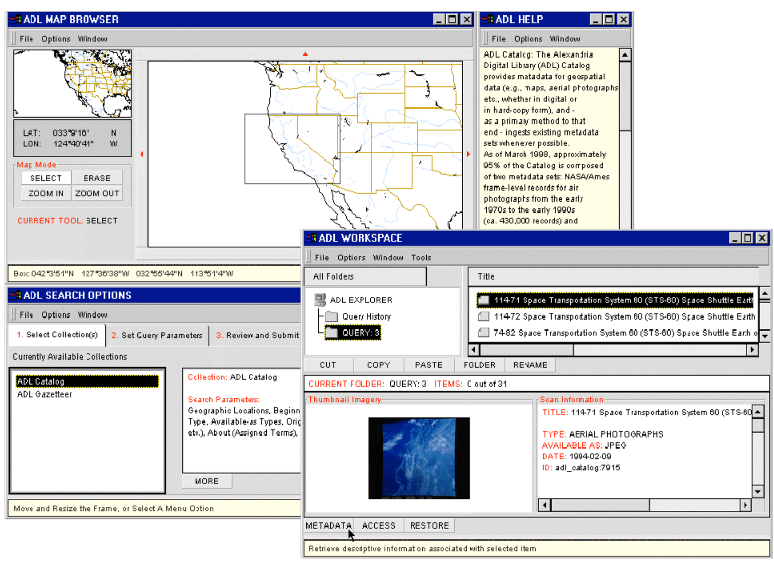
Given the lack of data available to the cataloguer, one can image how the difficulties are compounded for the information seeker. Few visual collections permit searching by facets that make sense to the end user's individual needs but even generalists cannot locate materials by searching. For example, a search of the term "calligraphy" on Harvard's visual collection, returned (July 20, 2005) almost 1,100 hits. Browsing some of the thumbnails revealed *most* images were of manuscripts displaying calligraphy some forms of Arabic, Chinese, Japanese, Korean writing, but some included unexpected results, such as a woman dressed as a geisha in a room that happened to have a small calligraphic artwork on the wall. In short, the end-user must browse.

Browsing for visual materials could be performed "offline" as it were, where the person handles each photograph or slide. This technique can be physically gratifying in a way – handling old objects – but of course is tedious and the original items risks being damaged. Online browsing using thumbnails is tedious, too, and is compounded by the technical issues of bandwidth, server speed, the html page design, scanned image quality and screen resolution.

GEOSPATIAL INFORMATION RETRIEVAL

Geospatial data are especially fascinating because end-users can see the image and (most of the time) download the source data that created the image so that they can use in their own software applications. A noteworthy geospatial project is the NSF-funded Alexandria Digital Library, UC Santa Barbara (Hill, Frew, & Zheng, 1999). This is a collection of terrestrial maps, aerial and satellite photos, astronomical maps, databases, and related textual information stored such that any data can be referenced by these "geographical footprints." This project is unusual, too, because it is both a test collection and a practical implementation in the UCSB map library.

Here is a screen dump of the Alexandria project's user interface:



This project combines the usual digitization and storing of related [here geospatial] information, which creates large files. In addition, there is multi-level decomposition of the image to give course grain and fine grain access. [Uses wavelets; first level is a small coarse image, extra levels provide greater detail.]

As one would suspect, the IR system must confront the user's interface. Small size of the computer displays really diminishes the usefulness of this project and slow performance in delivering the large files causes this project to retain the user's state through a session. [One advantage of the Internet is that is it usually *stateless*.]

In passing, we should note that map librarianship, growing up in isolation and adopting the practices of the field, creates *gazetteers*, a database and set of procedures that translate representations of geospatial references, such as place names, geographic features, coordinates, postal codes, census tracts, and so on. The search engine, then, must be tailored to the peculiarities of searching for place names. [Again, this is a deviation from traditional IR – a kind of default of fielded searching.]

The Alexandria Digital Library (ADL) has its own gazetteer, emphasizing the use of an object's location on Earth and its "geographic footprint." The footprint refers to the latitude and longitude values that represent a point, a bounding box, linear feature, or a complete polygonal boundary. Future research will no doubt use these data for automatic feature extraction using techniques such as *n*-nearest neighbor matching.

The gazetteer also acts as a controlled vocabulary, converting user's queries, such as "Where is Paris" into its location. Then the user can find objects by matching the footprint of a geographic name to the footprints of the collection objects. [Once again, this is merely the matching of tokens from query to collection!]

Through the interface users can select a region and have the system search based on the selection. For instance, a user can draw a box around an area on a map and find the hospitals or orchards or volcanoes in the area! Here is an example from the gazetteer.

Feature name	*State*	*County*	*Type*	*Latitude*	*Longitude*
Tulsa	OK	Tulsa	Pop pl	360914N	0955933W
Tulsa county	OK	Osage	Locale	360958N	0960012W
Tulsa county	OK	Tulsa	Civil	360600N	0955400W
Tulsa Helicop-ters, Inc.	OK	Tulsa	Airport Heliport	360500N	0955205W

One can see how a search of Tulsa returns the latitude and longitude for the city. By creating a bounding box around the area, say a 5 mile radius, one can determine which other elements of interest fit the query. Here one could search for helicopter companies in Tulsa.

Challenges

Map projects that include photographs necessarily inherit the problems to automatic feature extraction.

Bounding boxes can include too much territory. For instance, the bounding box for the State of California includes Nevada.

VIDEO

Unlike the other items discussed, video materials do not necessarily have any metadata and the source, say a TV news script, is not guaranteed to be available. Video poses interesting questions unique to its format. What should we consider as a video collection? What is the unit of analysis: the entire video or sections of it? How long of a section? Of course, some types of collections are given, such as the productions of a single TV show or news station's broadcasts. *Segmentation*, too, is fairly intuitive, such as an individual news cast. But the size of the files and the lack of text and metadata make searching video a long, tedious and resource-devouring task. For instance, imagine having hours of news films to review to find the one two-second segment of interest. Will you have sufficient computing resources: 4,000 hours of video is about 2 terabytes in size!

There are many video retrieval projects, such as the work of Gary Geisler, formerly of Simmons College, now at the University of Texas-Austin, and Informedia Digital Video Library project, lead by Howard Wactlar at Carnegie Mellon University. The researchers in these projects attempt to find techniques for browsing that are not too tiring on the user. For instance, Geisler's project showed test subjects video clips of varying length and speed. Some of the clips move so quickly the reviewers were nauseated. In addition, there are computing requirements that are demanding. MPEG-1 requires 1.2 Mbits/sec processing. Therefore, some kind of useful surrogate is needed along with novel, multi-modal approaches to IR.

Informedia has developed specialized tools for various aspects of image understanding. These include "scene break detection" through segmentation of the input video and selecting "icons" to watch for. Some images can be stored in a database and compared to shapes extracted from the video: this is image similarity matching. In the same way some police departments use video to capture the license plates of people who run red-lights, so video-OCR captures and recognizes text on screen.

Entrepreneurs have captured governments' attention through *face detection* software.

Multimodal IR

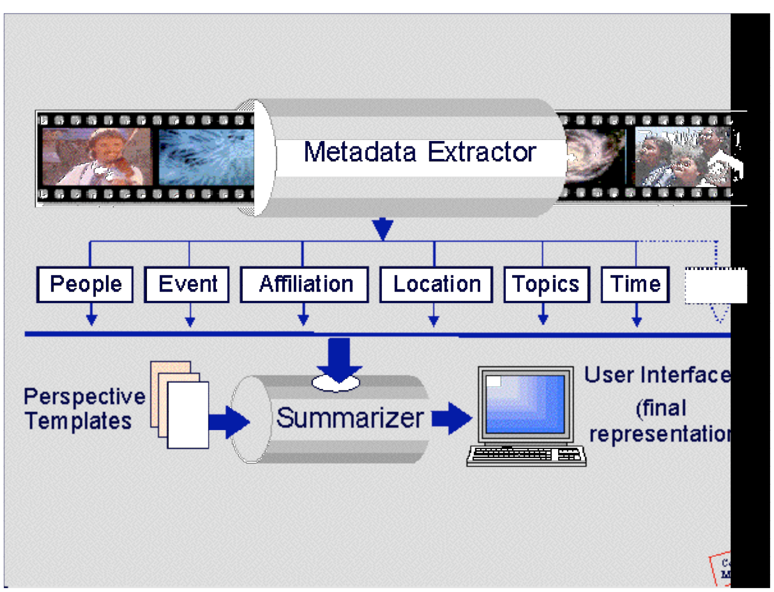
All animals respond to change. Consider the cat on the prowl; it notices only changes in the terrain and changes in sound. After identifying the whole scene, the cat is successful as a hunter because it processes *only changes* from the original scene and then acts accordingly. Similarly, video IR looks for clues inherent in film and video.

Using methods from artificial intelligence (speech recognition, NLP, image recognition), video processing can scan the digitized image quickly looking for changes, such as shifts in color, or anything else that might be useful, and then extract video scenes.

Video and film also have additional data. In film, sound tracks are not part of the visual image itself, but are stored next to image on the film. Closed captioning data, too, are stored outside the image and can be extracted for processing.

One avenue to video IR is to isolate the sources of data into three streams: video, audio, and text. Then applying AI techniques (speech recognition, image extraction, NLP) to create *segments* (equivalent to the retrieval set) and then apply automatically generated or human supplied metadata to the segment.

Having extracted data from the processing software, metadata can be created and a summary (a document in itself) presented to the user (multimodal metadata extraction):



When the resulting representation is presented to the user through the interface, a different kind of interface design comes into play. These include visual icons (one-line headlines, film strip views, video skims, transcripts following audio track), collages of images, semantic zooming, and skimming, among others. The next image shows the query input screen (upper left), the search results (the thumbnails; relevancy is represented by the red bar on the left of the thumb nail), the selected video (upper right on the screen), and finally, accompanying text that (probably) was searched, too.

Thumbnail	A single image that illustrates the content of a video
Filmstrip	A sequence of thumbnails that illustrate the flow of a video segment
Video skim	A short video that summarizes the content of a longer sequence, by combining shorter sequences of video and sound to provide an overview of the full sequence.

To make video IR work at all, there is a bit of preparation before the user's can search. To create a filmstrip, the video sequences are separated into single shots. Use techniques from image recognition applications to identify dramatic changes in the scene on the current shot. Frames with similar color characteristics are assumed to be part of a single shot. For example, imagine a video clip of a summer sky: mostly blue and some clouds. The clouds may move during the video clip – this in informative. The clouds, too, may change color – this is very informative, both in terms of video IR (because there has been a dramatic change in scene) and for the humans who may get rained on. A sample frame is now selected: the default is the middle frame from the shot. Now there's a known time length of the filmstrip. End-users provide feedback based on the point during their viewing of the filmstrip at which they decide it is relevant.

As video collections get larger, the problems of IR increase. One source of problems is technical: the storage and bandwidth requirements great considerably. Also, as a video collection grows, there are far more potential objects in the content which makes it difficult to tune heuristics. Finally, the increase in variety of what could be shown to the user as well as the volume of video skims and filmstrips places a great burden on the browser, similar to still pictures, for browsing.

In short, video IR considers searching and browsing as part of a single user behavior. Designers of video IR must consider *data* (content and metadata), the *computer system* (search engine) and the *user interface* as equal parts. To address incomplete or error-prone data, video IR relies on *multi-modal methods*.

TEXT EXTRACTION:

Some research projects attempt to create written documents (perhaps for further processing as a regular IR text) using speech recognition software, such as Sphinx II and III. These projects work with unrestricted vocabularies – the speaker says whatever he or she wants to and each speaker is independent of other speakers. This means one can't apply Bayes's Theorem but suggests strongly Markov models. [See Nanette Veuilleux's work.] The error rate is higher than 25%.

Closed captions are digitally encoded text. The intention is to assist deaf and hard-of-hearing people who watch TV shows. Closed captioning is done in real-time so there are a lot of typos, making the resulting extracted text rather inaccurate. Moreover, not all video include closed-captioning.

Text extraction is also applied to queries. This is the same technique applied to source (sound) documents.

References: Hill, L. L., Frew, J., & Zheng, Q. (1999, January). Geographic names: the implementation of a gazetteer in a georeferenced digital library. *D-Lib, 5*(1). http://www.dlig.org/dlib/january99/hill/01hill.html.