

9. Retrieval Using Controlled Vocabularies

Vocabulary resources, with their synonyms, hierarchical structures, and other conceptual relationships, can provide extremely powerful tools for retrieval across disparate data resources residing in different places and even in different languages, enabling users to obtain meaningful results in their online searches. It only remains for those whose mission is to deliver high-quality information to tap the vocabularies' immense potential.

Of all the information in a catalog record for an art object, the fields for the names of people, places, and things are the most obvious targets where vocabularies should be used in retrieval. Terms and names used to index art and cultural heritage information can vary widely, even when the same concept is being referenced.

In retrieval, users do not always know what a person, place, or thing is called. Nonexpert users often do not know the term used by a specialist to index a work. For example, an expert would call a particular vessel a *rhyton*, but a nonexpert would call it a *drinking horn* or even a generic *vessel* (if they did not know the purpose of the vessel). A controlled vocabulary allows such users to browse or search for data using familiar terms or other criteria in order to discover relevant information. Expert users will know the specialized terms for a work, but different specialists may use different terms to refer to the same person, place, or thing. Thus, no matter who the user is, vocabularies are critical in gathering these equivalent terms, relationships, and other information together and using them to launch searches across disparate data sets or even within a single database.

9.1. Identifying the Focus of Retrieval

The discussion of retrieval with and of controlled vocabularies covers two activities: retrieval using vocabularies versus retrieval of vocabulary terms themselves.

The primary end-user activity is retrieving work records or other content objects using vocabularies. In this activity, the user searches across content objects, often by typing a search term. The vocabulary is used, often behind the scenes—for example, to broaden the search by adding synonyms to the query. How to employ vocabularies to broaden retrieval

and how to display content objects as results to end users are topics that have been extensively studied and written about for several decades.

However, another activity often involves searching a controlled vocabulary itself. In this case, a user first approaches a controlled vocabulary in order to locate desired terms for use in searching or indexing. This activity involves retrieving the vocabulary records themselves, with the goal of either finding vocabulary records as an end or using the retrieved vocabulary records to in turn retrieve or index content objects. How to retrieve and display the controlled vocabularies themselves is a field of study in itself, discussed in ISO, NISO, and other thesaurus standards.

Given that these two activities are so intimately connected and overlapping, they are discussed together in this chapter. Issues surrounding interoperability of multiple vocabularies in retrieval are discussed in **Chapter 5: Using Multiple Vocabularies**.

9.2. User Intervention or Behind the Scenes

How end users will conduct searches using vocabularies is an important issue. The end users may be guided in their searches by showing them the expert terminology from which to choose, in a process known as *user intervention* or *mediation*. If searchers are offered nearest equivalents for their search term and, preferably, also broader and narrower terms, they can choose those that best match their retrieval requirements.

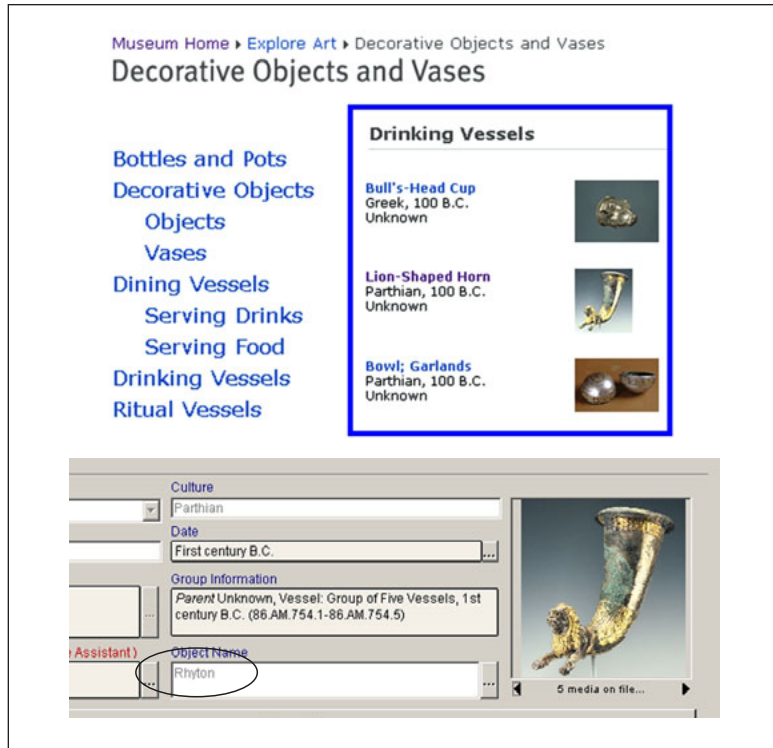
Another approach is to apply vocabulary terms to a user's query entirely *behind the scenes*, with no overt user intervention. In interfaces where users are the general public, this approach is often likely to be less confusing and more satisfying to most users. However, this approach limits the ability of the user to control the search criteria, which can be frustrating to more technically sophisticated users and content experts.

Ideally, a vocabulary designed specifically for retrieval (separate from the indexing vocabulary) accommodates nonexpert searches. End users should be provided with a vocabulary designed specifically for nonexperts, which is linked to the specialist vocabulary that was used for indexing. In the example on the opposite page, users are presented with a short browsing list of nonexpert terms, which allow access to records that were indexed with expert terminology.

9.2.1. Retrieval by Browsing

In online retrieval, *browsing* refers to the activity of looking through various entries to make a selection, such as a list of terms or hypertext links. Browsing should allow users to follow links on a Web page and explore the content as if they were scanning titles on the shelves of a library or thumbing through an encyclopedia. Entries may be arranged

Fig. 60. Examples of an end-user browser display and the editing screen for a rhyton in the J. Paul Getty Museum. In the upper image, nonexpert terms such as *Bottles and Pots* provide access to the general public. In the lower image, what the public has seen as a *Lion-Shaped Horn* is indexed with the expert vocabulary term *rhyton* in the cataloging database.



by alphabetical lists, short pull-down lists, or in other arrangements. In the example on the following page, both pull-down lists and a more extensive alphabetical display are provided.

The lists of terms in a browsing interface and their organization may be derived from the indexing terms that have been used to catalog the works or other content objects. With browsing, retrieval is generally not accomplished by variant names; authorized terms or names must be located in the provided lists. Occasionally, such lists may have *see references*, but typically they do not. If users do not know how to spell the name, they will have difficulty in finding the content they seek. Thus many art-information sites allow retrieval via search boxes in addition to via browsing. The most successful use of browsing allows discovery by users who wish to have the broadest of overviews of the collection, generally helpful to those who do not know enough about the content to search for particular artists or works.

The term *browsing* may also refer to other examples of lists in a system or on the Web, as where users scan a results lists for the desired content or move through a hierarchical display scanning for appropriate terms.

Explore Art

Learn more about many of the works of art on display at the Getty Center and the Getty Villa.

Artists
Browse artists by name:

A B C D E F G H I J K L M N
O P Q R S T U V W X Y Z

Types of Art
Browse by object type or medium:

Select a Type of Art

- Select a Type of Art
- Architecture and Room Elements
- Decorative Objects and Vases
- Drawings
- Furniture
- Implements and Costume
- Manuscripts
- Paintings
- Photographs
- Sculpture

Browse Artist by Name: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

1 - 10 of 122 Artists [Next ▶](#)









Babbitt active: 1853 - 1870 American		Bacchiacca b. 1494 d. 1557 Italian	
Baciccio b. 1639 d. 1709 Italian		Backer b. 1608 d. 1651 Dutch	
Bagnacavallo b. 1494 d. 1542 Italian		Bailleul le jeune active: 1740 - 1750 French	
Baillie b. 1723 d. 1818 British		Bailly b. 1584 d. 1657 Dutch	

Fig. 61. Examples of browser displays from the J. Paul Getty Museum Web site. The upper image illustrates how users may browse alphabetical lists and pull-down lists. The second image illustrates a detail of the alphabetical listing for artist names beginning with *B*.

9.2.2. Retrieval via Search Box

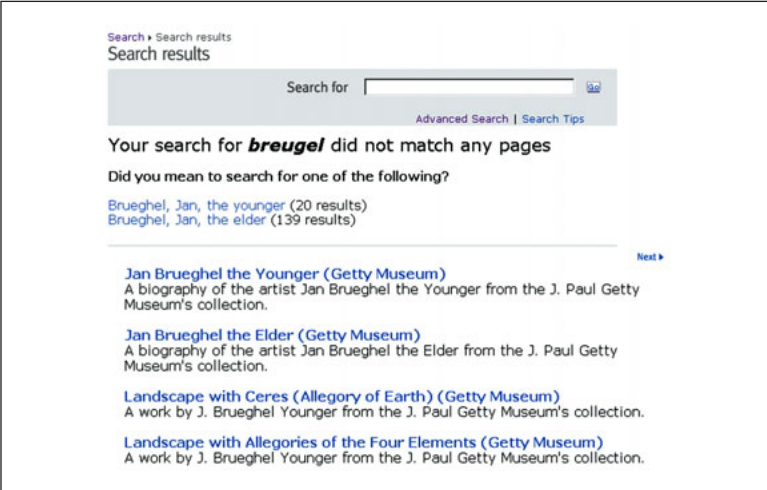
A search box is a field or other method by which users can enter terms and compose searches. When they search for a term, they expect to retrieve all occurrences of the term (and its synonyms) throughout the database or site. Ideally, the searching interface would use a vocabulary behind the scenes to provide users with alternate terminology choices when their search is unsuccessful or the results are ambiguous.

In the example below, the user typed a term that retrieved no results on the pages searched; however, the user's term was found in a vocabulary (the *ULAN*) used for retrieval behind the scenes. Based on the matches in the vocabulary, the user is offered two choices for terminology that meet the criteria and bring back results on the site.


Ideally, the search engine will not offer the user choices of terms from the vocabulary that would retrieve zero results from the target data being searched (called *blind references*). In the example below, the term *breugel* was actually found in several vocabulary records, but only two of the artists were represented on the target site; therefore, the vocabulary terms that would have brought back no hits were suppressed from the user in this view.

If the search term does not match any preferred or variant term in the vocabulary, the system could offer the user additional options by displaying the terms that are alphabetically close to the search term. For example, the system could display a list of terms from the vocabularies that alphabetically would precede and follow the search term entered by the user, as is common in online dictionaries.

Fig. 62. Example of a search box that allows minimal user interaction. The retrieval results are displayed along with a prompt allowing the user to narrow the results by one of the two available artists named *Jan Brueghel* on this site.



Search > Search results
Search results

Search for 

[Advanced Search](#) | [Search Tips](#)

Your search for **breugel** did not match any pages

Did you mean to search for one of the following?

[Brueghel, Jan, the younger \(20 results\)](#)
[Brueghel, Jan, the elder \(139 results\)](#)

[Next >](#)

[Jan Brueghel the Younger \(Getty Museum\)](#)
A biography of the artist Jan Brueghel the Younger from the J. Paul Getty Museum's collection.

[Jan Brueghel the Elder \(Getty Museum\)](#)
A biography of the artist Jan Brueghel the Elder from the J. Paul Getty Museum's collection.

[Landscape with Ceres \(Allegory of Earth\) \(Getty Museum\)](#)
A work by J. Brueghel Younger from the J. Paul Getty Museum's collection.

[Landscape with Allegories of the Four Elements \(Getty Museum\)](#)
A work by J. Brueghel Younger from the J. Paul Getty Museum's collection.

9.2.3. Retrieval by Querying in a Database

Which is better for the users, a simple search or an expanded query? Users may be offered the choice of a simple search across the entire data set or a fielded query, which is a search on individual fields in a database. In the example of a simple search on the previous page, the search terms were gathered from the vocabulary and used across all Web pages available on the site. This approach provides a simple searching interface typically suitable as a default search option for most members of the general public. They receive the benefit of vocabulary-assisted searching without having to worry about the difference between types of information; they may search for the artist's name in the same search box as they would for the medium of the work.

However, a more technically sophisticated user or a content expert will probably be dissatisfied with such a broad search. An alternative method is to use vocabulary to search individual fields of a database. The terminology made available for each field should be appropriate for that field (e.g., fields for artists' names should be linked to vocabulary for artists' names, fields for materials should be linked to vocabularies appropriate for materials, etc.). The results of searching data fields is more accurate and more precise than with a simple search across all content.

In the example below, pull-down lists for some fields are combined with search boxes that allow users to type in the artist's name and the title of the work. The *artist's last name* search box is linked to a

Fig. 63. Example of an expanded query form for the National Gallery of Art Web site, which allows the user to choose values for several specific fields, including artist, title, and medium.

Expanded Search

Search the National Gallery's collection by the artist's last name, key words in the title, school, style, date, medium, and/or short list of popular subjects. It is not necessary to fill all of the boxes. Type search words or select terms from pull-down menus, then click on the **Search** button. [Search Tips](#)

Artist's Last Name
example: monet | [index of artists](#)

Key Words in Title
example: rouen cathedral

School

Style

Year Created from: to:

Medium

All Media
 Decorative Art
 Drawing
 Media Art
 Painting
 Portfolio
 Photograph
 Print
 Sculpture
 Technical Material

Choose multiple options by holding down the "Control" (PC) or

Name Authority, allowing the user to access works by that artist via his or her preferred name or any variant name.

Search boxes for retrieval are also utilized in systems used for cataloging artworks. Catalogers must be able to retrieve sets of work records for editing, comparison, and other purposes. In a search for work records, the cataloging system (typically a *collection management system*) should allow catalogers to incorporate variant terms and hierarchical

Fig. 64. A search screen and results from the J. Paul Getty Museum cataloging system illustrating how a query may be formulated using equivalents and narrower terms from a thesaurus. The query for works with the medium *metal* returned a work record with the medium *bronze*, for which *metal* is the broader context in the thesaurus.

The figure consists of two screenshots from the J. Paul Getty Museum cataloging system.

The top screenshot is a "Query Assistant" dialog box. It prompts the user to "Please enter the term to perform your search." The search term "metal" is entered in the text box. Below the text box are three checkboxes: "Include Thesaurus equivalents" (checked), "Include narrower terms" (checked), and "Proximity search" (unchecked). There are two groups of radio buttons for logical operators: "At least one word (logical OR)" (selected) and "ALL words (logical AND)". There are also two radio buttons for search matching: "Match using exact term search" (selected) and "Match using approximate term search". A link "Here is your query so far:" is visible. At the bottom are buttons for "< Back", "Next >", "Help", "Clear", and "Cancel". The title "Query Assistant" is at the bottom center.

The bottom screenshot is the main "Objects" interface. The title bar says "Objects". The menu bar includes File, Edit, Display Mode, Query, Report, Registration, Related, Tools, Maintenance, and Help. The status bar shows "Curator Approved", "Public Access" (checked), "On View", "Accountability" (checked), and "Virtual Object". The main content area shows "The Museum System" with "NUMBER CHECKED Antiquities" and "55.AB.12". The record details include:

- Classification:** Sculpture (dropdown), Culture: Etruscan
- Period:** Date: about 480 B.C.
- Constituent(s):** Artist: Unknown
- Title:** (3 titles, see Title Assistant); Statuette of a Bearded Man, Probably Tinia
- Object Name:** Mythological Figure
- Medium:** Bronze (circled in red)
- Dimensions:** Object: H: 17.2 x W: 9 x D: 4 cm (6 3/4 x 3 9/16 x 1 9/16 in.)
- Description (Object):** Bronze statuette of Etruscan God Tinia (Latin: Jupiter, Greek: ...)
- Current Location:** 55.AB.12: JPG V MUSEUM, 202 [Villa, Museum, Gallery 202, Changing Exhibitions] [Jul-09-2009], Display

 A small image of the statuette is shown on the right. At the bottom are tabs for "Front Card", "Notes", "Documentation", "Media", "Context", "Related", and "Other".

relationships from the controlled vocabularies. In the example on the previous page, the collection management system gives users the option of including terms from the thesaurus and narrower concepts in the query.

Search boxes may be combined with the ability to truncate terms or to add Boolean operators and other facilities to allow users to make versatile and powerful searches. In the example below, an advanced search interface allows term truncation, Boolean operators, and searching for ranges of dates.

Fig. 65. An advanced search screen formerly used on the Metropolitan Museum of Art Web site illustrating how the user may narrow the query to a particular database (the collection or the provenance research project) as well as query an extensive list of fields, including artist, title, country of origin, and others. The Boolean operators AND and OR are available. The Metropolitan Museum of Art, www.metmuseum.org. Copyright © 2000–2009 The Metropolitan Museum of Art. All rights reserved.

The screenshot shows an advanced search interface with a yellow background. It includes the following sections:

- Area of Search:**
 - Entire Site or select one or more
 - Collection Database
 - Provenance Research Project
- Artist or Maker:**
 - Enter a last name, full name, or partial name (such as Brueghel, Jan Brueghel, or Brueg*):
 - Text input field with an AND dropdown menu.
- Title of Work or Type of Object:**
 - Enter a word, partial word, or phrase (such as dancer, danc*, dance class, vessel, or ritual and vessel):
 - Text input field with an AND dropdown menu.
- Country or Culture of Origin:**
 - Enter the name or partial name of a country, city, or culture (such as Mexico, Mexic*, or Olmec):
 - Text input field with an AND dropdown menu.
- Medium or Materials:**
 - Enter a word, partial word, or phrase:
 - Text input field with an AND dropdown menu.
- Date of Creation:**
 - Enter a date range. For a single year, enter a number in the first box only:
 - Between [] C.E. and [] C.E.
- Credit Line or Reproduction Rights:**
 - Enter a word, partial word, or phrase:
 - Text input field with an AND dropdown menu.
- Description:**
 - Enter a word, partial word, or phrase:
 - Text input field with an AND dropdown menu.
- Provenance:**
 - Enter a word, partial word, or phrase:
 - Text input field with an AND dropdown menu.
- Inscription or Marks:**
 - Enter a word, partial word, or phrase:
 - Text input field with an AND dropdown menu.
- Accession Number:**
 - Text input field with an AND dropdown menu.
- Keyword:**
 - Text input field.

At the bottom left, there is a search button with a magnifying glass icon.

9.2.3.1. Reports and Ad Hoc Queries of the Database

For maintainers of the vocabulary data and other authorized advanced users, predefined reports must be supplied and ad hoc queries on the database must be allowed. A *predefined report* is a query and a format for output that has been written in advance and is used for queries that are asked frequently. This type of report may or may not have variables that can be changed by the user. An *ad hoc query* allows a qualified user to use a query language to access all of the underlying data tables without going through a user interface that limits access to only certain fields and predefined query logic. In the examples below, users construct queries targeting various data tables and columns in a relational database.

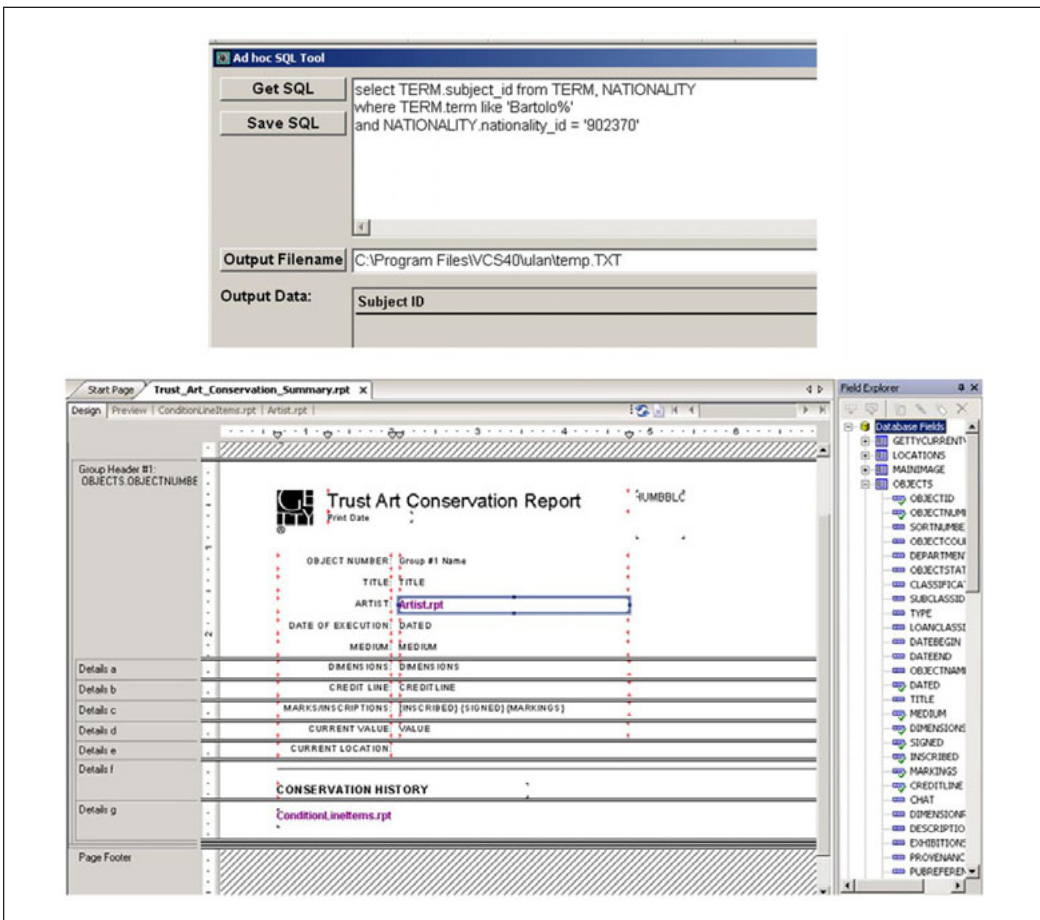


Fig. 66. Examples of query screens for constructing reports in VCS (the Getty Vocabulary Program's editorial system) and TMS (The Museum System). The upper image illustrates an ad hoc query in VCS, where the editor has constructed an SQL (Structured Query Language) query based on values in the relational database of vocabulary data. The lower image illustrates a query under construction in a report writer in TMS.

9.2.4. Querying across Multiple Databases

When querying across multiple databases, implementers must resolve several issues having to do with target data and vocabularies. Data located on the surface Web (or *visible Web*), even if it is derived from multiple databases, may be rendered accessible to local search engines and public retrieval tools, including Google. However, other data can be hard to retrieve in concert across multiple databases. The various databases may be located at different institutions or even at the same institution, but they may be on different servers, different platforms, they may have different interfaces, and the data fields, rules, and data values may not be compatible. Such data may be visible on the Web in certain views, but if it is located in the deep Web (or *invisible Web*), the information is hidden or generally inaccessible through traditional search methods.

As a first step in resolving these issues, disparate databases must be mapped to each other or to a separate standard set of fields. In addition, deep Web data generally must be made accessible to a common search engine, either by copying all the data onto a common location or somehow making the data available from its native environments. If both of these criteria are met, controlled vocabularies may be applied during searching to minimize retrieval problems caused by the original data having been cataloged using different vocabularies. For issues surrounding the use of multiple vocabularies for retrieval, see **Chapter 5: Using Multiple Vocabularies**. For a set of fields for exchanging work data, see the *CDWA Lite* XML schema.

9.2.5. Seeding Tags with Vocabulary Terms

Another way in which vocabularies can improve retrieval of Web content is by seeding *meta tags* located in the source code of a Web page with synonyms and broader contexts. HTML (Hypertext Markup Language) is a programming and markup language used to create documents for display on the World Wide Web. These Web documents are presented in a specific tagging language, where the data values, formatting, and other information necessary to display the page appear between opening and closing tags in angled brackets.

In the example below, variant names for an artist have been taken from a vocabulary and added to the keywords for a Web page. This allows the page to be retrieved by search engines by any of the variant names for this artist.

```
<META NAME="keywords" CONTENT="El Lissitzky,
Lissitsky, Lisickij, Lisitski, Lisitskii, Lisitsky, Lissickij, avant-
```

garde art, avantgarde, book design, Yiddish book design, children's books, Futurism, Futurist Art, Modernism, Modernist, Modernists, Proun, Russian art, Soviet art">

9.3. Processing Vocabulary Data for Retrieval

Retrieval of vocabularies must accommodate the special needs of the vocabulary data; it should not necessarily be limited by the functionality of off-the-shelf software and standard search algorithms. Efficient retrieval of vocabulary terms and names requires processing and algorithms suited to the unique characteristics of the data, which is unlike natural language. Vocabulary data includes proper names, generic terms, compound terms, historical terms, term inversions, and variations representing all possible languages. Standard searching methods are optimized for uncontrolled free text and often do not work well with terminology from a controlled vocabulary. The methods discussed in this chapter are primarily intended for thesauri. For a discussion of other types of vocabularies that may be optimized for retrieval, including synonym ring lists and ontologies, see **Chapter 2: What Are Controlled Vocabularies?**

As discussed earlier, the requirements of vocabularies intended for indexing generally differ from those of vocabularies intended for retrieval. A vocabulary for indexing focuses on warrant, correct usage, and authorized spellings of terms, while a vocabulary for retrieval allows less strict parameters to accomplish broader retrieval. However, in many institutions, the same vocabularies must be used for both purposes. This issue may be largely resolved by processing or preprocessing the indexing vocabulary for optimal use in retrieval.

In this context, *data preprocessing* may refer to any type of processing performed on data to prepare it for a processing procedure different from that for which it was originally compiled. Preprocessing of vocabulary terms translates the data into a format that is more easily and effectively processed for the use of the search engine or end-user displays.

Terms and other data may be preprocessed and stored in indexes or tables specific to the retrieval application, or the terms and data may be processed for retrieval as needed on the fly. For large and complex data sets, it is typically more efficient to store the preprocessed terms and other data, rather than constructing them on the fly. For example, data in a complex relational database designed for an editorial system could be packaged so that it could be displayed faster and more easily on the Web for end users. The packaging could include the precoordination of parent

strings from the hierarchical structure, preconcatenating them so they do not need to be constructed on the fly in the Web interface for end users.

9.3.1. Know Your Audience

Defining the user audience is critical to most issues discussed in this book, but it is particularly relevant in the context of retrieving, processing, and sorting names or terms. In this book, the audience is assumed to be an international audience familiar with English, which is the common default language of the computing community and the Web. It is necessary to have a default language because the vocabularies discussed here are often multilingual; thus one language must be preferred as a base language because it is impractical to have dozens of alternative sets of rules in a single vocabulary to deal with all possible languages. The scenarios and rules discussed here are generic, intended for a multilingual vocabulary accessible to an international audience.

However, if the audience is restricted to a particular location and language, and if it is certain that the data will never be shared with the broader user community, rules for normalization, processing, and sorting terms may differ from those described here. For example, if a vocabulary contains only German terms and the audience is and always will be restricted to speakers of German, then rules can be established that are applicable specifically to the German alphabet, keyboards, etc. Characteristics of Unicode and other issues are discussed below.

9.3.2. Using Names for Retrieval

Although the hierarchy and other information in a vocabulary record may sometimes be used in queries, querying by name or term is the method used most often to access records in a controlled vocabulary. Basic access by all terms or names for a given vocabulary record is critical. The main purpose of adding variant names and synonyms is to allow access to the vocabulary data by any linked term. Any retrieval system should search for any and all variant terms and names for the person, place, thing, or concept. In the example below, if the user searches for *ushabti* (small ancient Egyptian funerary figures), all work records, pages, or other content objects in the target database that contain *shawtabys* and the other synonyms should also be retrieved.

ushabti (preferred, descriptor)

ushabtis (used for term)

shabti (used for term)

shawabti (used for term)

shawtaby (used for term)

shawtabys (*used for term*)

ushabtiu (*used for term*)

ushabty (*used for term*)

ushabtys (*used for term*)

Access must be allowed through official and correct names as well as through nicknames, pseudonyms, and other unofficial names. These names will probably be included in the authoritative vocabulary used for indexing. For example, the twentieth-century architect *Charles Édouard Jeanneret-Gris* was known by his pseudonym *Le Corbusier*; both names should be included in a vocabulary record for this artist. Even common misspellings may be included in the indexing vocabulary to improve access, particularly when these misspellings are published. For example, the twentieth-century painter *Georgia O'Keeffe* is frequently if incorrectly listed as *O'Keefe* (with only one *f*). This common published misspelling should be included in the vocabulary and used to aid retrieval.

On the retrieval side, consideration should also be given to additional misspellings and name variations, even if they are not found in a published source. These misspellings would not be appropriate for the authoritative indexing vocabulary, but they should be used behind the scenes for retrieval, hidden from the end user to prevent confusion. For example, by using a hidden index or other method, it may be advantageous to allow end users who enter *Richard Meyer* to retrieve information about the contemporary architect *Richard Meier*, even though the users have misspelled his name.

9.3.3. Truncating Names

Users should be able to access terms and names by *truncation*: truncation involves the user employing a wildcard symbol (often an asterisk, question mark, or percent sign, or another method) to search for a string of characters regardless of what other characters follow (or sometimes, precede) that string. *Right-hand* truncation is used to match terms starting with the same letters; for example, searching for *arch** retrieves *arch*, *arches*, *architrave*, *architecture*, *architectural history*, etc.

For names and terms, querying must allow, at minimum, right-hand truncation on strings and on key words. Truncation should be allowed in combination with Boolean operators, as in the following example.

Edinburg*

Jan Cornel*

gar* AND eldon

The employment of the wildcard symbol at the middle or left-hand side of the string is helpful as well, allowing retrieval when exact spelling is unknown. However, due to the impact on processing, left-hand and middle character truncation are often impractical when querying large sets of terms.

Pyeitawinzu Myanm* Nain*
*durrahim

9.3.4. Keyword Searching


Keyword searching is a method of computer searching ultimately based on natural language texts rather than controlled vocabulary; however, it should be adapted to searching vocabularies. Keyword searching refers to searching for individual words or combinations of words; this is useful for searching vocabularies that may contain names and terms comprising multiple words. In standard retrieval, keywords are often determined on the fly during a search; however, creating indexes that contain normalized keywords and other normalized strings is a recommended strategy for vocabulary data (see also 9.3.5. **Normalizing Terms**).

Electronic controlled vocabularies should provide keyword access to all words of all the terms in the vocabulary. Keyword searching

Fig. 67. Example of a results list for a keyword search on *buttresses* in the AAT.

Find Name: **buttresses**
Logic:
Note: 8 results

[View Selected Records](#) [Select All Records](#) [Clear All](#) First Previous Page: 1 Next Last

Click the  icon to view the hierarchy.
Check boxes to view multiple records at once.







1.  **angle buttresses**
(<corner buttresses>, buttresses, ... Components (Hierarchy Name)) [300000901]
buttresses, angle
2.  **buttresses**
(<supporting and resisting elements>, <structural elements>, ... Components (Hierarchy Name)) [300000891]
3.  **clasp buttresses**
(<corner buttresses>, buttresses, ... Components (Hierarchy Name)) [300000902]
buttresses, clasp
4.  **<corner buttresses>**
(buttresses, <supporting and resisting elements>, ... Components (Hierarchy Name)) [300000900]
5.  **diagonal buttresses**
(<corner buttresses>, buttresses, ... Components (Hierarchy Name)) [300000903]
buttresses, diagonal
6.  **flying buttresses**
(buttresses, <supporting and resisting elements>, ... Components (Hierarchy Name)) [300000905]
buttresses, flying
buttresses, arched



Fig. 68. Example of the results for an exact string match on the term *window* in the *AAT*. In this application, the user has placed quotes around *window* to search for the exact term rather than for keywords; searching for the keyword *window* would have returned over ninety results.

thus serves the same purpose as the permuted and rotated indexes that are common in print formats.

The process of keyword searching typically uses spaces and punctuation between words to determine which elements of a term are separate words. For the term *flying buttresses*, the space would be used to identify *flying* and *buttresses* as separate keywords. If a user searched for the keyword *buttresses*, this term and any others with the word *buttresses* would be returned.

While keyword searching is useful as a default search strategy for end users, the user must be able to search for the full normalized string instead of the keywords when necessary. A common way of designating the string as opposed to keywords in searches is to enclose the string in quotes (e.g., "*flying buttresses*").

For example, it must be possible to find the exact term *window* without returning the dozens of other terms that have *window* as a keyword.

9.3.5. Normalizing Terms

This section discusses the normalization of terms in the context of vocabulary retrieval. This differs from *database normalization*, which is the process of organizing data in a database by reducing a complex data structure into its simplest structure, creating tables, establishing relationships between tables based on set rules, eliminating data redundancy, and converting Unicode text into a standardized form, among other things.

In the context of this book, *normalizing terms* refers to the process of removing or ignoring spaces, punctuation, diacritics, and case sensitivity in terms. The purpose of such normalization is to allow comparison of terms on the basic character strings, regardless of minor or superficial differences.

Data storage and searching methods typically differ between the editorial system used to create and maintain the vocabulary

and the system optimized for end-user access. Maintainers and creators of the vocabulary data need to search by normalized terms, but they also require the option of searching for an exact match on the full name string, with diacritics, punctuation, and capitalization remaining intact. However, searches for normalized strings and keywords are typically the preferred and only methods required for indexers and end users.

Data should be stored in a way that allows its translation into other encoding schemes. One way to match on normalized terms is to establish normalization routines or create automated indexes of normalized terms. Normalization should be done on the user's query string, on the terms and names in the target vocabulary, and possibly in the database or Web pages being queried. In the example below, the terms have been normalized to all capitals, although normalizing to all lowercase would work equally well.

Name: Atakora, Chaîne de l'
Normalized string: ATAKORACHAINEDEL

Name: Carlos María de Borbón
Normalized string: CARLOSMARIADEBORBON

Term: Ayios Onouphrios ware
Normalized string: AYIOSONOUPHRIOSWARE

The suggestions in this section refer to the preprocessing and normalization of terms in an index to be used for retrieval; both normalized keywords and normalized strings would be stored together for use in searching. In the example below, the name *d'Or, Castel* has been normalized to create six separate entries in the index. The methods used to create these entries are discussed below.

The following are normalized keywords and strings for *d'Or, Castel*:

DORCASTEL
 D
 OR
 CASTEL
 DOR
 CASTELDOR

9.3.5.1. Case Insensitivity in Retrieval

A retrieval system should accommodate end-user queries, no matter what case they use. For example, if an end user searches for *Bartolo Di Fredi* or *BARTOLO DI FREDI*, he or she should retrieve records containing the name *Bartolo di Fredi*.

9.3.5.2. Compound Terms and Names in Retrieval

A retrieval system should accommodate compound terms and names that may be spelled with or without a space. For example, an end user's search for *Le Duc* should retrieve records for both *Charles Leduc* and *Johan le Duc*; a search for *Westwood* should retrieve the record for *West Wood*.

9.3.5.3. Diacritics and Punctuation in Retrieval

A retrieval system should accommodate both the end user's *use* of diacritics and punctuation and his or her *omission* of diacritics and punctuation. For example, if the end user searches for *Jean Simeon Chardin* (without the hyphen and diacritic), he or she should retrieve records containing the name *Jean-Siméon Chardin*.

Given that end users may use a variety of codes or alphabets in searching, and given that most users expect results to sort in a particular way (ignoring diacritics), diacritics should be stripped or mapped to normalized strings in order to achieve adequate retrieval and satisfactory sorting in the results.

A user may type a search string containing a character encoding set different from the one used by the native vocabulary data, or with the diacritics stripped (e.g., typing an *o* when the character in the data against which the query is run contains an *o* with circumflex, *ô*). One way to allow retrieval is to map diacritics or Unicode characters to the corresponding nondiacritic ASCII character, no matter which diacritics are typed by the user.

Possible keyword queries from users:

Amazônica
 Amazónica
 amazonica

Keyword values in the vocabulary data base:

Amazónica
 Amazônica

Keywords in the normalized table omit the diacritics:

AMAZONICA
 AMAZONICA

Terms retrieved in the query:

Región Amazónica
 Amazônica Brasileira
 Amazônica, Região
 Hoya Amazónica
 Bacia Amazônica

Various issues surround the retrieval and display of diacritics, particularly those outside the Latin 1 character set. More and more art institutions are using Unicode, which is a set of codes for diacritics and characters in various alphabets. The Unicode Standard is maintained by the Unicode Consortium in cooperation with the World Wide Web Consortium (W3C) and ISO, the latter of which controls the character set defined in *ISO/IEC 10646:2003: Information Technology—Universal Multiple-Octet Coded Character Set (UCS)*.

Outstanding issues include the following: Unicode is still an evolving standard subject to occasional changes in encoding and protocol of usage. In addition, some art institutions are still using technology that cannot accommodate Unicode, meaning their data needs to be mapped to the Unicode character set in a common environment of data sharing. Furthermore, using Unicode in a multilingual environment presents challenges simply because most systems will expect to be told to use one particular language, not many languages simultaneously.

It is not necessary to store data in Unicode. However, it is very important that data is stored in a way that allows it to be translated to UTF-8 (8-bit UCS/Unicode Transformation Format) or any other relevant encoding scheme.

9.3.5.4. Phonetic Matching

Phonetic matching involves retrieval based on the matching of two words that presumably sound alike. It is common in many search engines. However, rather than using standard phonetic matching for art terminology, specialized normalization and searching algorithms are recommended. Although standard phonetic matching is not very useful for art information, it is discussed here so that readers will understand what it is and why it does not work well in multilingual controlled vocabularies.

A *phonetic algorithm* is an algorithm used to index words by their pronunciation. Words with the presumed same pronunciation are encoded to the same code or string so that they can presumably be matched despite minor differences in spelling. Among the best known of the dozens of such phonetic algorithms are Soundex and Metaphone. Soundex is a phonetic algorithm for encoding names by sound as pronounced in English, with the goal of matching names with the same pronunciation despite minor differences in spelling. Metaphone is a similar algorithm, attempting to improve on Soundex.

The primary problem with such phonetic algorithms is that they were developed for use with standard English. They are complex algorithms with many rules and exceptions that attempt to account for irregularities of spelling and pronunciation in English. They do not work well

with historical words, words in other languages, or most proper names. For the vocabularies discussed in this book, such algorithms do not work well because historical terms, proper names, and terms and names in all languages (not only English) may be represented; furthermore, name inversions and punctuation idiosyncrasies cause complications not found in standard texts in English.

9.3.5.5. Singulars and Plurals in Retrieval

A retrieval system should accommodate the end user entering either the singular or plural form of the term (or any other grammatical variant), whenever possible. While automating this facility does not produce useful results for all languages, it is useful to target languages that will significantly improve matching in the application.

For example, if an end user searches for the plural *portals*, he or she should retrieve records containing the singular term *portal*. One method of accomplishing this is through automatic *stemming*, a common retrieval feature that retrieves the term and all of its grammatical variants (e.g., stemming on *frame* would also retrieve *frames*, *framing*, and *framed*). While stemming improves access to natural-language texts in English (which can contain any English word representing all parts of speech), it is less useful for art vocabularies recorded as fielded data (which tend to contain specialized terms, primarily nouns, formulated according to precise rules).

Rather than using off-the-shelf stemming routines, a more efficient method to deal with singulars and plurals is to formulate special algorithms that better suit the content and rules employed in the target vocabulary data. For example, adding and subtracting the letter *s* aids in matching singular and plural terms in English, Spanish, and a few other languages. In the *AAT*, terms may exist in either plural or singular forms. However, due to constraints of practicality, the singular forms have typically not been added for all *used for* terms. Therefore, creating a special routine to subtract and add a final *s* to both the existing *AAT* data and the incoming user queries increases retrieval for many terms. These terms will not be automatically added to the authoritative *AAT* database, but they are used in the special normalized index created for the retrieval process. For example, in a query for *Turkish dome*, the search engine would look to normalized constructed keywords and strings to which *s* has been added or subtracted:

For the term *domes*, *Turkish* versions of the normalized strings with the *s* subtracted are included:

DOMES
 TURKISH
 DOMESTURKISH
 TURKISHDOMES
 DOME
 TURKISHDOME

Another way to improve the retrieval of singular and plural terms would be to automatically truncate words to find a match (e.g., *dome** AND *turkish**); this may help with plural forms, but automatic truncation can adversely affect accuracy of retrieval overall.

9.3.5.6. Abbreviations

In cases where the terminology may regularly contain abbreviations, or where users may expect to search using abbreviations, common abbreviations may be mapped to the full word in the term to increase retrieval accuracy. For example, users may expect to retrieve a town by the term *W Lafayette*, but if the value in the vocabulary is *West Lafayette*, the correct record will not be retrieved. An index may be created, mapping words that have common abbreviations to the abbreviation values, in order to add the abbreviated variant (*W Lafayette*) for the purposes of retrieval.

St. Louis
 W Lafayette
 Mt Everest
 Moskovskaya Ob

9.3.5.7. Trunk Names

Some terms or names consist of a core or trunk word or phrase, combined sometimes—but not always—with a modifying word to form a name or term. This happens often with geographic names and certain other classes of terms. Consideration must be given to allow access regardless of whether or not the modifier of the *trunk* or core element of the name is included in the user's query.

For example, depending upon which published atlas or gazetteer the user consults, the name of a particular mountain/volcano can be *Mount Etna*, *Berg Etna*, *Monte Etna*, *Mt Etna*, or simply *Etna*, with the words *Mount*, *Monte*, or *Berg* omitted as descriptive phrases that are not truly part of the name. Therefore, an efficient retrieval interface would allow users who enter *Berg Etna* to find the correct place, even if the vocabulary includes only the term *Mount Etna*. This could be done by maintaining a table of descriptive words and phrases that could be added or omitted from the trunk name.

9.3.5.8. Form and Syntax of the Name

Names referring to the same concept may be recorded according to a variety of syntactical conventions. The jumble of information on the Web includes texts in which names occur in natural order, along with indexes, catalogs, and other structured data resources in which the standard syntax may be in inverted order. Access should be accommodated no matter what the syntax of the name is in the target data. The retrieval system should accommodate end users' use of terms and names in either natural or inverted order. For example, a search on *Wellesley, Arthur, Duke of Wellington* should retrieve records containing *Arthur Wellesley, Duke of Wellington*.

Searching by keywords accomplishes this to some extent. However, accuracy is increased with the adoption of a routine that creates variant names by pivoting on the comma.

9.3.5.8.1. First and Last Names

Most of the vocabularies discussed in this book do not parse first and last names into separate fields. A single field is used to store the value of the terms and names; commas are used to create inverted forms of the names and terms. The reason for this is that a large percentage of names and terms used for art information could not appropriately be parsed into separate last names and first names, because the use of first and last names is a relatively modern Western custom. Non-Western and early Western artists may not have a first and last name, such as those with qualifiers that are patronymics (as in *Bartolo di Fredi*, meaning "Bartolo son of Fredi") or place name qualifiers (as in *Gentile da Fabriano*, meaning "Gentile from Fabriano").

First and last names do not apply to geographic, generic concept, and subject terminology; however, these names and terms may be inverted in ways similar to people's names. In addition, it is convenient for maintenance and retrieval if all vocabularies used in an institution have the same or very similar data structures.

In all cases, retrieval should not require users to distinguish first and last names. However, retrieval systems must still account for users who may try to look for people by last name and who may look for other vocabularies' terms in similar ways.

9.3.5.8.2. Pivoting on the Comma

Special processing of the terms based on commas is advantageous in retrieval, given the wide variety of possibilities in forming inverted names by using commas and because the vocabulary may contain proper names, generic terms, and words in all languages.

Useful variations of names and terms may be created by establishing algorithms that use the comma as a pivot; this should be used behind the scenes in retrieval only and should not be visible to the end user (because some of the variants created in this way will be nonsense).

Using the comma as a pivot, values are flipped on either side of the comma and other punctuation to create an indexing term. For example, for *Atakora, Chaîne de l'*, an algorithm can create the flipped term *Chaîne de l'Atakora*. Both of the terms may then be normalized, removing case sensitivity, spaces, punctuation, and diacritics—for example, *ATAKORACHAINEDEL* and *CHAINEDELATAKORA*.

Fig. 69. Names such as *Jan Breughel the Elder* may be preprocessed and indexed to allow more satisfactory retrieval.

Jan Brueghel the Elder (Flemish, 1568–1625); *The Entry of the Animals into Noah's Ark*; 1613; oil on panel; 54.6 × 83.8 cm (21½ × 33 inches); J. Paul Getty Museum (Los Angeles, California); 92.PB.82.

9.3.5.8.3. Multiple Commas

In cases where a name or term has two or three commas, algorithms should be developed to flip the portions of the term into two or more reasonable formulations. In inverted names, common conventions are not consistent regarding which parts of the name may be at the far right of the inverted phrase containing multiple commas. Even though one string might be nonsense, making variants by pivoting on multiple commas results in useful combinations in half the cases; the strings do not display to end users. The resulting strings would then be normalized, removing punctuation, case sensitivity, spaces, and diacritics.



Inverted name with two commas:

Breughel, Jan, the elder

Two indexing strings created by pivoting on the commas:

Jan Breughel the Elder
the Elder Jan Breughel

Values added to the normalized index for retrieval of this one name

JAN
BREUGHEL
THE
ELDER
JANBREUGHELTHEELDER
THEELDERJANBREUGHEL

Inverted name with two commas:

Wren, Christopher, Sir

Two indexing strings created by pivoting on the commas:

Sir Christopher Wren
Christopher Wren Sir

Values added to the normalized index for retrieval:

CHRISTOPHER
WREN
SIR
SIRCHRISTOPHERWREN
CHRISTOPHERWRENSIR

9.3.5.9. Articles and Prepositions

Additional normalized combinations of words in names and terms should be created to account for differences in the treatment of articles and prepositions in various languages. For example, processing may involve the construction of additional keywords by an algorithm that grabs any lowercase word to the right of the comma to make a last name to the left of the comma. For instance, even though the name in the vocabulary is inverted *Gogh, Vincent van*, users may consider his last name to be *Van Gogh*. Apostrophes, hyphens, and other designated punctuation may also be considered pivots to create additional last names or joined keywords. Once the additional terms and strings have been compiled, they should be added to the normalized index for retrieval.

Normalized strings and keywords for the name *Gogh*,*Vincent van:*

GOGHVINCENTVAN
 GOGH
 VINCENT
 VAN
 VINCENTVANGOGH
 VANGOGHVINCENT
 VANGOGH

Normalized strings and keywords for the name *Atakora*,*Chaîne de l':*

ATAKORA
 CHAINE
 DE
 L
 ATAKORACHAINEDEL
 CHAINEDELATAKORA
 LATAKORA
 DELATAKORA

9.3.6. Reserved Character Sets

Certain punctuation and words are used by query languages to designate specific aspects of the underlying logic of formulating queries. When these reserved words and nonalphabetic characters are part of the actual content of the vocabulary, care must be taken that they do not conflict with the same special characters required in search commands. For example, if parentheses, other special characters, or the words *or* and *and* are used in the term field, it should be possible to avoid their being interpreted as nesting indicators or Boolean indicators in a search statement. Where the potential for such ambiguity exists, programming algorithms or another method, such as substituting the problematic characters, should be adopted. For example, the Boolean operators may be expressed in all capitals to distinguish them from terms containing *and* or *or*. In the example below, the term *William and Mary* is a term referring to an English style.

The following search phrase includes the term *William and Mary* and the Boolean *OR*:

William and Mary OR Jacobean

9.3.7. Stop Lists

Stop lists contain words that are ignored in queries. In standard search processing, typical stop lists include articles and prepositions in English. However, for a vocabulary database, these words are not meaningful in a stop list because, unlike in natural language, they do not occur with great frequency in terms and names. In fact, articles and prepositions are critical components of certain names and terms that must not be ignored in searches. For example, the term *clerks of the works* refers to architectural workers and must be retrievable in the *AAT*; *Master of the Encarnación* must be retrievable in the *ULAN*.

The purpose of stop lists is to avoid retrieving impractically large sets of results, particularly in keyword tables. If it is necessary to devise stop lists in vocabulary databases, words that are appropriate for the terminology should be used. At the same time, users must be able to search for words that are on the stop list if they wish to pursue it. Users should be prompted to use quotes or to narrow their search with additional criteria. For example, in a geographic vocabulary, including the word *lake* in the keyword table could result in too many hits on the tens of thousands of lakes in the database that have the word *lake* in their name. Users would be prompted to use *lake* with another keyword to narrow the search. However, there are towns named simply *Lake*, so it must be possible to retrieve them, even if the word *lake* is on a stop list. One solution is to allow retrieval of *Lake* as an exact phrase (not a keyword), enclosed in quotes.

9.3.8. Boolean Operators

Boolean operators are logical operators used as modifiers to refine the relationship between terms in a search. The three Boolean operators most commonly used are AND, OR, and NOT. For names and terms, a minimum requirement is that complex AND and OR queries must be allowed. They should be used with parentheses and other punctuation to form logical groupings of criteria in queries.

Bay of Biscay OR Biscay, Bay of
(Castillo OR Rancho) AND Diego
Monte AND Oliv*

9.3.9. Context of Terms in Retrieval

In addition to names and terms, other information in the vocabulary can be used to aid in retrieval. Most importantly, the context of the term in the vocabulary is often important in assuring accurate and meaningful retrieval.

9.3.9.1. Qualifiers in Retrieval

In some vocabularies, the qualifier (a word or phrase used to disambiguate homographs) may be recorded in the same field with the term, perhaps separated by parentheses. However, ideally, the qualifier should be located in a separate field, thus allowing qualifiers to be easily processed separately from the terms in retrieval, sorting, and other situations.

Automatically including the qualifier with the term in searching reduces efficiency in retrieval. Qualifiers are intended to disambiguate homographs when the term is displayed, but they can cause a number of problems in retrieval. Consider *drums (column components)*, *drums (membranophones)*, and *drums (walls)*. In displays, the parenthetical qualifiers distinguish (a) cylinders of stone that form the shaft of a column, from (b) objects with a resonating cavity covered at one or both ends by a membrane, which is sounded by striking, from (c) the vertical walls that carry a dome.

On the one hand, if the qualifiers are automatically included in query phrases across disparate databases, it is unlikely that results will be good, unless the target databases used exactly the same vocabulary resource at the data capture phase. On the other hand, given all the homographs in art information, terms are less meaningful when taken out of context, and retrieving on the name or term alone may result in imprecise results. Qualifiers and broader contexts should then be used with the user's discretion to narrow results as necessary. In the examples below, *Edo* is the name of both an African culture and a Japanese period; *stretcher* is a masonry unit, a furniture component, equipment for mounting and framing, and a conveyance. Allowing users to add the qualifier (or a word from the qualifier) may narrow a search that has returned results that are too large and unwieldy.

Edo (African culture)

Edo (Japanese period)

stretcher (masonry unit)

stretcher (furniture components)

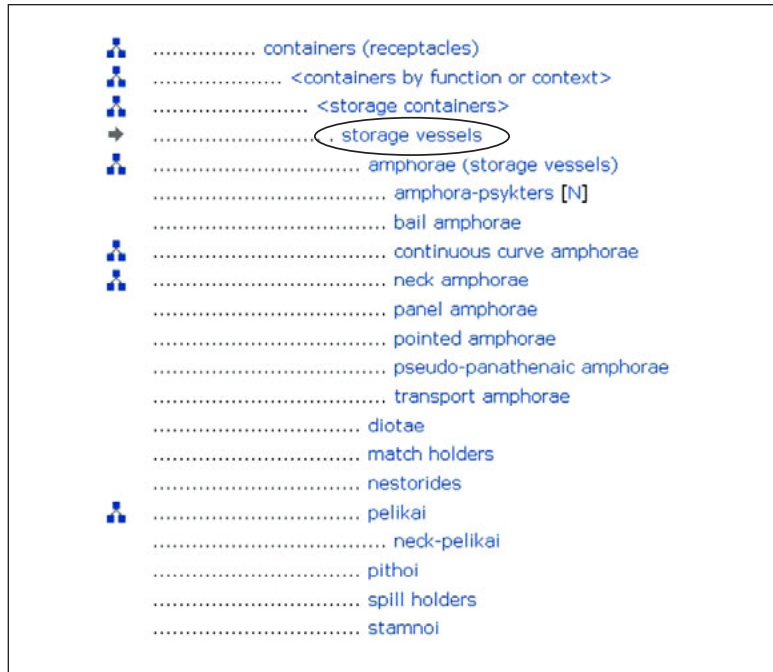
stretcher (framing and mounting equipment)

stretcher (conveyances)

9.3.9.2. Hierarchical Relationships in Retrieval

In addition to qualifiers, hierarchical relationships may also be used to provide context in order to narrow search results. For example, there are many homographs in geographic information; thus, querying on a common name, such as *Paris*, may retrieve too many results. However,

Fig. 70. Example of narrower and broader terms for *storage vessels*. If a user is searching for storage vessels, it may usually be assumed that he or she also wishes to find information for specific types of storage vessels, such as *amphorae* and *pithoi*. However, it should not be assumed that he or she would want information on all broader terms, such as *containers*.



providing a broader context to the query could narrow the results; for example, add one or more of the “parents” of Paris, which are *Europe*, *France*, and *Île-de-France*, to the search criteria.

Hierarchies are also powerful aids in expanding searches. The rules surrounding construction of thesaural relationships are largely driven by the eventuality of employing hierarchical relationships to enhance retrieval. Retrieving down a hierarchy is highly desirable; if users ask for a term, the search engine should give them the option of also including the terms for all the children of that concept (with their respective descriptors and variant terms) in the query. For example, if a user asks for *storage vessels*, most likely he or she also wants to retrieve all specific types of *storage vessels*. Therefore, the user should have the option of including all the narrower concepts for *storage vessels* in the search, such as *amphorae*, *diotae*, and *pithoi*. A user interested in Tuscany, Italy, may want to look for data associated with names of any city or town in Tuscany; the hierarchical vocabulary can provide a list of these names to be used in a search.

Retrieving up the hierarchy and retrieving siblings are not typically expected by the user and generally should not be employed. If the user searches specifically for *decanters*, he or she does not expect to

additionally receive all other types of serving vessels and their broader contexts. However, allowing the user the option of including broader contexts and siblings may be useful in certain situations.

9.3.9.3. Associative Relationships in Retrieval

Expanding a search with associative relationships may be desirable but should typically be done only when requested by the user. However, the option should be available. For example, a user interested in the wall paintings known as *frescoes* may also be interested in the related concept of *sinopie* (the drawings under a fresco), which would be linked through an associative relationship. A user interested in the French manufactory *Manufacture nationale des Gobelins* (which produced tapestries, furniture, *pietre dure*, and other items) may also be interested in information about the artists of the manufactory. The search engine could offer the user the option of also looking for Gobelins artists linked through associative relationships, including *Marc de Comans* and *François de la Planche*, among others. The following is an example of associative relationships for an architectural firm.

Richard Meier & Partners

Associative relationships:

members are:

Richard Meier

Bernhard Karpf

Michael Palladino

Reynolds Logan

James R. Crawford

9.4. Other Data Used in Retrieval

In addition to queries by names and terms, qualifiers, and hierarchical relationships, additional search criteria can be used to retrieve vocabulary records.

9.4.1. Unique Identifiers as Search Criteria

In a local or controlled environment, the unique numeric identifier for a concept can provide a link between the content being queried and the vocabulary used to aid retrieval (e.g., the seven-digit number *7008038* is the unique identifier of *Paris, France* in the *TGN*). For instance, the identifier for a particular object or concept in a controlled vocabulary could be placed in an object record by a cataloger (presumably in an

automatic way aided by the cataloging system) and could thus be linked to the vocabulary to provide extremely accurate retrieval through the use of variants and other data.

This method, of course, does not work when querying across disparate databases that do not all use the numeric identifier or when querying across the Web at large. Generally, in these cases, the vocabularies can be used to suggest terminology to users for queries or to broaden queries automatically behind the scenes; however, they cannot guarantee refined, precise results.

9.4.2. Other Vocabulary Data Used in Retrieval

Controlled descriptive information in the vocabulary record may be used for retrieval. For example, the place type for geographic information or the life roles of people are controlled lists that would be helpful in narrowing searches. The nationality of a person, geographic coordinates of a place, or associated dates would also be useful in retrieval. Such criteria would typically be used in a query in combination with names or other information.

For example, with geographic information, users could find all villages within a certain set of coordinates. With artist information, a user may want to find vocabulary records for all people who were English watercolorists (*English* is the nationality, *watercolorist* is the role); once these records are retrieved, the names in the vocabulary records could be gathered to use in a search against work records or other content objects. On the following page are examples of search interfaces using information in addition to names for retrieval.

9.5. Results Lists

A critical issue related to querying vocabulary data is how to display the information once it is retrieved. Since vocabularies can be very rich and complex, decisions must be made regarding how to display the information without confusing or overwhelming the user. An initial results list should include matches on the terms or names used in the query as well as a brief reference to each concept (e.g., for *TGN*, a preferred name, place type, and hierarchical context). From there, the user can either view the full record for the concept or view the concept in the full hierarchical display. Displays are designed with the goal of presenting as much information as necessary in a clear and coherent way. See **Chapter 7: Constructing a Vocabulary or Authority** for further discussion of displays.

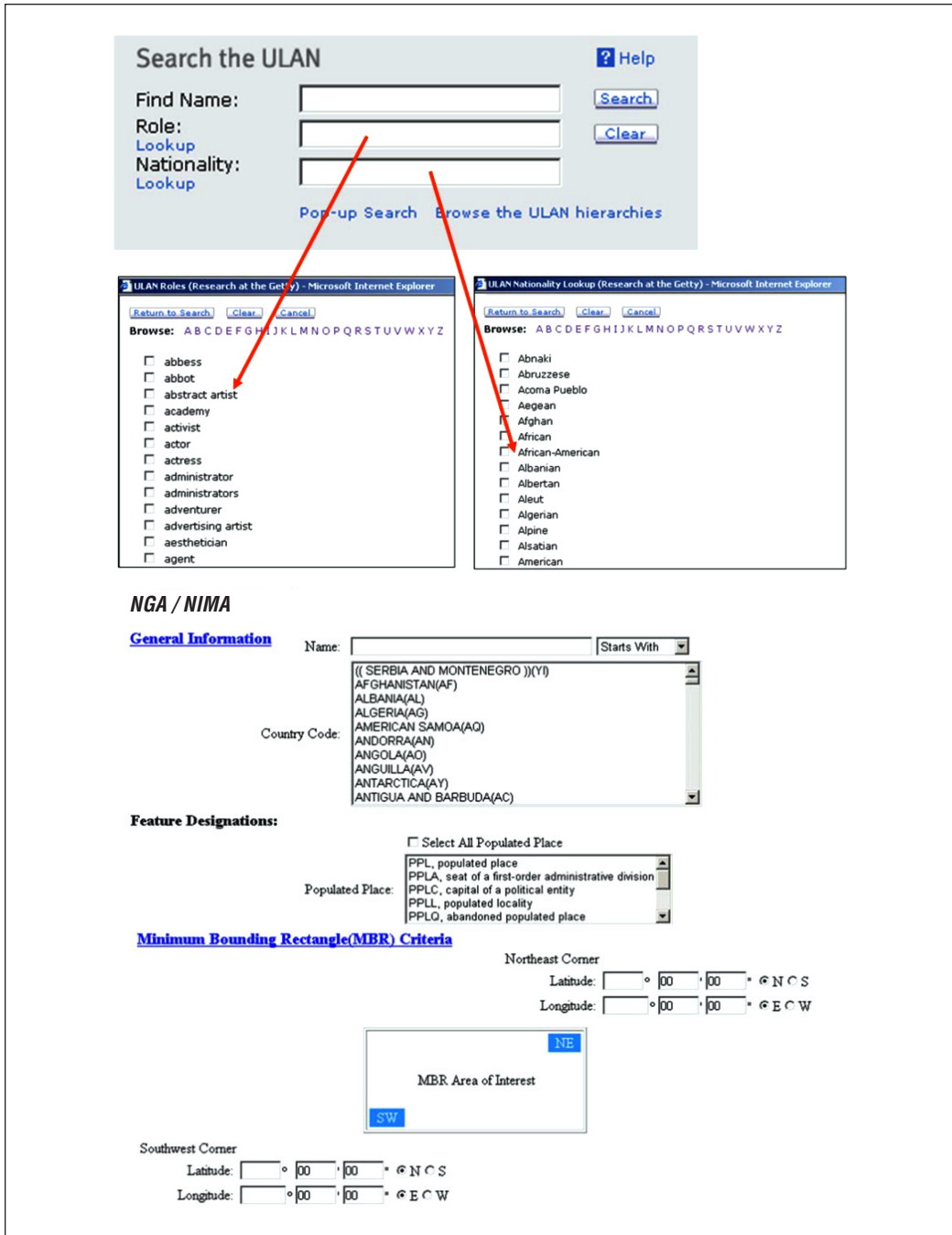


Fig. 71. Examples of search screens with various criteria for queries in the *ULAN* and National Geospatial-Intelligence Agency (NGA) *GEOnet Names Server (GNS)*. In the upper image, criteria for a search in the *ULAN* include the name, nationality, and role (e.g., *abstract artist*). In the lower image, a search on the NGA data includes criteria such as the name, the nation to which the place belongs, place type, and coordinates.