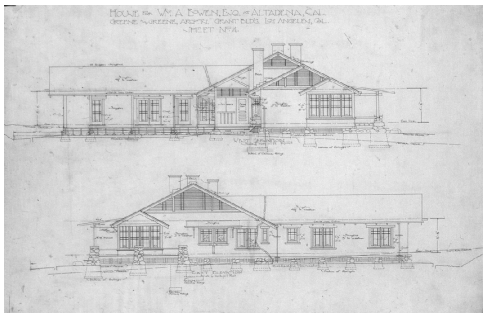
CLiMB

Computational Linguistics for Metadata Building



Center for Research on Information Access
Columbia University

Funded by the Andrew W. Mellon Foundation
March 2002–June 2004

Progress Report
Second Year Cumulative

CLiMB Progress Report
Second Year Cumulative

Acknowledgements

This report was a collaborative effort of the entire CLiMB project team.

For reference purposes, authors are:

Judith Klavans, PI
Roberta Blitz
Peter Davis
Stephen Davis
David Elson
Angela Giral
Amy Heinrich
Vera Horvath
David Magier
Rebecca Passonneau
Patricia Renfro
Robert Scott
Mark Weber
Robert Wolven

Executive Summary

Goals of the CLiMB Project

The CLiMB project at Columbia University aimed to discover to what extent and under which circumstances automatic techniques could be used to extract descriptive metadata from texts associated with image collections. Ordinarily, descriptive metadata (in the form of catalog records and indexes) are compiled manually, a process that is slow, expensive, and often tailored to the purpose of a given collection. Our goal as a research project was to explore the potential for employing computational linguistic techniques to alleviate some of the cataloging bottleneck by enhancing descriptive metadata with automatic procedures. Among our objectives was the creation of a set of tools that could ultimately and easily be used by other projects, in order to enable sophisticated cataloging procedures for large digital image collections. At the end of the two-year grant period, CLiMB developed a prototype Toolkit that enabled users to identify and extract metadata from written material associated with the images in our test collections. The second year of CLiMB also included a two-phase evaluation component, which measured the effectiveness of the tools and the usability of the Web-based interface. To our knowledge, CLiMB's approach to issues of automatic metadata extraction from selected authoritative texts is unique.

Highlights of Progress—The Second Year of CLiMB

The first year of the CLiMB project resulted in a set of criteria for collection selection, an initial Toolkit aimed at catalogers, and a preliminary approach for use and evaluation. Though it is our intention to make this report a complete account of the project's achievements over the full two years, some detail relating to the first year of CLiMB has been abbreviated in order to focus on second-year achievements. The first year report is still available as a printed document and as a PDF file (http://www.columbia.edu/cu/libraries/inside/projects/climb/reports/CLiMB.1year.report.pdf). During the second year, CLiMB shifted focus from selecting and developing tools for extraction to making those tools available for targeted user groups in evaluation sessions. As such, we formed a new group for Evaluation. We finalized the collections for testing, and confirmed the selection criteria. The prototype CLiMB Toolkit as it now exists represents the combined efforts of these three project groups: Curatorial, Technical, and Evaluation.

Collections

- On the advice of CLiMB's External Advisory Board, selected and developed an additional test collection. The collection is based on Saskia images available at the Columbia Libraries and text from the North Carolina Museum of Art's *Handbook of the Collections*. Significant features of this collection are:
  - a clear relationship between text and images,
  - well-edited concise text,
  - English-only text to avoid translation and transliteration issues.
- Developed Collection Guidelines for future selection of collections.

Technical

- Built the system back-end to support the modular processing of texts, target object identifiers (TOIs) and authority lists (see Section 3.3).
- Completed computational linguistics research on TOI-Finder and Sectioner modules.
- Constructed an elaborate Web-based front-end "CLiMB Toolkit" interface which makes the functionality accessible to many different user groups, including lay and technical groups.
- Proofread, documented, packaged, and released the code for future development.

Evaluation

- Held a Formative Evaluation Meeting in October of 2003:
  - refined the initial goals to target catalogers,
  - confirmed differences in usability based on user background and goals.
- Assembled a group of target users for Toolkit Evaluation in April of 2004:
  - based on responses from Visual Resources Association (VRA) presentation,
  - built a solid core for feedback on usefulness,
  - results showed promise for tools.
- Employed user feedback while developing the final version of the CLiMB Toolkit prototype.
- Summarized findings which serve as the basis for future development.

Preface: Elaborating the CLiMB Problem

The innovative contribution of the CLiMB project is its use of computational linguistic tools in order to extract metadata from text associated with images in large digital collections—hence **C**omputational **Li**nguistics for **M**etadata **B**uilding. The project applies techniques developed in computational linguistics to the problem of cataloging and describing the images in a given collection; it does this by putting to use information that already exists in the form of written material that is related to those images. Although manual cataloging is an established field, what is novel about the CLiMB approach is the notion that some cataloging might be accomplished automatically. The CLiMB project seeks to address the cataloging bottleneck that arises as the volume of available data increases with new technology; the project was also driven by the possible benefit to end-users, enabling access to information across collections and vocabularies. As a research project, CLiMB sought to create a platform in which to determine whether such an approach was indeed useful, and to what extent results could be incorporated into existing metadata schemes.

When a collection of images is cataloged, art librarians may provide intellectual access to the material by selecting relevant name and subject terms for those items from existing authority lists and controlled vocabularies. If, for example, a collection contains images of artworks, an art cataloger can employ his or her knowledge of the material to select the applicable terms from among such lists and vocabularies, which can then be entered into a catalog or index. These terms answer a fundamental set of questions that might be put to the collection by potential users when they are searching for images: *What does this image depict? What is the object made out of? What other physical characteristics does it have? Who created it? What is its historical and cultural significance?* The benefit of using experts to perform this task is that they possess an extensive and invaluable understanding of the material. However, the increasing size of digital collections now being created makes it both too expensive and too time-consuming to accomplish this manually with the full level of detail needed to provide maximum user access.

In many cases, researchers have already described aspects of collection images in contexts such as scholarly monographs and subject-specific encyclopedias. CLiMB employs text sources that closely mirror the content of a given digital image collection to automatically extract descriptive metadata from those texts—in effect, using the writings of specialist scholars to enrich the catalog. The challenge is to identify the *meaningful* facts (or metadata) in the written material and distinguish them from among the thousands of other words that make up a text in its original form. CLiMB undertook the task of determining whether this could be accomplished automatically or semi-automatically. The deliverable result of the project is the prototype of the CLiMB Toolkit, which contains a select group of computational linguistic software tools that work together beneath (and are adjustable through) a Web-based user interface. During the second year of the project we were able to test an early version of this Toolkit with our potential users, primarily catalogers and digital resource specialists.

This report describes progress in three main areas—Collections, Technical, and Evaluation—but the report is organized chronologically rather than thematically. This was done, first of all, in order to make the process over time clear to the reader. Just as importantly, it allows the report to proceed naturally through the two years without difficult cross-referencing; the report can be read straight through, though each section should be comprehensible on its own. Readers who are interested in

seeing the specific developments of the first year should consult the *CLiMB First Year Report*, which is available online as a clean PDF at http://www.columbia.edu/cu/libraries/inside/projects/climb/reports/CLiMB.1year.report.pdf.pdf (it can also be found under the heading of *CLiMB One Year Report to the Mellon Foundation* on our main presentations and publications page, http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html).

Section 1 of the present report is devoted to collections. During the first year of the project, collections with varying characteristics were selected in order to ensure that our research would permit us to discover how different techniques might produce a range of results for distinct collections. In the second year, we developed an additional test collection, which was our fourth, based upon criteria we elaborated earlier in the project. This additional collection has characteristics that might more accurately represent the kinds of texts and images to which CLiMB software would be successfully applied on a larger scale in the future.

Section 2 of the report is devoted to technical developments. During the entire first year of the CLiMB project, and for a significant portion of the second year, the Technical group was concerned largely with testing and developing software tools. This testing was designed to assess the capabilities of various computational linguistic techniques in the context of extracting metadata for catalog records. Decisions about which tools were helpful, and how to use those tools in conjunction with one another, were made in advance of placing them into a user interface.

Section 3 discusses the first major aspect of CLiMB's evaluation process, the Formative Evaluation Meeting, which was held in October of 2003. For this meeting we invited a group of experts from diverse fields related to the CLiMB project to address two important questions: first, we asked what sort of data CLiMB software should be harvesting, and second, we asked how we might evaluate that data in a full scale evaluation in the future.

With results and feedback from the first phase of evaluation, technical attention was turned to developing a user interface that made the tools workable for a specific set of users, namely catalogers. Building on earlier versions of a browser interface, a prototype version of the CLiMB Toolkit was developed before the end of the project grant period. **Section 4** provides an overview of this Toolkit and discusses the integration of the software tools into the interface.

Finally, **Section 5** of this report covers the CLiMB Toolkit Evaluation, which was held in April of 2004. For this evaluation meeting, we selected a group of target users from the community of digital resource professionals and had them work with the Toolkit in a controlled environment that would allow us to collect data on usability and effectiveness. It is largely as a result of this evaluation that we were able to make final changes to the look and function of the Toolkit.

1. CLiMB Collections

1.1 Initial Test Bed Collections

Early in the CLiMB project the Curatorial Committee selected three test bed collections comprised of images and associated texts with which to begin the development and testing of CLiMB software tools. Two of these collections, the Greene & Greene Architectural Records and Papers Collection, and the Anne S. Goodrich Chinese Paper Gods Collection, are housed at Columbia University. The third collection, South Asian Temples, consists of digitized photographs and metadata from the American Institute of Indian Studies as part of the Digital South Asia Library (http://dsal.uchicago.edu/images/aiis/about.html). For each image collection the Curatorial Group compiled a bibliography of associated texts. A complete list of these can be found in **Appendix B: Collections and Related Material**.

The Greene & Greene Architectural Records and Papers Collection at the Avery Architectural and Fine Arts Library consists chiefly of architectural drawings by the brothers Charles Sumner Greene (1868–1957) and Henry Mather Greene (1870–1954), architects based in Pasadena, California. **Figure 1.1** below provides an example of one of these images. The CLiMB process presupposes the existence of at least minimal level catalog records (creator and title) into which harvested metadata terms can be inserted. A primary reason for the selection of Greene & Greene as a test bed collection was that in addition to containing 5,000 digital images of architectural drawings and photographs, it included detailed MARC (MAchine-Readable Cataloging) records. These records were created as part of AVIADOR (Avery Videodisc Indexing of Architectural Drawings Online in RLIN), an early project that pioneered the linkage of images on videodisc to descriptive cataloging records in the Research Libraries Information Network database. The Curatorial Committee's bibliography of associated texts contained seven monographs about the Greenes. All titles were scanned and made available for use by the Technical Group, with development focusing on Edward R. Bosley's book *Greene & Greene* (London: Phaidon, 2000).

An interesting problem posed by this collection is the relationship between the text and the images. All of the authors of the Greene & Greene books write about the architectural projects as fully realized buildings, yet the bulk of the collection consists of architectural drawings—e.g., plans, elevations, and sections, as seen in **Figure 1.1**. Furthermore, although the texts might focus on a particular project, references both to other Greene & Greene projects and to related architectural developments are

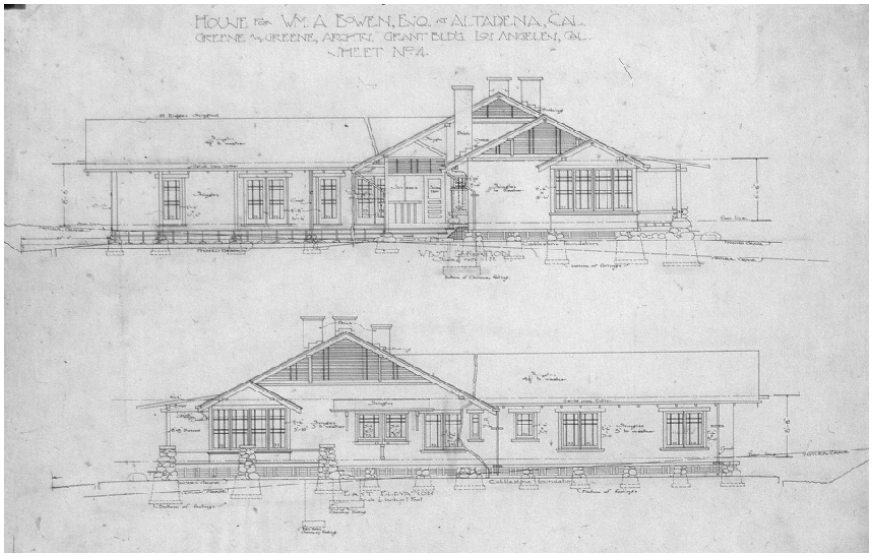


Figure 1.1: William A. Bowen House (Altadena, Calif.); west elevation, east elevation: sheet no. 4 (Greene & Greene Architectural Records and Papers Collection, Avery Library, Columbia University, NYDA.1960.001.00512).

frequently woven throughout lengthy passages of text. Thus, identifying precisely which text refers to which image presents a challenge for automatic techniques. The research question raised here is how closely matched images and texts need to be in order for the software to be able to yield useful descriptors.

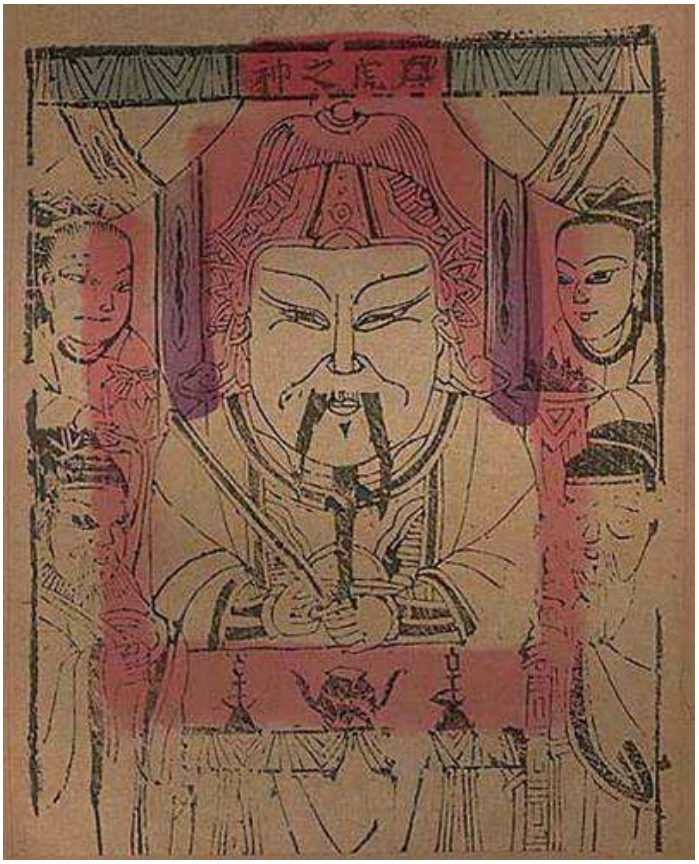


Figure 1.2: *Pan-hu chih shen* (Anne S. Goodrich Collection, C.V. Starr East Asian Library, Columbia University).

The Chinese Paper Gods Collection in the C.V. Starr East Asian Library contains 256 woodblock prints depicting popular Chinese deities; **Figure 1.2** is an example of one of these prints. In contrast to Greene & Greene, this collection's strength is the close relationship between images and text, because the author of one of the major monographs on the subject, Anne S. Goodrich, was also the donor of the collection. Her book *Peking Paper Gods: A Look at Home Worship* (Nettetal: Steyler Verlag, 1991), describes many of the individual images in the collection at the Starr Library. However, this collection presented new challenges: alternate transcription systems from the original Chinese, and numerous variants for the names of the deities.

CLiMB's third test bed collection was to consist of selections from the American Institute of Indian Studies (AIIS) Photo Archives, available as part of the Digital South Asia Library (DSAL), a collaborative project designed to provide scholarly digital materials on South Asia. The AIIS col-lection is large, containing over 125,000 photographs, and includes the categories of architecture, sculpture, terracotta, painting, and numismatics. This CLiMB test bed collection was referred to as "South Asian Temples," as it was to focus only on temple sites in the AIIS collection. Development of this collection was postponed for a variety of reasons. Like the Chinese Paper Gods, South Asian Temples posed a challenge due to the occurrence of variant name forms. In this case the names of particular temples were often generic, which made it necessary to qualify each site geographically. However, these geographical names themselves (designating the location of a particular site) often underwent changes over the course of time. Also, as with Chinese Paper Gods, transliteration was a problem and the heavy use of diacritics presented an obstacle for the optical character recognition software used for scanning in the project. Finally, due to the large number of images in the AIIS database, it proved difficult to identify a workable subset for use on the project.

1.2 Fourth Test Bed Collection

In the second year of the CLiMB project, an additional collection was selected, largely following criteria that derived from the input of the CLiMB External Advisory Board, which held its inaugural meeting in June 2003. After detailed presentations on the collections and technical progress, the Board advised the creation of an additional collection that purposefully avoided some of the challenges inherent to the initial test bed collections. To facilitate the development and testing of CLiMB tools, it was necessary for this collection to present a close association between text and images, contain a limited number of variant name forms, and preclude the various issues surrounding multilingual transcription. Three types of publications were suggested by the Board as possible texts for this collection: museum collection catalogs, exhibition catalogs, and catalogs raisonnés. Using this type of highly structured publication would also enable the project team to test a current working assumption that the choice of a CLiMB collection should be driven as much by text—both in terms of content and structure—as by images; the CLiMB Toolkit could potentially make its most important contribution using museum publications. The scholarly nature of these works and the regularized presentation of metadata make them good candidates for computational linguistic techniques, more so than a discursively arranged text such as Bosley's *Greene & Greene*.

Digital images from the North Carolina Museum of Art were selected as the additional collection for this phase of the project. In addition to fulfilling the requirements outlined by the Board regarding multiple languages and transcription, this collection was developed because forty-two works of art described in the museum's collection catalog, the *North Carolina Museum of Art: Handbook of the Collections* (ed. Rebecca Martin Nagy [Raleigh, NC: NCMA, Hudson Hills Press, 1998]) were available for CLiMB testing through Columbia's licensing of the Saskia digital image archive. One such image is shown here in **Figure 1.3**.

Figure 1.3: Claude Monet. *The Cliff, Etretat, Sunset*. 1883. North Carolina Museum of Art (67.24.1).

As part of the April 14, 2004 CLiMB Toolkit Evaluation, participants were asked to perform a metadata selection task using images and texts about three works from the CLiMB subset of the North Carolina collection—Jan Jansz den Uyl's *Banquet Piece*, Jan Brueghel the Elder's *Harbor Scene with St. Paul's Departure from Caesarea*, and a statue of *Neptune* attributed to Benvenuto Cellini. They were instructed to use the prototype of the Toolkit to locate relevant passages from the *NCMA Handbook*, to choose text display options, and finally to select descriptive terms for a catalog record. Specific discussion of the Toolkit Evaluation can be found in Section 5 of this report.

1.3 The Target Object Identifier (TOI) Concept

The conceptualization and development of new methodologies and new software often require new concepts; an interdisciplinary project like this one often demonstrates this by introducing new terminology or redefining old terminology. A key concept and term for the CLiMB project is Target Object Identifier (TOI). A group of TOIs represents a de facto "authority" list of the names of the entities that require subject descriptors in a given collection. The TOI provides a unique reference

for an object so that CLiMB software can automatically locate descriptive terms for it within a given text. After the CLiMB Toolkit locates a TOI in a text, it can then more accurately identify the descriptive words and phrases associated with the object to which the TOI refers. The TOI concept differs from the standard named entity concept in computational linguistics in that the TOI assumes an authoritative status.

A TOI contains one or more fields, depending upon the nature of the collection. For example, all the drawings in the Greene & Greene collection are by the architects Greene & Greene, so the TOI consists only of a single field, which is the name of the architectural project: for instance, *Charles M. Pratt house* or *William A. Bowen house* (as in **Figure 1.1** above). The cataloger employs the CLiMB Toolkit to first find occurrences of a TOI and its variants (Pratt house, the house) within a text; the Toolkit is then employed to locate common nouns, proper nouns, and terms from an external vocabulary, such as the Getty Art & Architecture Thesaurus, that might be associated with the TOI.

Similar to the Greene & Greene collection, the TOI list for the Chinese Paper Gods collection contains only a single field, the name of the depicted deity (the creators of these prints are anonymous). However, a TOI will contain multiple fields when necessary, as is the case with the North Carolina collection. There, the name of the creator of the object and the title of the work are needed to identify an object uniquely: for instance, "Jan Jansz den Uyl; *Banquet Piece*" or "Claude Monet, *The Cliff, Etretat, Sunset*." Conceivably a third (or even fourth) field could be required to identify a work of art distinctively. Participants in the Roundtable Discussion at the April 2004 Toolkit Evaluation raised the example of John Singleton Copley's *Watson and the Shark,* a work that exists in multiple versions. Further qualification by repository or date would be needed to establish a TOI to identify such an object.

The TOI list for a collection can be compiled in a number of ways. For Greene & Greene and Chinese Paper Gods the list was extracted from the titles supplied by the cataloger (MARC field 240). The TOIs for North Carolina collection were generated from local TEI markup specifications for the museum collection catalog. Other possible methods of compiling TOI lists include using external sources, such as image vendor lists and back-of-the-book indexes.

1.4 Next Steps

Initial work with the test bed collections has enabled project members to begin to identify the elements that make a collection a potentially fruitful candidate for the CLiMB approach. More testing would be needed before authoring definitive guidelines on selecting candidate collections; however, at present the set of characteristics employed to select the fourth collection (particularly as distinguished from the first three) represent a major step in this direction. The guidelines thus far developed could be used in further CLiMB research, or by other, similar projects.

Additionally, one project follow-up would be in-depth testing of the Toolkit with museum publications, one of the text categories CLiMB found to be particularly promising. Collaboration with partners in the art museum community may prove to be very productive, particularly if the institutions are able to digitize their exhibition catalogs and other scholarly publications about their collections. Such a partnership could yield practical results, since there is a real need to enhance image access through the use of subject descriptors, as museums and other cultural institutions engage in large-scale digital imaging projects.

2. CLiMB Tools

2.1 Modeling CLiMB As a Conceptual Process

CLiMB is an interdisciplinary project that aims to use computational linguistics to solve a problem in library science. As such, the Technical Group, with its background in computer science, was charged with bringing that field's techniques to bear on the abstract problem described by the Curatorial Group.

As in all challenges, the initial, and most significant, step in implementing a solution is clearly defining the problem. In this case, to define the problem is to model CLiMB as a conceptual process, grounded in prior work in computational linguistics. In other words, we wanted to recast the CLiMB problem as an extension and application of an existing class of problems in computer science. If done correctly, we could then extend and apply the processes and tools that have been developed to solve the similar problems in the class. An appropriate model of the CLiMB problem would:

1) take no more input than what was available from the Curatorial Group's conception of the problem (that is, associated texts, TOIs, and authority lists);
2) give no less output than what was required by the Curatorial Group's conception of the problem (that is, subject access terms for each TOI);
3) make maximum use of existing tools and processes in computer science; and
4) be implementable in a reasonable time frame with maximal portability (that is, avoiding the use of proprietary or commercial tools).

The Technical Group considered several models before settling on the one it finally implemented.

Information Retrieval. Many computer science people, upon hearing about CLiMB, initially liken it to document indexing and retrieval, an active area of research in computer science within information retrieval (IR). IR is the problem of finding and retrieving documents relevant to particular topics, from a sea of unrelated documents. Search engines such as Google, for instance, are commercial IR services that allow users to find Web pages about particular topics from across the Internet.

Though there is a large body of recent work addressing document indexing and retrieval, CLiMB does not quite fit into its paradigm. IR, for example, assumes that a user has issued a single query and that the system must respond with documents or excerpts of documents. In CLiMB's case, all the computation is offline, as we seek to extract subject access metadata that a hypothetical user may search once the records are mounted in an image search platform. We do not seek to return a list of entire documents for an arbitrary query, as does Google, but rather to look within a text, reason about it, and carefully extract words and phrases about which we are very confident.

Put another way, there is a trade-off between the number of documents one can consider and the knowledge and specialization one can bring to analyzing each document. IR techniques are coarse tools geared toward getting the best results, on average, across hundreds of thousands of documents

about which the tool knows very little. CLiMB, by contrast, has to do a much finer analysis on a few select texts about which it knows a great deal more.

Finding little directly applicable prior work in IR on which to build, the Technical Group decided on a different model: one of text classification.

Classification. Classification is also a well-researched class of problems in computer science. As it has been applied to hundreds if not thousands of problems across many domains over the years, the processes and tools are well-established. The basic paradigm is that there is a large set of data points which all fall into one of several classes. A classifier must collect all available information (in technical terms, "features") about each data point, and hypothesize a formula that best predicts the class for any new, unseen data points.

For example, suppose one wanted to automatically determine the political affiliation of a U.S. senator. In this case, the data points are the 100 senators, and there are two classes (Democrat and Republican). The classifier would collect all the features—their votes on various types of bills—and hypothesize a formula for guessing an arbitrary senator's affiliation. The formula might be: "If the senator votes YES for gun-control bills and NO on bills limiting abortion, the senator is in the Democrat class. Otherwise, the senator is in the Republican class."

Classifiers are widely used in many real-world applications. For example, credit card companies seek to minimize fraud by raising an alert when a transaction is probably fraudulent (indicating that the card has been stolen). The data points are transactions; the classes are Valid and Fraudulent; the features are basic information about the transaction (such as time, size, and location). A simple formula would be: "If the transaction is in a different location from most of the customer's other transactions, and is over $1000 in value, and occurs soon after other large purchases, it is probably Fraudulent. Otherwise, it is Valid."

The Technical Group was able to apply this model to CLiMB because of a key observation: every word or phrase that we would like to extract from the text is a noun phrase. Though not all noun phrases are suitable for extraction, by considering only the noun phrases in a text, we should never miss anything we want.

Given this assumption, CLiMB fit naturally as a classification problem: all of the noun phrases that exist in a text are the data points; the classes become "OK for extraction" and "Bad for extraction." Once the classifier has a formula for knowing which of the two classes fits for each noun phrase, it can go ahead and extract the "OK" ones. In other words, all the noun phrases in the text become "candidates" and CLiMB gives them all a fair chance, before extracting the ones that best fit the formula.

In order for this to be possible, there must be good features with which the formula can work. As in all classification problems, the more features there are, and the better they divide up the data points in meaningful ways, the more accurate the formula can become. Thus, the challenge for CLiMB was to collect as much information about each noun phrase as possible, in order to give users maximal flexibility when designing extraction formulas.

The Technical Group originally intended to use machine learning classifiers to create formulas automatically. In the end, though, for a variety of reasons, it instead hand-crafted a reasonable

formula for the CLiMB Toolkit. The Toolkit allows users to modify the formula easily, and create new formulas to achieve the best extraction for each situation. This is explained further below.

2.2 Defining the Process: From Conceptual Model to Computational Pipeline

The next challenge for the Technical Group was to apply the conceptual model to a process flow for a CLiMB software application. After some deliberation, the Group settled on the following pipeline for the user to experience:

1) Data collection. From the point of view of CLiMB software, all the user's texts must already be in machine-readable form. The first step for CLiMB must be to support an open text encoding standard so that it can easily read user texts.
2) Candidate identification. Since the Group decided to model the problem as one of classification, with noun phrases as the candidates, the second step in the pipeline is to identify these candidates automatically.
3) TOI consideration. One key requirement of the CLiMB application is the capability to identify which noun phrases are "about" which TOIs (see Section 1.3). For a noun phrase to be useful as descriptive metadata for a TOI, it not only needs to be judged as "good metadata" by the classifier, but also as good with respect to a particular TOI. In other words, it is insufficient when dealing with most collections simply to extract the most "suitable" metadata from a text. CLiMB must extract different sets of terms, each of which is related to one TOI.

   This last design requirement manifests itself in the application specifications as such:

   3a) Specification. The software must be able to exchange TOI information with other applications, so that users can easily upload lists of TOIs into the Toolkit from other sources. For example, the user should be able to import TOIs from collections encoded in open metadata standards such as MARC.

   3b) Finding associational context. In this crucial step, CLiMB software must associate paragraphs in the users' texts with the specified TOIs. That is, the classifier must know which candidate terms are "about" which TOIs. This is implemented in two stages: the TOI-Finder, a module which identifies those noun phrases that are direct references to TOIs, and the Sectioner, which uses the occurrences of TOIs in the text (as identified by the TOI-Finder) to decide which paragraphs are associated with which TOIs. (More discussion of these two stages can be found below.)

   3c) Application to extraction. There are two ways that the software could make use of associational context when its classifier selects noun phrases for extraction. One option is to add classes to the classifier: one for each TOI. (This would be akin to training the classifier from Section 2.1 not only to tell between Republicans and Democrats, but also Independents, Greens, Libertarians, and Whigs.) The other option is to partition the noun phrases by TOI and run the classifier separately within each partition in order to find that TOI's best candidates. This latter approach supplies superior results, and is the one that the Technical Group implemented.

4) Feature collection. CLiMB software would here need to collect as many features about the noun phrases as possible. The problem here is that many features depend on external data sources which may or may not be present. There is a certain set of features that can be determined with the text alone, but they may not be sufficiently precise for the user. Therefore, the software must be flexible enough to allow a non-technical user to supply his or her own external data source and teach the software how to use it to assign features to candidate noun phrases.

5) Ruleset creation. In the language of CLiMB software, the classifier's formula is a "Ruleset." As mentioned in Section 2.1, there is a Default Ruleset that encodes common-sense rules. But this "one size fits all" Ruleset leaves room for accuracy improvements when applied to particular texts in particular collections that the Technical Group cannot foresee. Thus, the software must offer non-technical users the ability to understand and manipulate Rulesets to achieve better results. Then, users can specialize their Rulesets to give optimal results for each text and each collection.

6) Ruleset interpretation. The software must include an interpreter for Rulesets, capable of understanding and reconciling logical rules to rank the candidate noun phrases in order of "fitness" for extraction.

7) Verification and export support. Once the software has selected noun phrases for extraction, it must allow users to verify the results and make any corrections. Once users are satisfied with those results, CLiMB software must allow them to take the results out of the software in open metadata formats.

2.3 Describing an Architecture: From Pipeline to Technical Specifications

Once the Technical Group settled on these high-level specifications, it went about designing a technical architecture and implementing solutions to many of the isolated problems.

It settled on the Perl programming language for the back-end. This unconventional choice was ideal for a research environment with its flexible, "try it and see" specifications. Moreover, Perl's greatest strengths, those in text processing, are a natural fit for this particular application.

The Technical Group found it essential to design CLiMB in a modular fashion. Since each of the operations in the pipeline described in Section 2.2 can be performed in isolation, the Group implemented them as separate Perl modules. One strong advantage of this approach was that it allowed members of the Group to work on small aspects of the system without touching the functionality of the rest. This allows for smooth concurrency, that is, work in parallel by different members of the Group on different modules.

The Group decided to use XML as the data format for encoding texts. XML is not a programming language, but a markup scheme for text files (one can think of it as a generalized HTML, the markup used for the Web). Since CLiMB deals with natural language text, XML is a natural match. One major factor here was the decision by the Curatorial Group to supply the testbed texts in TEI (Text Encoding Initiative), an XML recommendation used worldwide for the encoding of books and manuscripts in machine-readable form. Taking this cue, the Technical Group decided to embrace TEI as its data format throughout the CLiMB back-end. Each module inputs a loaded TEI-XML

file, performs some operation on it (such as finding and marking up noun phrases), and outputs the revised TEI-XML file.

The Group also wrote a front-end "backbone" program called RunCLiMB to provide a user interface to these modules. Using RunCLiMB, which featured a text-only interface on the command line, a user could pass TEI-XML files from one module to another. This made it possible to construct the "pipeline" described above simply by executing all the modules in the correct sequence. Although RunCLiMB was essential for the development and testing of the back-end CLiMB modules, it was not sufficiently user-friendly to satisfy the objectives of the CLiMB software suite. This led to the development of the CLiMB Toolkit, described in Section 4, which is a second front-end that allows users to access the same modules via a user-friendly Web interface.

The following are the approaches the Technical Group took during the development of selected CLiMB modules to solve the problems identified in Section 2.2.

Data import. CLiMB is able to read two types of encodings from the user for source texts: those in plain text, and those already in TEI-XML. The TEI specification includes a great many tags, of which CLiMB only uses a few. All the modules assume that the TEI files they receive only use tags within this subset (which the Technical Group termed "CLiMB TEI-XML"). Thus, the data import module must not only encode raw texts as new TEI-XML files, but also reformat incoming TEI files to ensure they conform to the CLiMB TEI restrictions.

Noun phrase identification. As there are already tools in computational linguistics to address the problem of noun phrase identification, the Technical Group tested and evaluated several for this module rather than re-inventing the wheel. In particular, the Group looked at Alembic WorkBench, LT Chunk from the University of Edinburgh, IBM's Nominator, and LinkIT, a product of Columbia's Natural Language Processing Group. The Group empirically determined that LT Chunk had the best results. This conclusion came naturally as the Group discovered that the rest were specialized for finding *proper* noun phrases. Since common noun phrases had to be considered valid candidates for extraction, this was not a permissible assumption. (A detailed account of the testing process can be found in CLiMB's First Year Report, which is available online at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.) Thus, the Group wrote an interpreter for LT Chunk output, and the software became an external dependency for CLiMB software.

TOI identification. This module represented one of the more original and exciting opportunities for research in computational linguistics. The challenge, similar but not identical to *coreference*, is to identify those noun phrases that are referents for the art objects identified by TOIs. For example, in the Greene & Greene testbed collection, one TOI was "David B. Gamble House (Pasadena, Calif.)." This exact string is not likely to appear anywhere in the text, but certain variations are, such as "Gamble House," "David Gamble House," and "the house." The problem becomes more complicated, though, when there are several such houses discussed in a text, each of which may be referred to in turn as "the house." To address this, the Group devised an algorithm for the disambiguation of those noun phrases that could refer to more than one TOI. This algorithm became the basis for a short paper accepted by the Third ACM/IEEE Joint Conference on Digital Libraries (JCDL) 2003: "Methods for Precise Named Entity Matching in Digital Collections."

Sectioning. The Technical Group attempted to take advantage of external tools for the module in charge of sectioning text, but in the end found that the problem was a particular, yet-to-be-invented wheel. Specifically, the Group ran projects such as Hearst's TextTiling and Choi's C99, as well as an in-house segmenter from Columbia's Natural Language Processing group. The predominant problem with these external tools was that they assumed that every part of the text *must* be in a segment, and that no paragraph can be in more than *one* segment. (They are specialized for finding topic changes in news articles and the like.) Neither of these assumptions hold true for CLiMB: not only can a paragraph be "about" more than one TOI, but it is almost certain that some paragraphs in the text will not be concerned with *any* of a collection's TOIs. Thus, the Group implemented its own tool that uses the output of the TOI-Finder—individual referent noun phrases—as "seeds" from which associational context grows.

TOI import. To meet the need of allowing users to import TOIs from any source, the Technical Group defined a data format recommendation known as the TOI Interchange Format. This open specification, published among the documentation of the final CLiMB software package, informs users on how to create text files that CLiMB software will correctly interpret and use as TOIs. (The CLiMB Toolkit, discussed in Section 4, also allows users to specify TOIs manually over a Web form.)

Feature identification and Ruleset creation. The challenge in implementing this module, as mentioned previously, is that one cannot rely on any data sources that the user will not have available in practice. Thus, the group took a three-phase approach:

1) create an optimal Ruleset with optimal features for its testbed texts, using resources which may not always be available;
2) use the experience gained in the first phase to design an optimal Ruleset with the resources that are always available;
3) ensure that the software is capable of allowing users to introduce and use their own external resources flexibly.

The optimal Ruleset was based upon six features:

1) Proper/common noun: whether the noun phrase is proper or common.
2) TOI references: whether the noun phrase is a TOI reference.
3) Frequency ratio: whether the noun phrase occurs much more frequently in the text than in average English usage.
4) Back of book indexes: whether the noun phrase appears in the book's index.
5) Frequent usage: whether the noun phrase is one of the most common in English usage.
6) AAT hits: whether the noun phrase appears in the Getty Art & Architecture Thesaurus as an art term.

While these features gave good results, the CLiMB Default Ruleset (the result of the second phase) takes a slightly different shape:

- If a noun phrase is a proper noun phrase, then it should be somewhat more likely to be extracted as descriptive metadata.
- If a noun phrase is made of numbers only, then it should not be extracted as descriptive metadata.
- If a noun phrase is very short (less than 5 characters), then it should be somewhat less likely to be extracted as descriptive metadata.
- If a noun phrase is a reference to *any* TOI, then it should not be extracted as descriptive metadata.
- If a noun phrase is not within a section about a TOI, then it should be much less likely to be extracted as descriptive metadata.
- If a noun phrase is inclusive of numbers, then it should not be extracted as descriptive metadata.

This Ruleset does not take advantage of either the statistical model of common English usage or the AAT, both utilized in the "ideal" Ruleset. However, the software still allows users to employ one or both of these tools if they are able to procure them. It does this through a set of modules which deal with "Controlled Vocabularies." CLiMB uses this term in a broad sense. It can refer to a pre-existing external authority list, back-of-the-book index, or any other list of terms created by the user. Once the user imports the list into the CLiMB software (using an interchange format similar to the TOI Interchange Format described earlier), he or she can find occurrences of terms on the list in the text, and then write a Rule in the Ruleset to deal with such hits.

This is how CLiMB software implements stoplists (terms which the user would like to prevent from being extracted). To create a stoplist, the user simply composes a list, imports it into CLiMB as a Controlled Vocabulary, and creates a Rule that states that terms on the list should "never be extracted as descriptive metadata."

Output formats. Once the user runs a Ruleset, CLiMB can export the terms in one of three formats. In addition to a CLiMB-native format, one can also take away a file in MARC (MAchine-Readable Cataloging) format, or in VRA (Visual Resources Association) Core format.

Thus the Technical Group was able to implement a software prototype capable of concretely meeting the abstract needs of the metadata-extraction problem. However, the CLiMB software would not be complete until the Group wrote an elaborate, easy-to-use interface for it. For more on this front-end, which is known as the CLiMB Toolkit, see Section 4.

3. CLiMB Formative Evaluation

3.1 Introduction

In order to design technology to assist, or perhaps ultimately to replicate the human process of selecting descriptive metadata for images, we must understand more about that selective process. For this reason CLiMB conducted a formative evaluation of the relationship between images, terms humans select to describe the images under a variety of conditions, and texts about the images. Three questions guided this evaluation. The first was whether we could develop a better understanding of how humans assign descriptive terms to images, and what differences might arise in the terms humans propose when they have an image search request, when they have a specific image, when they have texts about images, or when they have a list of terms. The second question was whether we could develop informed hypotheses about the potential impact of computational linguistic tools in support of the human process of culling subject metadata from texts. Third, we asked whether we could use observations derived from pursuing the first two questions to refine the CLiMB tools.

To begin to answer these questions, we first invited expert catalogers, image professionals, and computer scientists to participate in an initial formative evaluation, held on October 17, 2003, at Columbia University. All attendees were given one of two surveys that differed only in the order of materials. After completing the survey, attendees engaged in a discussion about the nature of the CLiMB task, and the types of solutions computational linguistic techniques might provide. The attendees and survey materials are described in more detail in Section 3.2 below, with detailed results in Section 3.3. Here we give an overview of the results.

Regarding the first question, humans were able to find more terms, and more specific terms, when presented with a checklist of CLiMB output, or with the original texts CLiMB processed, than when they had the image alone. This is unsurprising given that the images in question, which were taken from the Greene & Greene and Chinese Paper Gods collections, were unfamiliar; that is, the attendees had little relevant expertise for describing those images. In addition, catalogers and image professionals proposed more terms, and more terms that agreed with each other, than the computer scientists. An image of a cultural object that is well known requires less expertise to describe than an image of an unknown object. Furthermore, there are potentially numerous descriptors that could be assigned to any one image.

Regarding the second question, terms selected by humans were ranked on the basis of how many of the attendees proposed the term. CLiMB terms themselves are automatically associated with a rank, depending on how many of the possible factors discussed in Section 2.2 were associated with the term, and with what strength. Although there was no direct statistical correlation between CLiMB ranks and human ranks, we performed some exploratory analyses by cross-tabulating high-, mid-, and low-ranking CLiMB terms with terms selected by many, some or few of the human evaluators, using various thresholds. Results are in presented in Section 3.3.

As a result of the formative evaluation data, and our analysis of the relationship between terms that humans prefer from CLiMB checklists versus terms that humans select directly from discursive texts, we further refined the CLiMB rulesets that ultimately became integrated in the Toolkit, as described in Section 4. This study also served as a pilot for developing a more precise model of how

to enhance CLiMB technology through a similar evaluation conducted on a large scale and with a more homogenous group of experts. This is discussed in detail in Section 5.

3.2 Evaluation Method

3.21 Participants and Method

The formative evaluation of the metadata task was conducted as part of a multidisciplinary meeting of experts assembled by the CLiMB team. Attendees were given a paper survey, and were provided with time to complete the survey items. After completing the survey, respondents broke into two groups to discuss the task in the context of imagining a software application or environment in support of the task.

Thirteen attendees completed the survey materials for the evaluation. Table 1 gives the breakdown of those surveyed by background, indicating that the majority of respondents had expertise directly relevant to images or library metadata.

Background	**Total**	**Survey # Identifiers**
Art librarian/historian or librarian with some experience with image collections	8	44, 45, 50, 52, 53, 54, 55, 56
Computer scientist/engineer	5	40, 41, 42, 43, 51

Table 3.1: Breakdown of formative evaluation survey respondents by background.

3.22 Survey Materials

The survey questions were distributed across four tasks that were presented in different orders to three different groups of respondents. The two CLiMB collections that were used for the survey were the Greene & Greene Collection of Architectural Records and Papers, Avery Architectural and Fine Arts Library (henceforth G&G); and the Anne S. Goodrich Collection of Chinese Paper Gods, C.V. Starr East Asian Library (henceforth CPG). The four tasks are as follows:

1) **User Scenario**: In this task, the survey item contained the following hypothetical user scenarios. Respondents were asked to list keywords and phrases that could be used "to search for relevant images in an image database."
   a. I am writing a paper on domestic architecture in Southern California in the early part of the 20th century. I was told that there are homes with exteriors clad in a type of concrete or cement. How can I locate images?
   b. I am trying to locate an image of the Buddhist goddess of compassion. I can't remember the name but I know this deity, widely worshipped by women in China, originated as a male figure in India. She is often portrayed wearing a headdress, attended by other figures, and often some type of plant is depicted. Can you help me find a picture?
2) **Image**: This survey item contained an image. Respondents were given the following instructions: "Please write keywords and phrases that you would use to find this image in a database. You may write as many as you wish."

3) **Free Text**: This task contained a passage from one of the texts associated with G&G or CPG. Respondents were asked to "Suppose there is a collection of related images that needs metadata keywords and phrases. Please select the words and phrases in this text that you feel would be good metadata for the images.
   a. Please circle 10 words or phrases as your top choices
   b. Please underline 10 as your second tier choices."
4) **CLiMB Checklist**: Respondents were given a long list of words and phrases (194 G&G entries and 117 CPG entries) that had been extracted by CLiMB tools from the same texts presented in Task 3. Instructions were: "Please check off the words and phrases that you feel would be suitable metadata for the images in the collection."

Each of these four tasks contained both a G&G and a CPG component, meaning that the survey had eight total items. These eight survey items were presented in three different orders to various participants, as indicated in **Table 3.2**. Although we had too few respondents to fully investigate the effects of this ordering, it is consistent with the goal of conducting an exploratory pilot study that a tendency toward group differences might engender hypotheses that could be examined more systematically in future work.

Order	Group I: 40–45	Group II: 50–52	Group III: 53–56
1	G&G 3. Free Text	G&G 1. User Scenario	G&G 1. User Scenario
2	CPG 3. Free Text	G&G 3. Free Text	CPG 1. User Scenario
3	G&G 1. User Scenario	CPG 1. User Scenario	G&G 3. Free Text
4	CPG 1. User Scenario	CPG 3. Free Text	CPG 3. Free Text
5	G&G 4. CLiMB Checklist	G&G 2. Image	G&G 2. Image
6	CPG 4. CLiMB Checklist	G&G 4. CLiMB Checklist	CPG 2. Image
7	G&G 2. Image	CPG 2. Image	G&G 4. CLiMB Checklist
8	CPG 2. Image	CPG 4. CLiMB Checklist	CPG 4. CLiMB Checklist

Table 3.2: Three survey groups, differing by order of items.

For each noun phrase in a text, the CLiMB tools assign a numeric score, which is the sum of the weighted values assigned by each individual rule. Initially, the weights assigned in this manner were arrived at by hypothesizing the relative contribution of each rule, and by hypothesizing a linear relationship among rules. One purpose of the evaluation was to investigate whether a body of metadata collected from humans could be used to fine-tune the weights assigned to each rule individually, or the manner in which rules were combined.

3.3 Results

This section considers results primarily for G&G. The most informative results pertained to survey items 3 and 4, and exploratory analysis of these items turned out to be time-consuming.

Fewer terms were returned for this user scenario survey item than for other items. This reflects the relative lack of specific information from which to construct a query. Further anecdotal evidence that more information could lead to more terms in this item is the observation that in item 4

(CLiMB Checklist), three of the respondents (41, 53, 54) referred directly or indirectly to the "reference interview" that would be conducted as part of the "user scenario." For example, with regard to the G&G scenario, one wrote a note indicating that specific cities might be elicited from the user, then included as search terms; another wrote a note indicating that specific texts or knowledge about the works of specific architects could be recommended to the user, who could then return with more specifics.

As indicated in the description above, respondents were instructed to list words and phrases for a keyword search. The page for this item in the survey had a layout with seven bulleted lines where respondents could list their responses. To some degree, this predisposes respondents to try to produce something on the order of half a dozen terms. However, the instructions and layout were interpreted somewhat differently by different respondents. Some provided a list of individual terms, the implication being that any combination of these could be used in a Boolean search. Some provided lists of strings of terms, occasionally with minor variations within the list, the implication being that other possible combinations of the same terms were deliberately omitted.

On average, respondents found 10 terms for G&G (standard deviation = 3.3) and 7.6 terms for CPG (standard deviation = 3.6). There was a slight tendency for respondents in Group II to find the most terms for G&G, and for Group I to find the fewest; the same pattern, but with a greater disparity, occurred for CPG. It is not clear what to conclude from this, other than that respondents seemed to try harder to find more terms for item 1 if it (which was always followed by item 2) was the first item for a collection (Group II), or if both G&G and CPG item 1 preceded G&G and CPG item 3 (Group III).

We looked at exact matches of terms, with the exception of plurals (for instance, "house" and "houses"—see the third row in **Table 3.3**), and found 82 distinct terms used for G&G, and 56 for CPG. Most terms were suggested by at most a single respondent; **Table 3.3** shows those terms that were suggested by multiple respondents for G&G.

Respondents	Term
8	cement
8	concrete
7	house(s)
6	southern California
6	home(s)
5	California
5	domestic architecture
4	20th century
4	exterior
3	architecture
2	20th century domestic architecture
2	Southern California architecture
2	exterior cladding
2	images

Table 3.3: Terms suggested by multiple respondents for item 1, G&G.

The image and free text exploratory items gave results that showed how difficult users found generation of items de novo. In other words, the value of using text already describing an image was shown.

The conclusions of this formative evaluation showed that experts selected terms differently from non-experts, thus helping the CLiMB team to narrow our future testing on catalogers and image experts. It also showed that proper nouns hold a particularly important role, and that selected common nouns, especially if related to the AAT, were likely to be selected. The results of this evaluation formed the basis for our next evaluation.

4. CLiMB Toolkit

4.1 Specifications

Toward the end of the second year of CLiMB, many of the technical pieces of the software suite described in Section 3 were nearing completion. However, it took a great deal of technical expertise to use them. One had to know a particular syntax for constructing text commands to drive the various modules. In order to fulfill CLiMB's mission of a tool for visual resource professionals with lay technical expertise, the Technical Group had to devote significant resources toward building a "user-friendly" front-end interface for the modules.

Development of this front-end, which (in conjunction with underlying CLiMB software) is known as the CLiMB Toolkit, dominated the last six months of the Technical Group's work. The project began with the composition of specifications for this major CLiMB deliverable: what inputs and outputs it should support, and what major functions it should offer. The Group decided that the "pipeline" for the user experience it previously identified (see Section 2.2) would persist as a process flow for the Toolkit. It also made certain to fully document the Toolkit, both within the interface (user help) and without (development notes for future work with the code base).

The Group decided to implement the Toolkit as a Web-based application, operated from within a standard browser. This approach had three major advantages: it allowed users to use an application with which they already feel comfortable, it could be implemented without any rewriting of the back-end modules, and it allowed users to access the Toolkit remotely (without needing to install it on their own computers). The latter reason was particularly trenchant when the Group needed to provide access to the Toolkit in remote locations (such as conferences) or to many simultaneous users (such as at the second External Advisory Board meeting).

The Group completed the Toolkit shortly before the second Advisory Board meeting. Work on documentation and packaging continued until the official release of CLiMB v.1.0 on July 29, 2004.

4.2 A Tour of the CLiMB Toolkit

The following is a graphical tour of the Toolkit, with associated explanations of the Toolkit's functionality. Our scenario is the extraction of descriptive metadata from a set of image records from the Handbook of the North Carolina Museum of Art (see Chapter 1).

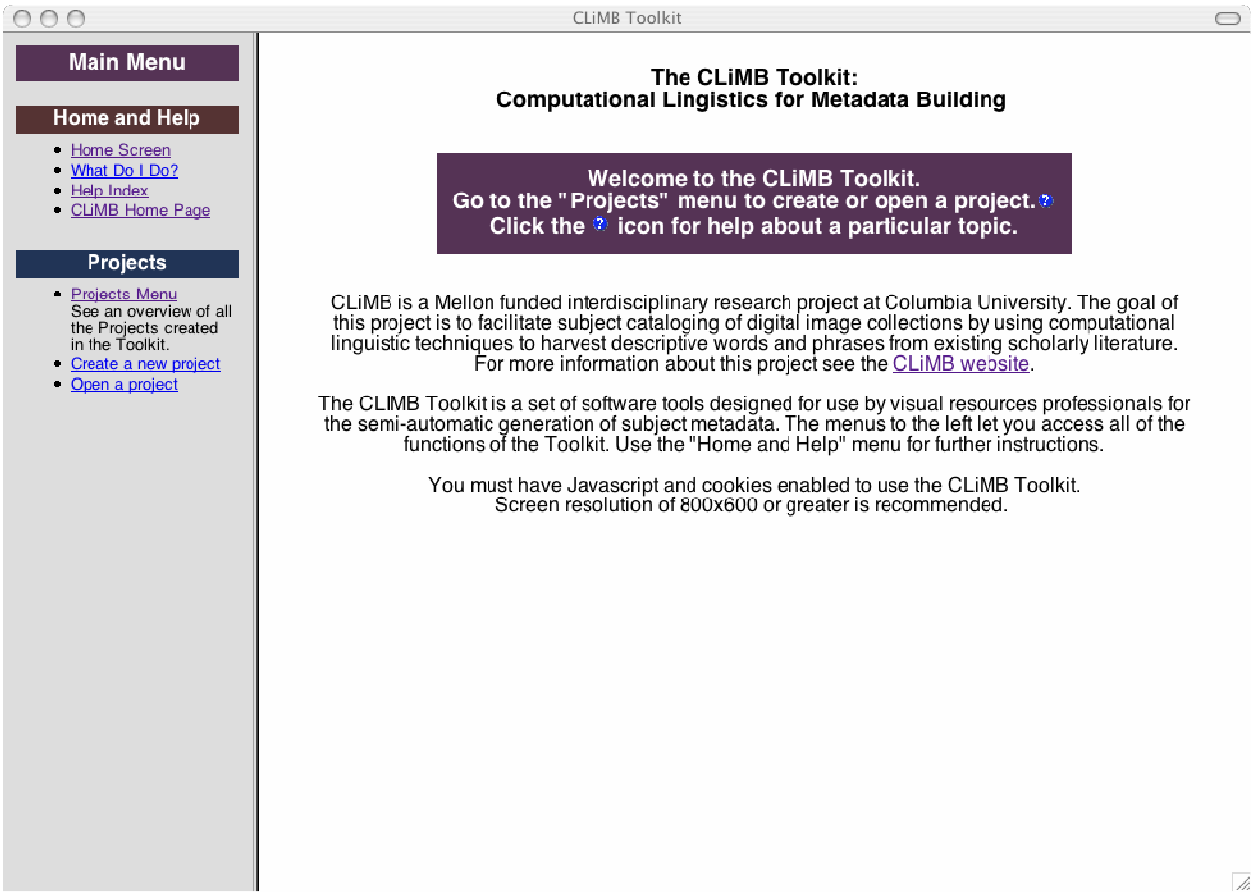


Figure 4.1 Toolkit Front Page

When one first accesses the Toolkit's front page, it appears as in Fig. 4.1.

The screen is divided into two areas: one for navigation, and one for content. A short introductory section introduces the Toolkit and its purpose. Users are invited to open, close, or create a new "project." The term refers to a set of texts, authority lists, TOIs, rules, and other materials associated with a particular image collection, as well as the current status of the user's work with these materials. A single project represents the process of extracting descriptive metadata from one or more texts for an image collection. Thus, the first objective is to "create a new project."

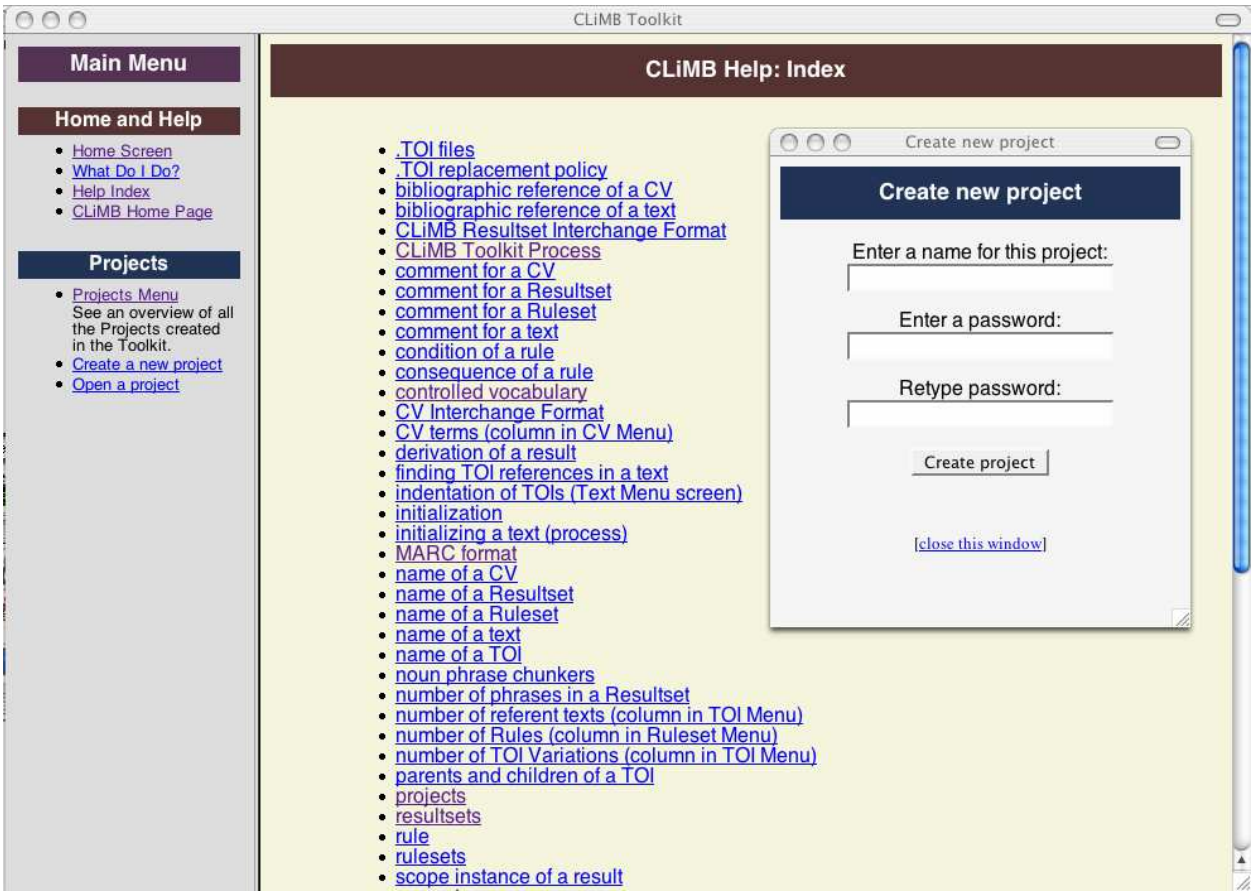


Figure 4.2 New Project and Help Page

Figure 4.2 shows the form allowing us to create a password-protected project, as well some of the index of available online Help articles.

We create a project "NCMA Records" and the screen changes to reveal all the functionality we can access within the context of an open project. The next step is to import our text into the Toolkit, which is made possible by a window allowing us to upload files from our computer. See Figure 4.3.

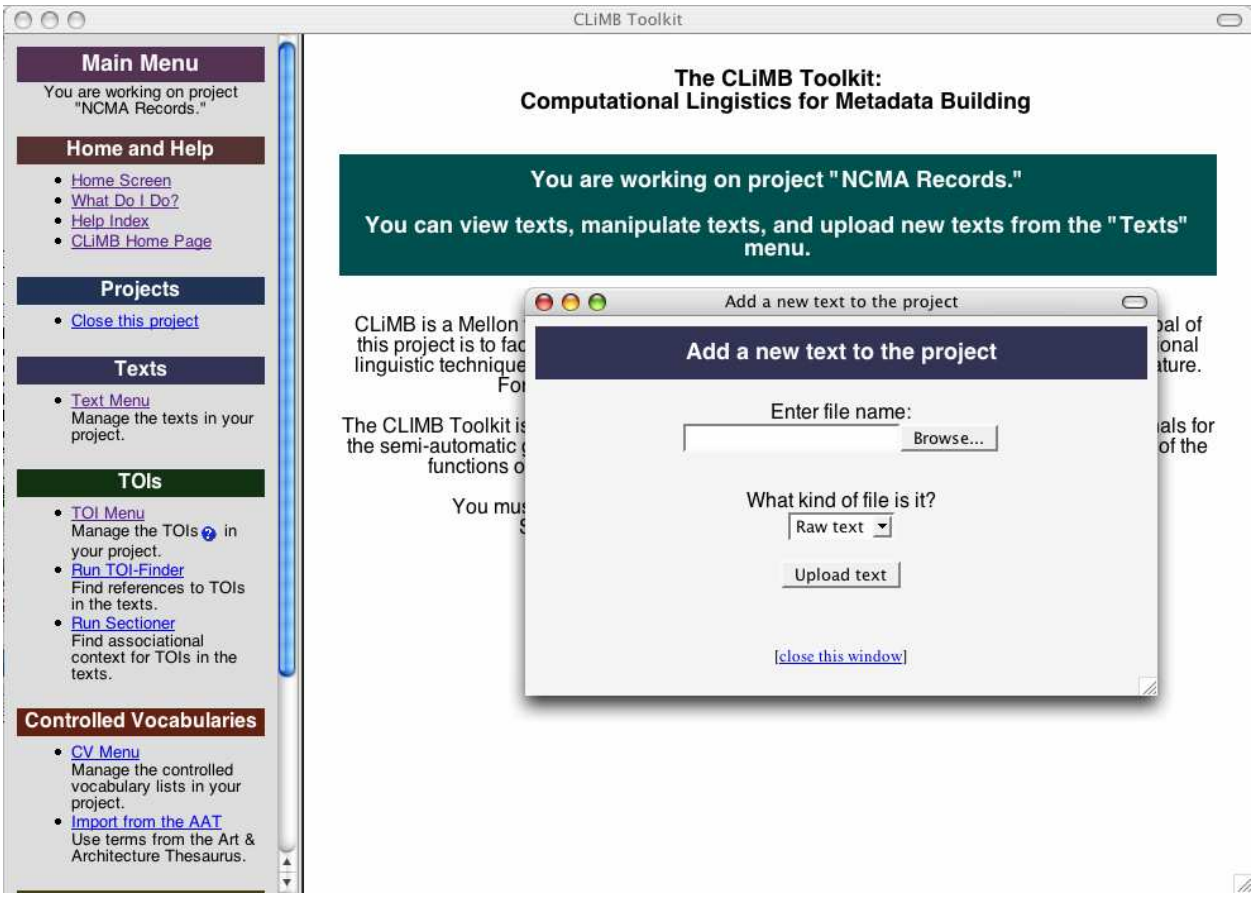


Figure 4.3 Upload Text

We upload the records of the NCMA collection as a single file in TEI XML (see Section 3.3). The software allows us to add metadata about the text such as bibliographic information and comments. It also informs us that we need to "initialize" it, i.e., impose the additional restrictions of CLiMB-TEI XML (see Section 3.3), before we can process it.

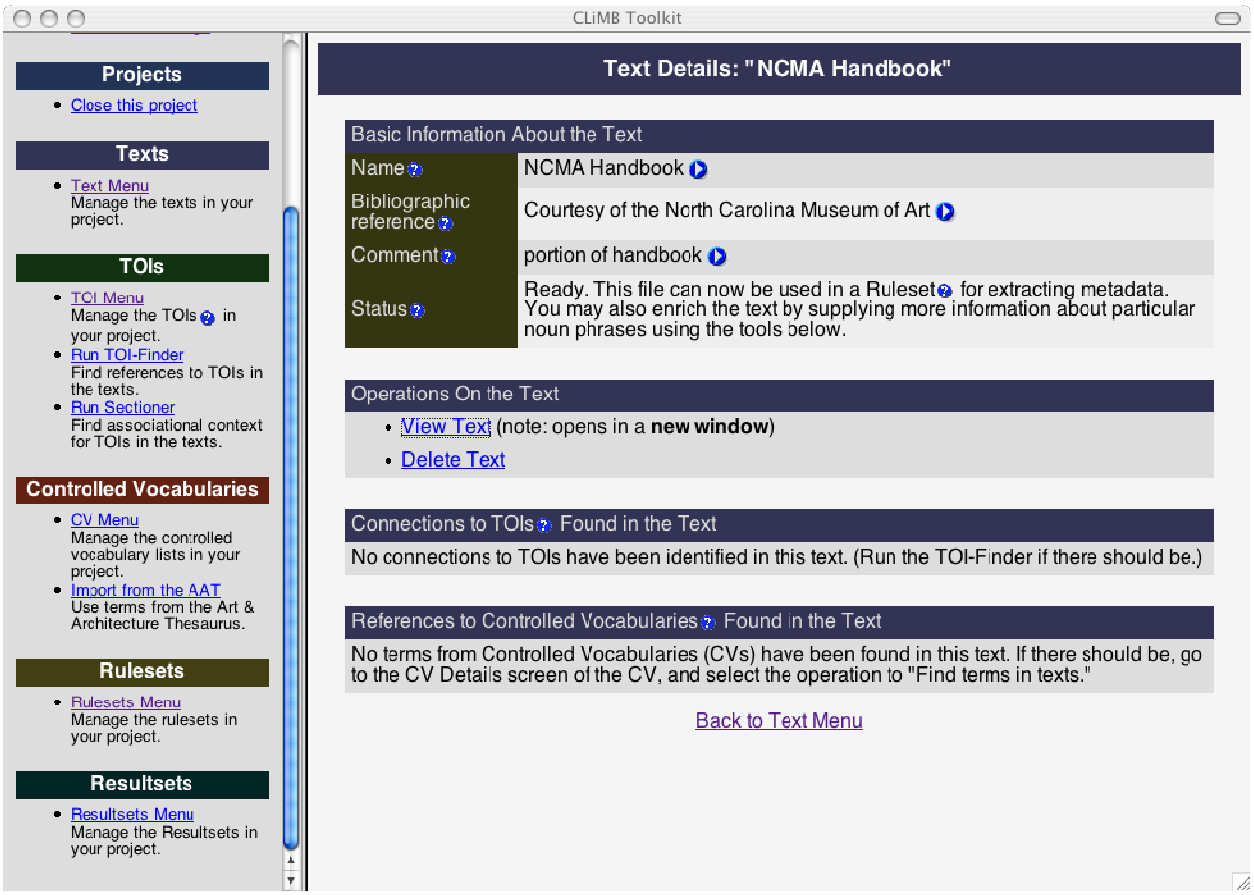


Figure 4.4 Text Details

Once the text is initialized, we must run LT Chunk, the noun phrase finder we employed for identifying the "candidate" noun phrases for extraction. After that happens, we see a screen as in Figure 4.4, which shows that the text is "ready" to be used for extraction.

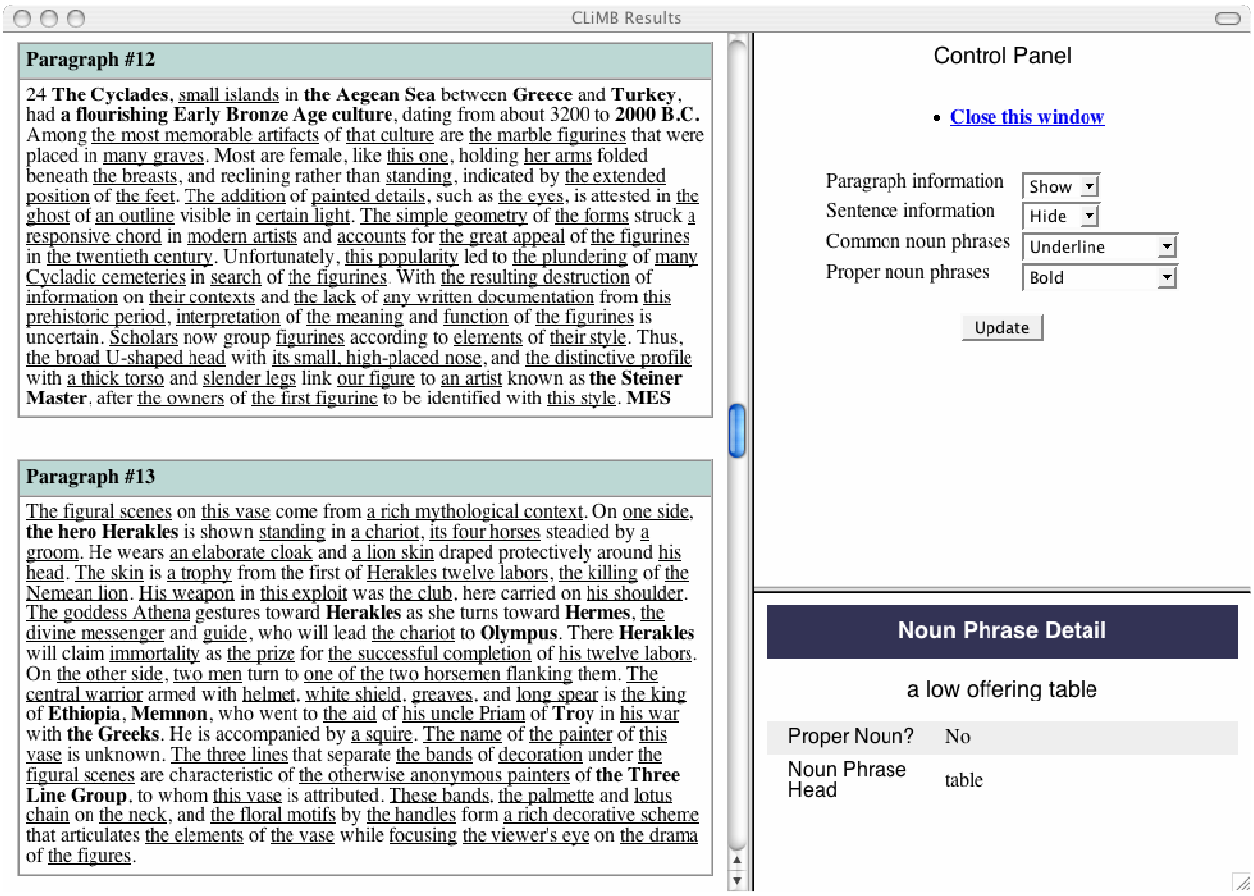
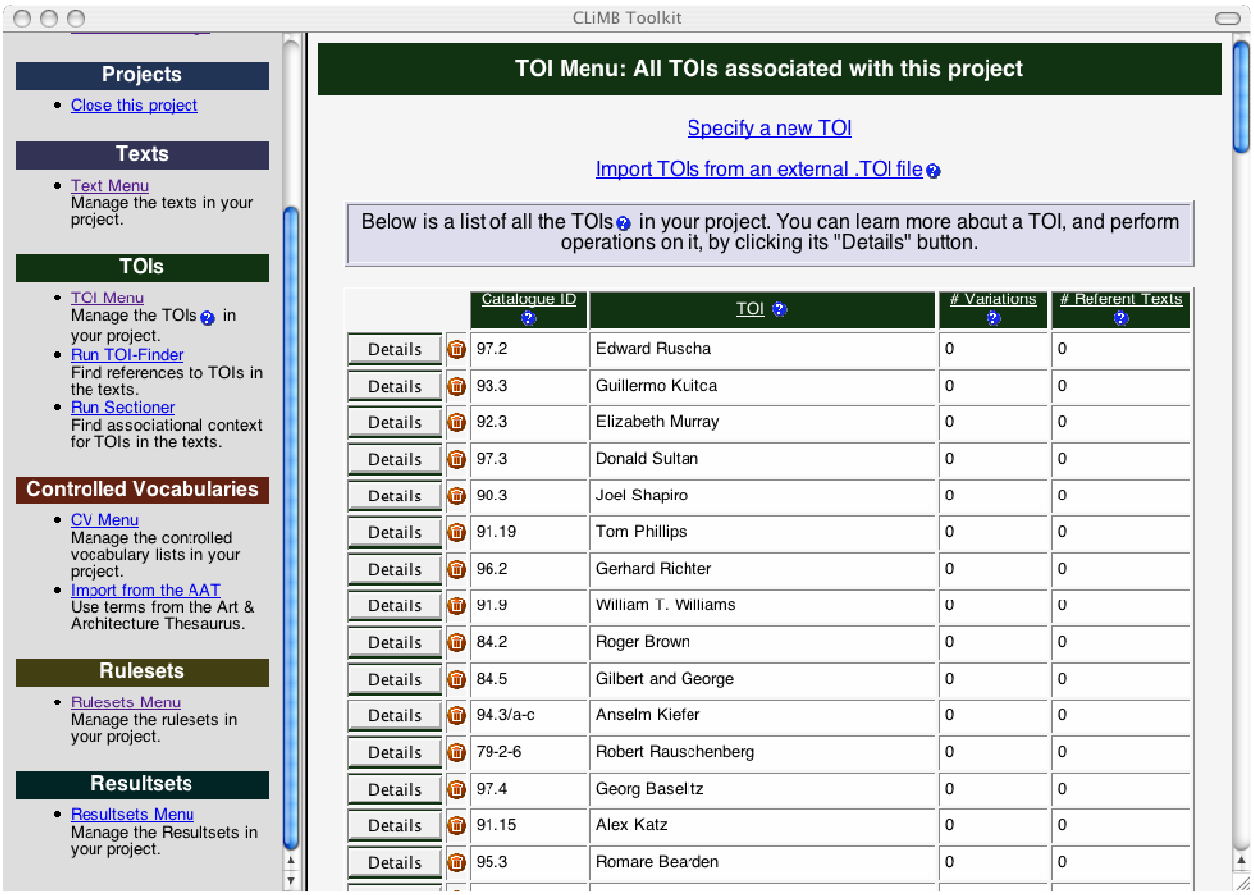


Figure 4.5 Chunked Text

We can examine the text at this point using a visualization tool the Technical Group calls the "Docviewer." (The Docviewer was implemented in the first year of the project, and adapted for use within the Toolkit in the second year.) It allows us to see the text with various permutations of content-based markup. In the case of Figure 4.5, we see the way the text has been separated out into paragraphs (corresponding to Handbook records), and we also see the noun phrases from which CLiMB will select descriptive metadata. Those that appear to be common noun phrases are underlined, whereas those that appear to be proper are in bold. We can also see all available information about a particular noun phrase with the Detail Panel on the lower right.



	Catalogue ID	TOI	# Variations	# Referent Texts
Details	97.2	Edward Ruscha	0	0
Details	93.3	Guillermo Kuitca	0	0
Details	92.3	Elizabeth Murray	0	0
Details	97.3	Donald Sultan	0	0
Details	90.3	Joel Shapiro	0	0
Details	91.19	Tom Phillips	0	0
Details	96.2	Gerhard Richter	0	0
Details	91.9	William T. Williams	0	0
Details	84.2	Roger Brown	0	0
Details	84.5	Gilbert and George	0	0
Details	94.3/a-c	Anselm Kiefer	0	0
Details	79-2-6	Robert Rauschenberg	0	0
Details	97.4	Georg Baselitz	0	0
Details	91.15	Alex Katz	0	0
Details	95.3	Romare Bearden	0	0

Figure 4.6 TOI List

The next step is to introduce TOIs into the project. We can upload a list of TOIs from an external file in the TOI Interchange Format (see Section 2.3) using a function similar to the one with which we uploaded a text. In this case, we have previously obtained the list of artists featured in the NCMA and transformed it into the Interchange Format. Now, once we have imported the TOIs, we can view them individually, or in list form as in Figure 4.6. Note the inclusion of NCMA accession numbers (as "catalog IDs"), which allow these TOIs to be cross-referenced elsewhere.

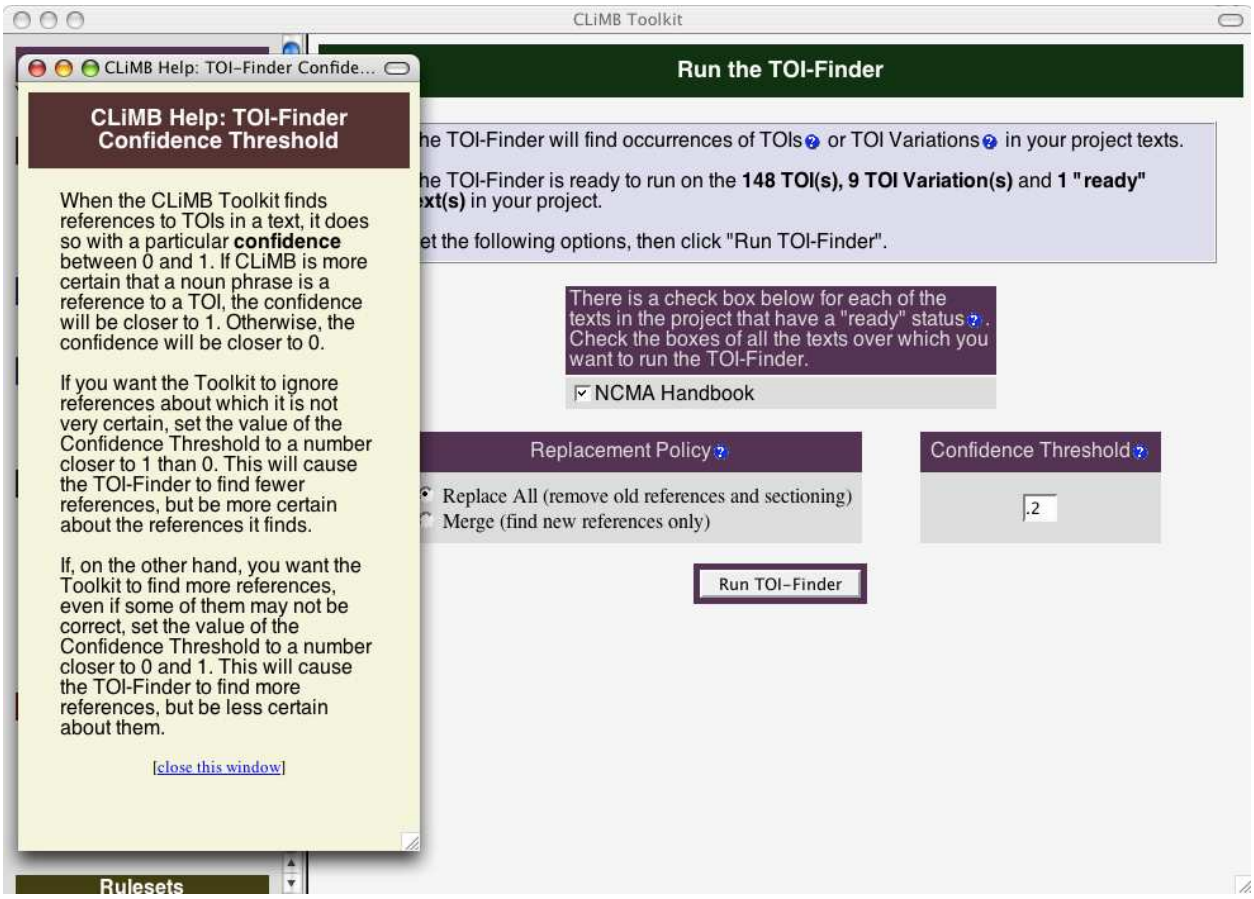


Figure 4.7 TOI Finder

We can now search for occurrences TOIs within the catalog text. Figure 4.7 shows the screen that prepares this function, as well as a relevant window from the online Help.

Next, we run the Sectioner, which is separate module from the TOI-Finder but builds upon its output. The Sectioner uses the TOI occurrences to determine which paragraphs in the Handbook are "about" which TOIs. See Figure 4.9 for the Sectioner's results.

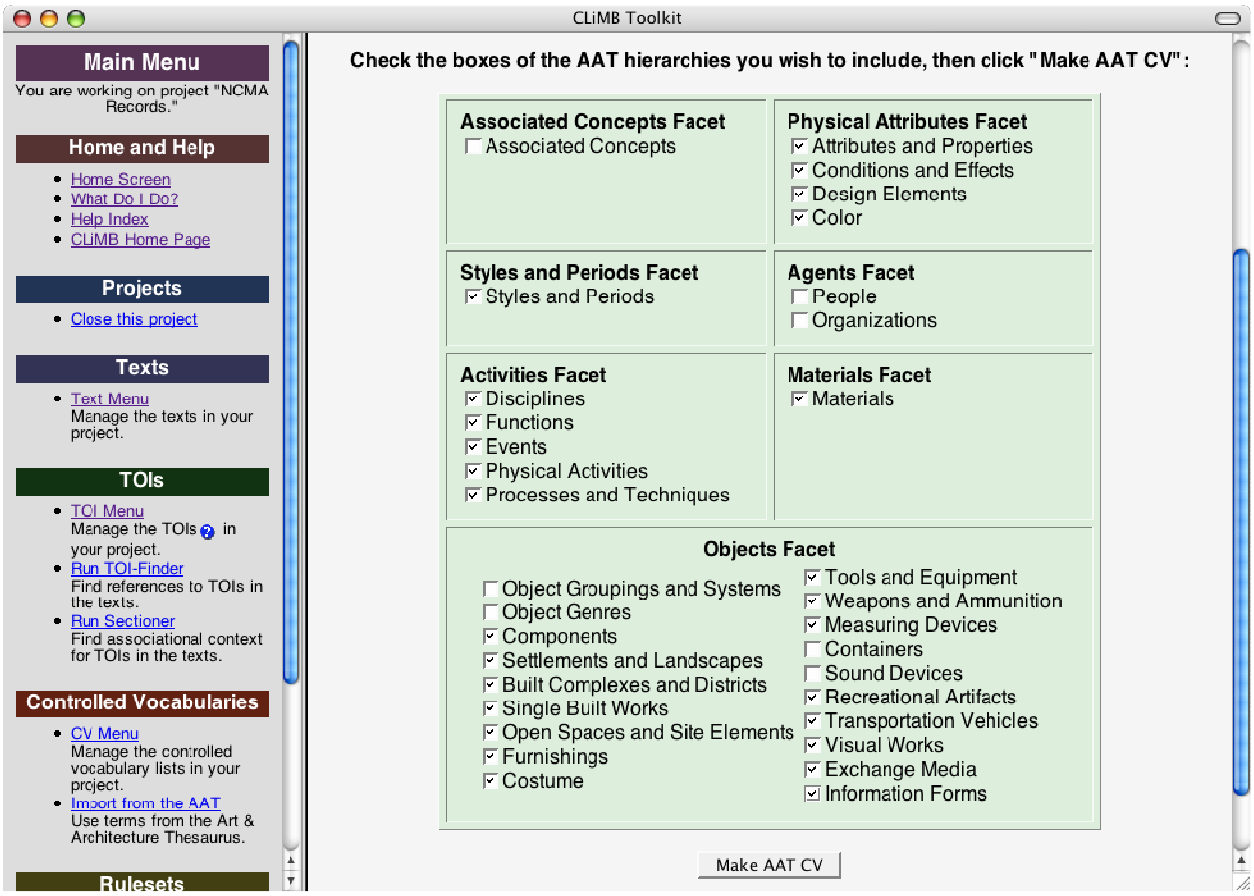


Figure 4.8 Hierarchy Selection

In order to help the Toolkit determine which noun phrases are most suitable for extraction, we have to give it a sense of what kinds of words we are looking for. Because the metadata will be descriptive of art objects, we hypothesize that terms that appear in the Getty's Art & Architecture Thesaurus are better than terms that do not. With the Getty's permission, we have incorporated the AAT into the Toolkit, where we can introduce it as a Controlled Vocabulary (CV) list (see Section 3.3). Figure 4.8 shows the process of selecting certain subsets of the AAT, known as "hierarchies," for inclusion in the CV.

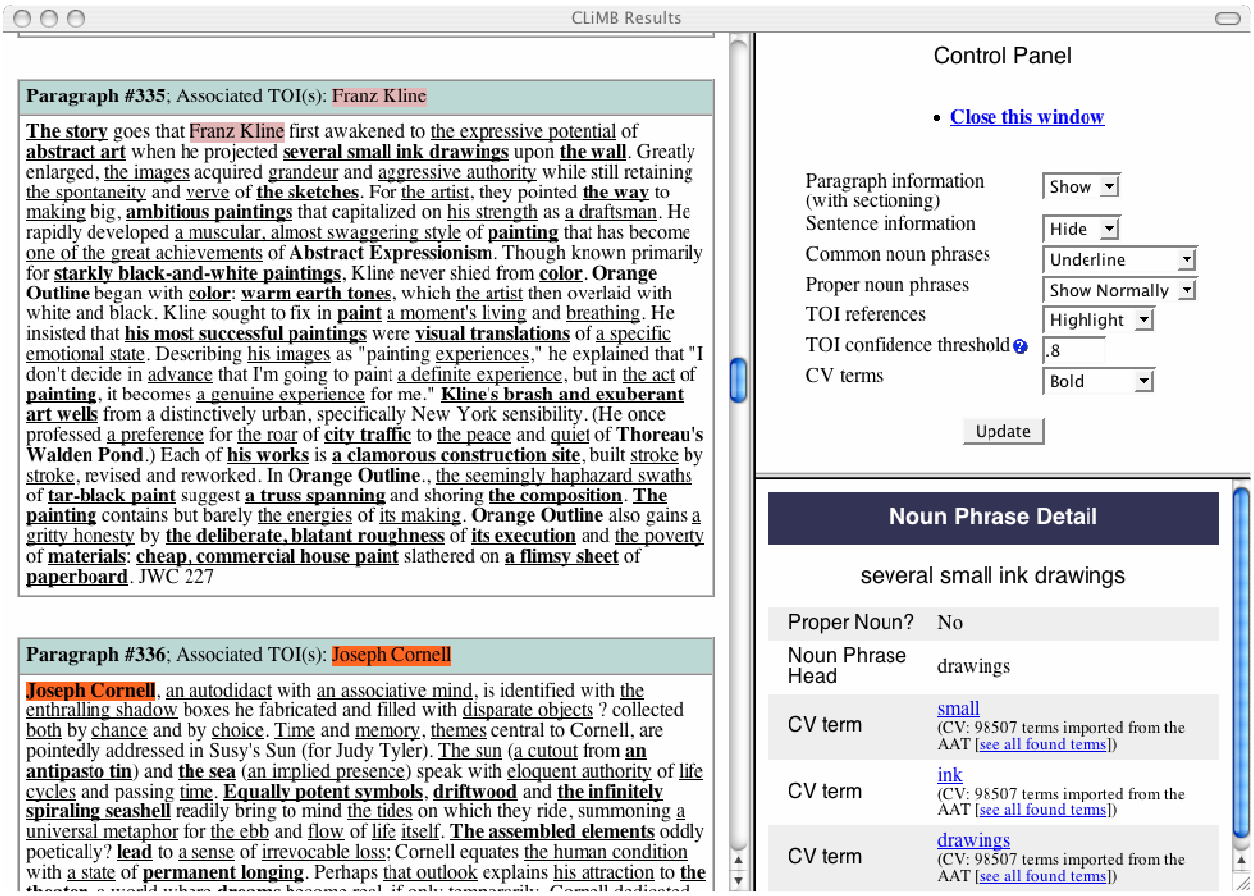


Figure 4.9 Sectioner Results

We run a function that finds the AAT's terms in the text, and then reload the Docviewer. Now, the Docviewer shows additional controls for viewing TOIs and CVs. In Figure 4.9, we have set the Docviewer to highlight TOIs, put AAT terms in bold and underline common noun phrases. By noticing the intersections between the types of markup, we can get a sense of how the CLiMB Rulesets (i.e., formulas for extraction) will choose the best phrases for extraction. Note the results of the Sectioner in each paragraph header. Additionally, we can now use the Detail Panel on the lower-right to see which AAT terms or TOI hits were found for a particular noun phrase.

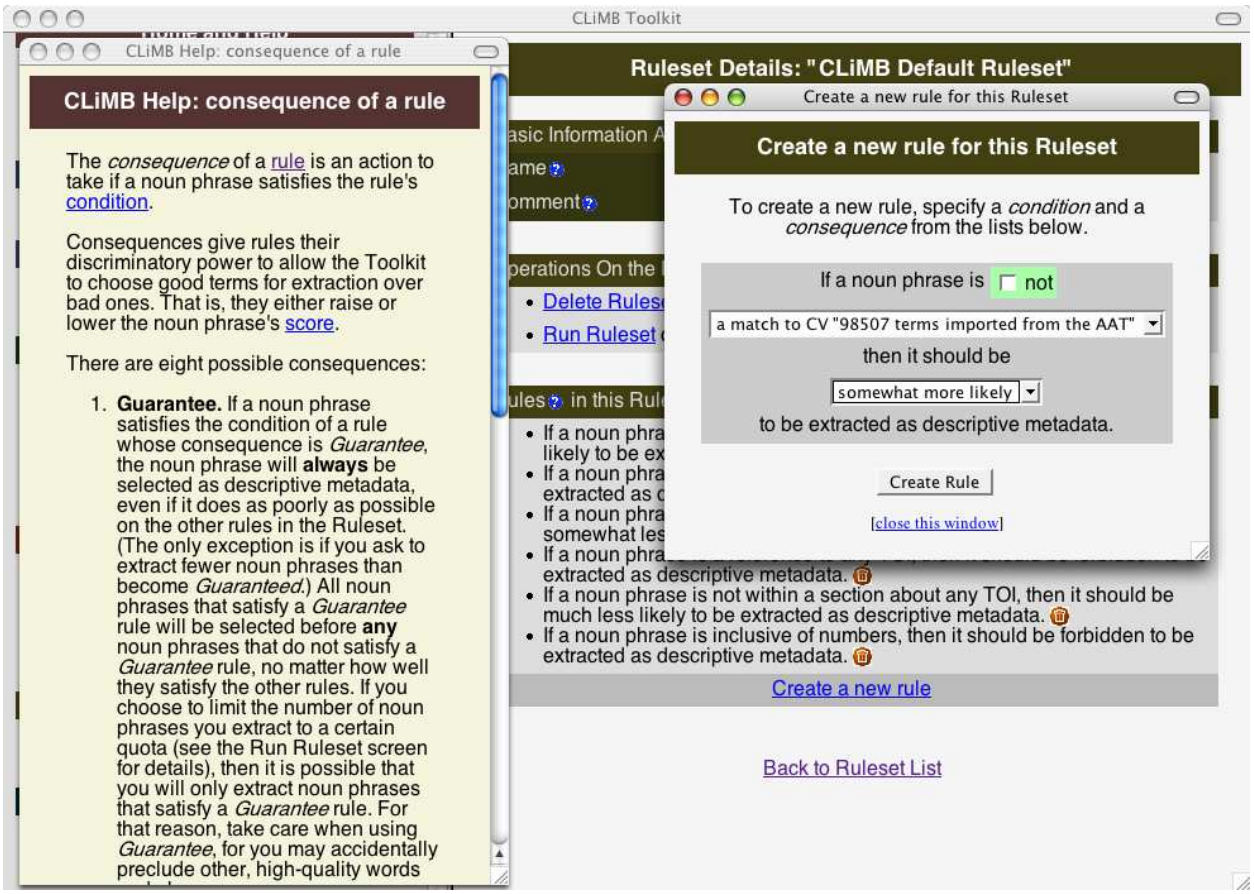


Figure 4.10 New Ruleset Rule

We're now ready to have the Toolkit use the information we have provided to select words and phrases. This is done with Rulesets that consider how "likely for extraction" a noun phrase should be if it satisfies certain conditions (such as being proper). We begin with the CLiMB Default Ruleset (see Section 3.3), and, with the assistance of a Help screen, create a new Rule to take advantage of the AAT. See Figure 4.10.

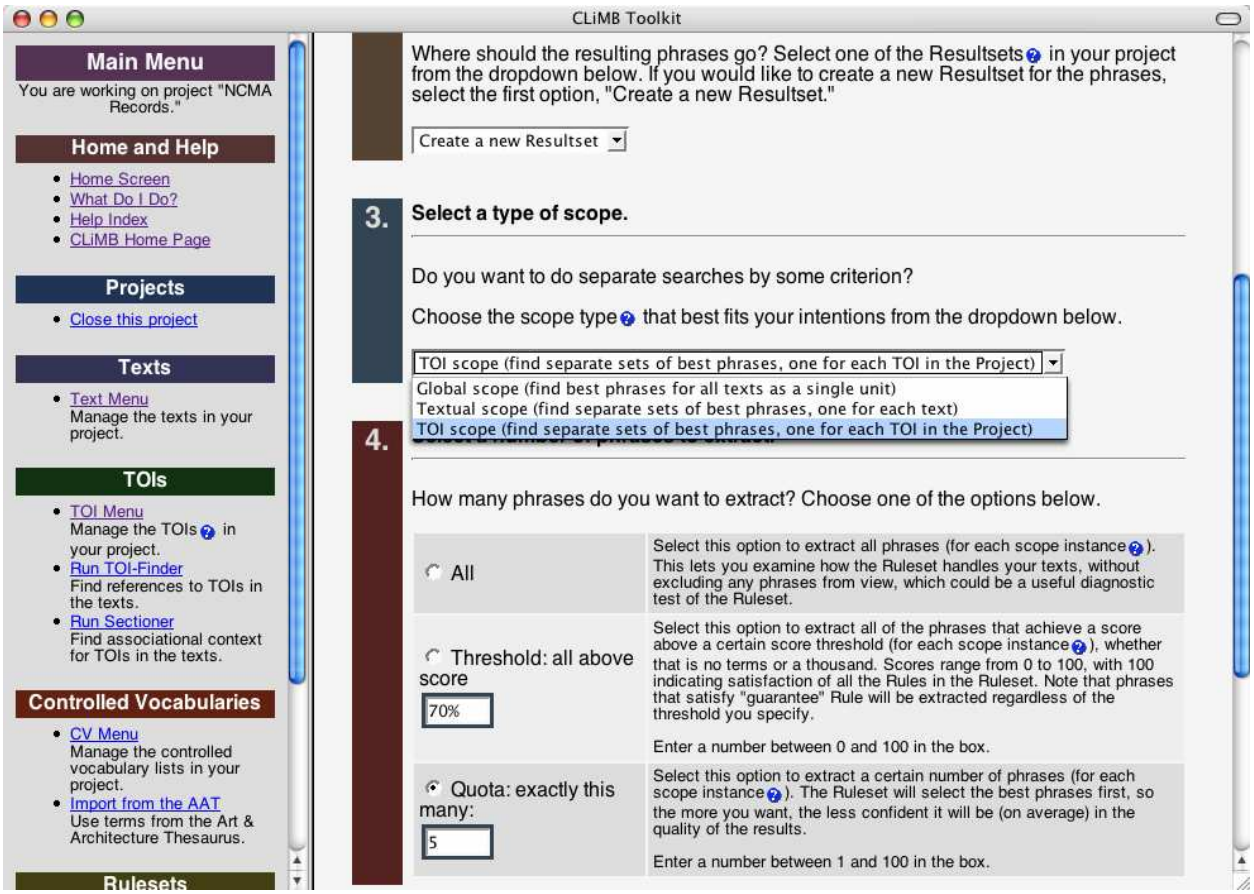


Figure 4.11 Ruleset Configuration

Now we are ready to "run" the Ruleset, i.e., have the Toolkit decide in practice which noun phrases to extract. This operation can be run with several different configurations; see Figure 4.11. For example, we can decide to extract all the noun phrases that perform above a certain threshold, or instead, fill a quota of top-performing candidates.

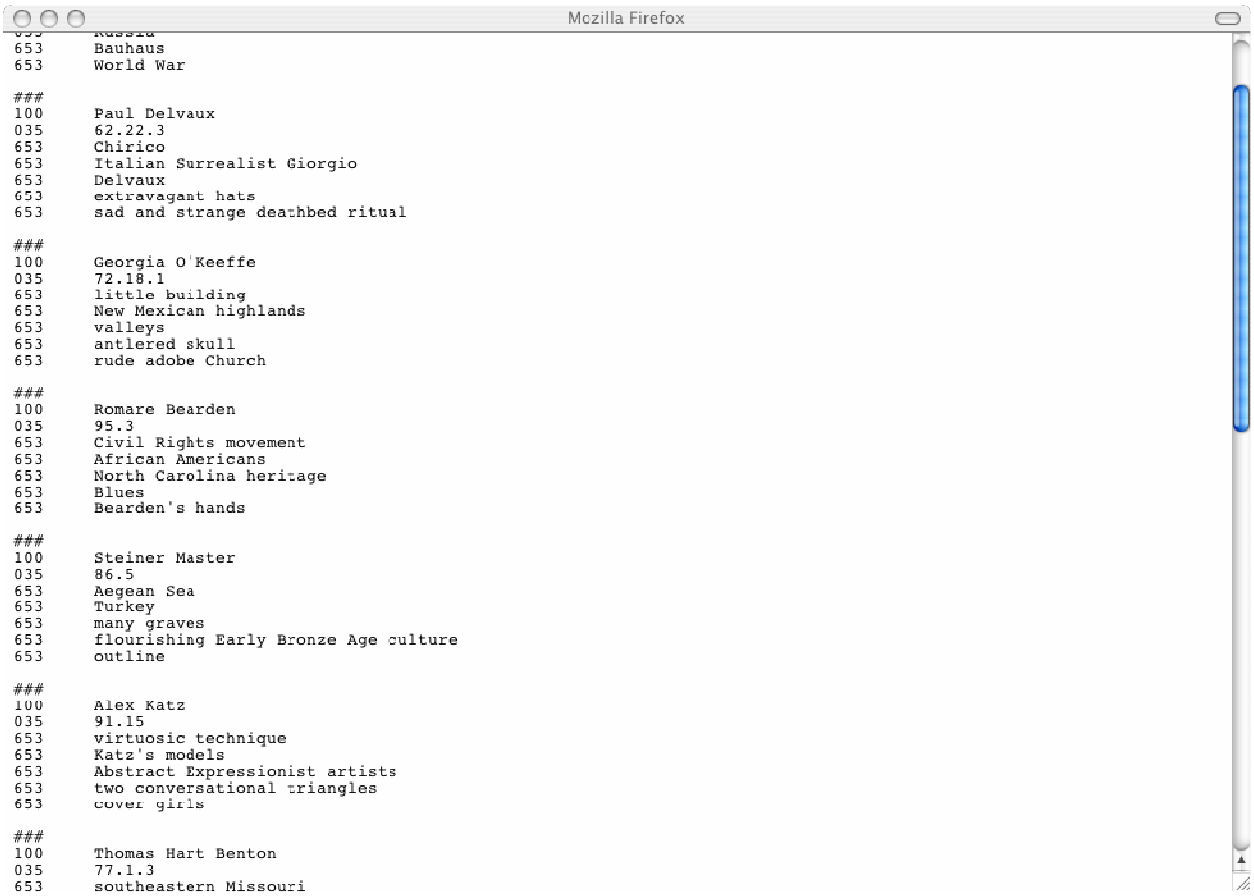


Figure 4.12 Updated MARC Record

We run the Ruleset so that it takes the top five highest-performing candidate noun phrases for each separate TOI. The results go in a "Resultset," a list of terms, which we can review and edit if necessary. Finally, we export the descriptive metadata from the Toolkit. The application supports two open standards: MARC and that of the VRA Core. Figure 4.12 shows new MARC records with "100" fields ("personal names," i.e., TOIs), "035" fields (accession number, i.e., the TOI "catalog IDs") and "653" fields (subject metadata, i.e., words and phrases from the text).

In the end, we have enriched the NCMA catalog records with appropriate words and phrases from the natural-language Handbook.

5. CLiMB Toolkit Evaluation

5.1 Introduction

On April 14, 2004, CLiMB conducted an evaluation of the CLiMB Toolkit prototype Version 0.9. This version of the Toolkit fully implements the first two stages of processing ("Load Text" and "Load TOI List") illustrated in **Figure 5.1** below, as well as parts of the third stage ("Analyze Text"). Details on Version 0.9 appear in Section 5.2. The purpose of this evaluation was to elicit feedback from two types of experts, image professionals and art librarians, at a stage early enough for us use in subsequent design stages, and late enough to provide a realistic view of the Toolkit environment and functionality. We sought feedback on three aspects of the Toolkit: the general habitability of the interface [Gibson, 1977; Norman, 1988]; specific functions pertaining to the task of interactively selecting metadata; and an initial, indirect view of the quality of the metadata provided by the evaluation.

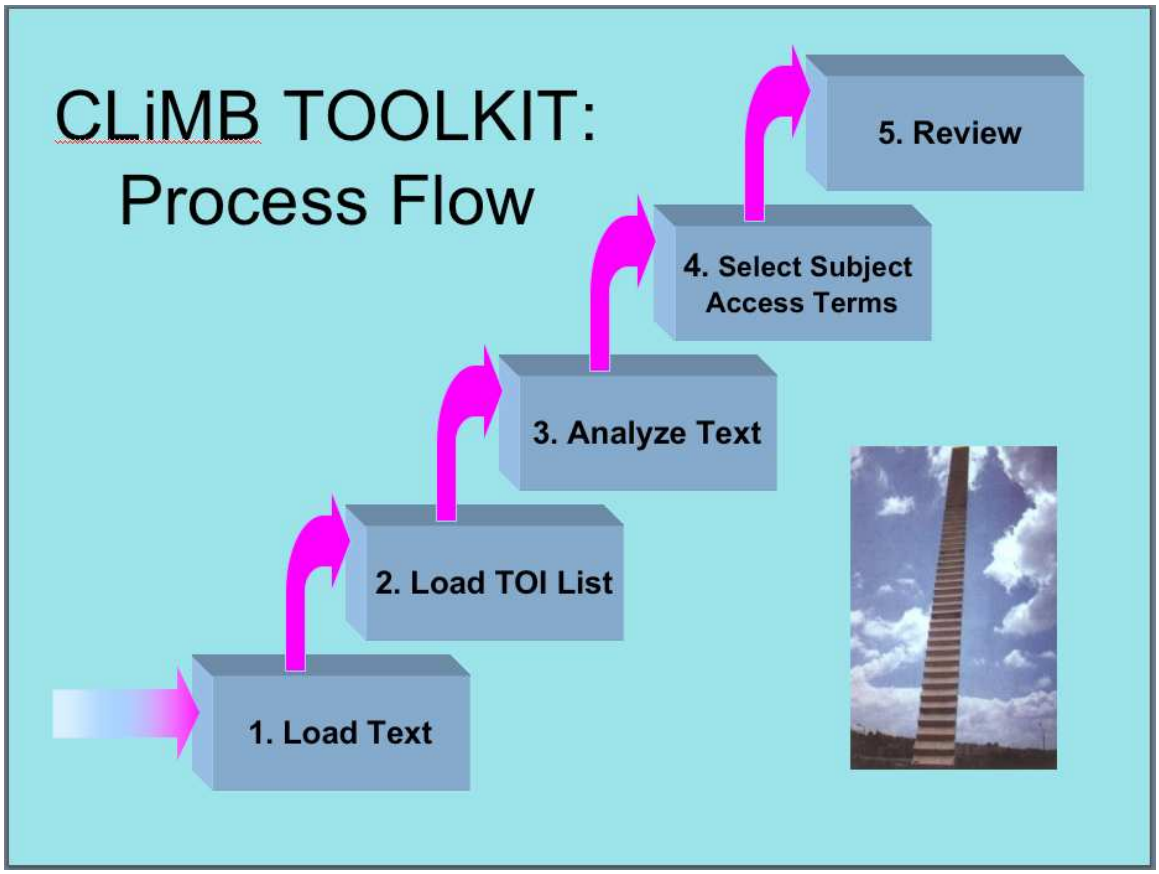


Figure 5.1: CLiMB Toolkit Process Flow

The design of the evaluation involved a mix of evaluation methods and question types in order to provide a comprehensive view of user responses to the prototype Toolkit. The three methods included:

1) Task-oriented steps to evaluate how well users were able to carry out the overall task of selecting metadata, as well as substeps contributing to this task;
2) User-satisfaction questions to evaluate the subjective response of users to the application environment and its functionality;
3) A roundtable discussion to help identify the concerns of users that they express spontaneously, and which are more or less shared by the entire group.

The general principle of using a mixed methodology for a formative evaluation is drawn from [Frechtling & Sharp, 1997; Patton, 1990]. The specific approach of combining a task-oriented and user-satisfaction study is drawn from [Walker et al., In submission]. Excerpts from some parts of the survey are included in the discussion below; a complete version can be consulted at http://www.columbia.edu/cu/libraries/inside/projects/climb/eval-questionnaire.pdf.

With respect to the three goals listed in paragraph one above, the evaluation was extremely successful. Users found the application interface very habitable overall as demonstrated by their ability to perform the steps presented in the questionnaire, combined with their responses to scaled questions [Likert, 1932] about the level of task difficulty (Likert scale questions are widely employed in usability studies and other types of user evaluations), and by their consistent level of satisfaction throughout the evaluation process. In regard to the specific functionality of finding metadata terms, users showed high satisfaction with the entire process in response to a question presented after they had selected terms for three images (see the discussion of response to question 20.33 in Section 5.3). The high average number of terms selected for all three images provides direct evidence that the Toolkit affords image professionals and art librarians the ability to select metadata interactively that would otherwise require scholarly assistance or special expertise. Furthermore, the amount of overlap in the terms selected provides indirect evidence of the quality of the metadata. (A direct test of the quality of extracted metadata awaits a study of the impact of CLiMB metadata in a retrieval environment). We describe the design of the evaluation in Section 5.2, present the quantitative and qualitative results in Section 5.3, and the results of the roundtable discussion in Section 5.4.

5.2 Overview of Evaluation

In effect, this report covers three versions of the CLiMB Toolkit: a preliminary version which was current for the CLiMB presentation to the Visual Resources Association in March, 2004 (0.8); the Toolkit evaluation version (0.9), which was employed for the April, 2004 survey, and is discussed here; and the final CLiMB Toolkit prototype (1.0), which addressed the results of our evaluation and represents the "final" Toolkit for this phase of the project. In Version 0.9, the following functions had been implemented:

- Loading and initialization of raw (ASCII) text was available (the functionality to make use of TEI markup had not yet been implemented);
- After initialization, text could be processed by a noun phrase chunker (termed "chunking") that locates the beginnings and ends of noun phrases (such as the underlined expressions in this sentence from a *North Carolina Museum of Art: Handbook of the Collections* passage about Jan Jansz den Uyl's painting entitled *Banquet Piece*: "At the same time, symbolically charged elements such as the empty glass, burned-down candle, and lute at the far left hint at deeper meanings.");

- A TOI list could be loaded, or TOIs could be manually created (refer to Section 1.3 or to the Glossary of CLiMB Terms in Appendix E for a definition of TOI);
- The TOI Finder could be run to locate references to TOIs in the loaded texts;
- Texts that had been processed by the TOI Finder could also be sectioned into associational contexts correlated with specific TOIs;
- Lists of Controlled Vocabulary could be loaded—included in this feature users were provided access to the Getty Art & Architecture Thesaurus (AAT), and the capability of selecting specific subsets from the AAT;
- The Noun Phrase detail frame was available.

In sum, participants in the evaluation could load and process texts and TOIs, could view text in a variety of ways in order to locate relevant sections (for instance, text a specific image or TOI), and isolate noun phrases likely to be good candidates for metadata terms.

The ten experts listed in Appendix F participated in the evaluation. They included five image professionals (Focht, Fry, Keefe, Wees, Williams), three art librarians (Clarke, Pisciotta, Rose), and two others. Of these last two, one has expertise in art terminology and metadata (Morgan), and one is a representative of the interactive design approach to digital image resources (Shelton). The participants were also approximately balanced by gender: four males (Clarke, Morgan, Pisciotta, Wees), six females (Fry, Keefe, Focht, Rose, Williams, Shelton). They were evenly balanced with respect to attendance at the CLiMB session at the March 2004 meeting of the Visual Resources Association (session 6; [Blitz et al., 2004]): five attended (Clarke, Morgan, Pisciotta, Rose, Shelton) and five did not. Thus we can classify results by background, gender, and degree of familiarity with the CLiMB project as it was presented at the VRA meeting.

The evaluation consisted of two components. In the first, participants were guided through a designated sequence of steps for using the CLiMB Toolkit, throughout which they answered a questionnaire asking them to evaluate the experience of using the Toolkit. In the second component, participants engaged in a roundtable discussion directed by Rebecca Passonneau.

The questionnaire (http://www.columbia.edu/cu/libraries/inside/projects/climb/eval-questionnaire.pdf) directed participants through a series of steps. There were forty-one questions eliciting user responses interspersed among twenty-two steps. Eighteen of the questions asked respondents to supply answers on a 5-point Likert scale [Likert, 1932] where 5 was the most negative and 1 the most positive. Frequently, a qualitative question followed the Likert question so that we could compare quantitative and qualitative answers. For example, users were asked to explain in more detail why they selected a particular value on the Likert scale, as in Question 9.10, which is shown in the next section in **Figure 5.2**.

The questionnaire consisted of three parts. In Part I, users were given an introductory overview of the Toolkit, exemplified by a preloaded project. In Parts II and III, users created their own projects using texts and TOI lists we provided; we also provided image gallery pages on the Web with images that corresponded to the texts. Parts II and III each dealt with a different image collection. In Part II, users were asked to create their own project using texts and TOIs for the *North Carolina Museum of Art: Handbook of the Collection* (*NCMA Handbook*). Part III pertained to the Greene & Greene

Collection at Avery Architectural and Fine Arts Library, Columbia University. All participants completed Parts I and II; only one completed the entire questionnaire.

5.3 Quantitative Results

The quantitative results provide a variety of measures of user satisfaction. That is, with the Likert questions, we asked users to tell us directly how they felt about aspects of the Toolkit. The focus of Section 5.33 is on the quality of the metadata that users selected, and thus provides detailed data on task success. The summary view of task success, however, is that eight participants successfully completed the metadata selection task for all three images in Part II, and the resulting metadata had sufficient overlap among evaluators as to indicate relatively high quality of metadata. All ten participants completed the task for the first image, which is the one discussed in detail in Section 5.33.

5.31 Overview

We give an overview of the quantitative results by examining three specific measures. 1) We present the average score on all Likert questions, which provides a single, summary metric of user satisfaction with the Toolkit. 2) We compare the questions that had the highest and lowest scores, which gives a view of the range of responses, and also specifics on what features were most and least satisfactory. 3) We compare the responses to a repeated question about overall user satisfaction that was presented at three points during the evaluation.

Analysis of the Likert questions in which responses fall on a 5-point scale, where 1 is the most positive, yields a quantitative view of the respondent's subject evaluations; eighteen of the forty-one questions on the survey were Likert questions. An example that will be discussed further is provided in **Figure 5.2**. We received answers from all evaluators on sixteen of the Liker questions; however, respondents treated one of these questions as eliciting multiple responses, so we quantify only the remaining fifteen questions. These questions are given in **Table 5.1**; the quantitative results for these Likert questions are presented in **Table 5.2**. As shown in **Table 5.2**, the average (AVG) for all responses on all Likert questions was 2.0 (the standard deviation, SD, a measure of variability around the average, is 1.8). This indicates that overall, people were satisfied with the experience of using the Toolkit.

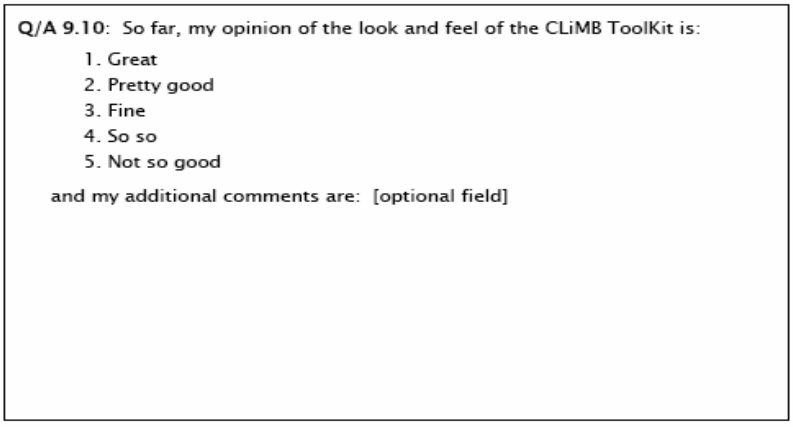


Figure 5.2: Sample Likert scale item followed by optional free text response. This instance of the question refers to Step 9, which asked users to enter a new TOI manually for one of the images in the image gallery.

#	**Text of question:**
3.2	Finding another access point to the "project" help text was: **1.** Completely obvious, . . . , **5.** Pretty difficult
6.6	How long did it take you to figure out how to close the project? **1.** No time, . . . , **5.** Too much time
7.7	I _____ what happens when I click on a table heading. **1.** Really like, . . . , **5.** Don't understand
8.8	So far, the concept of a CLiMB TOI is: **1.** Very clear, . . . , **5.** Pretty confusing
9.9	Entering a new TOI: **1.** Was easy, . . . , **5.** Was difficult
9.10	So far, my opinion of the look and feel of the CLiMB Toolkit is: **1.** Great, . . . , **5.** Not so good
11.12	Figuring out how to view the text: **1.** Was easy, . . . , **5.** Was difficult
11.13	Changing the text display options: **1.** Was easy, . . . , **5.** Was difficult
11.15	So far, my opinion of the look and feel of the CLiMB Toolkit is: **1.** Great, . . . , **5.** Not so good
15.18	Understanding the notion of a CLiMB "project" is: **1.** Very easy, . . . , **5.** Confusing
16.20	I was able to follow the above steps to get my new project to this point: **1.** Very easily, . . . , **5.** With difficulty
16.21	I find it _____ to understand why the TOI-Finder applies to the whole project, and why the sectioner applies to an individual text in a project. **1.** Easy, . . . , **5.** Difficult or impossible
17.22	I found these 4 substeps for finding AAT terms in the project texts: **1.** Very easy, . . . , **5.** Difficult or impossible
20.32	So far, my opinion of the look and feel of the CLiMB Toolkit is: **1.** Great, . . . , **5.** Not so good
20.33	I was _____ the process of selecting descriptive metadata. **1.** very pleased with, . . . , **5.** very displeased with

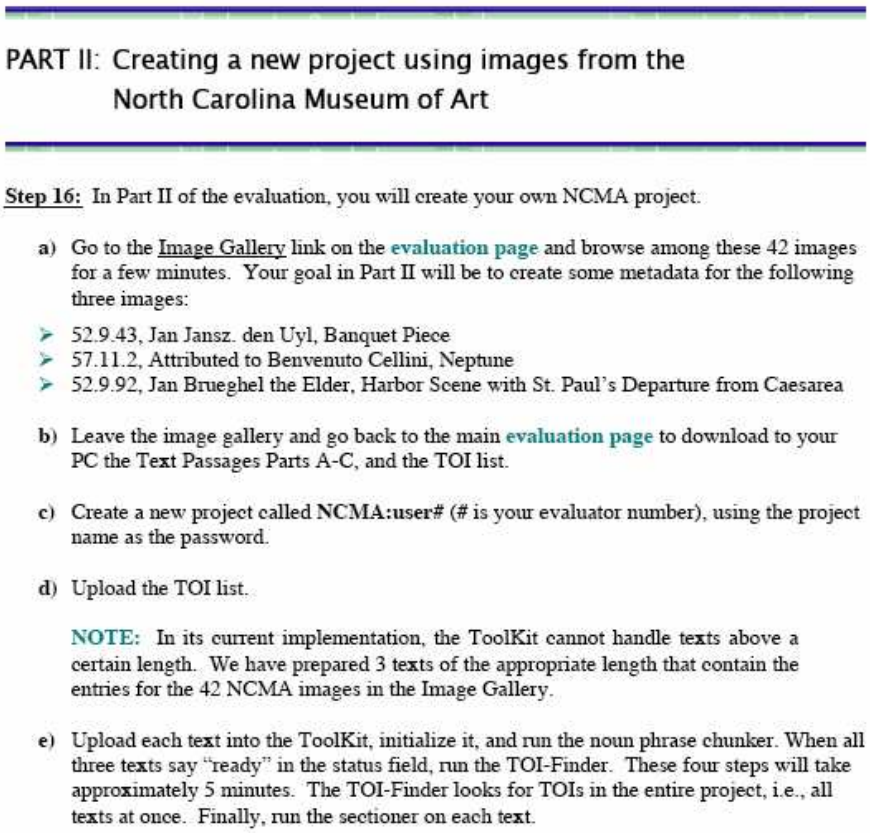
Table 5.1: Text of questions with Likert responses.

	Evaluator											
Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	AVG	SD
3.2	1	4	1	3	1	1	1	2	2	2	**1.8**	**1.09**
6.6	1	1	1	1	1	1	1	1	1	1	**1**	**0.00**
7.7	1	2	3	2	2	2	1	2	2	2	**1.9**	**0.60**
8.8	1	2	3	2	1	2	1	3	4	2	**2.1**	**1.05**
9.9	2	1	2	1	1	5	3	2	1	4	**2.2**	**1.32**
9.10	2	2	3	2	1	3		2	1	2	**2**	**0.76**
11.12	1	1	1	1	1	1	1	1	2	2	**1.2**	**0.33**
11.13	2	1	1	1	1	1	1	1	1	2	**1.2**	**0.33**
11.15	2	1	3	2	1	1	3	2	2	3	**2**	**0.78**
15.18	2	2	2	2	1	1	2	3	2	2	**1.9**	**0.60**
16.20	2	2	3	3	3	5	4	4	2	5	**3.3**	**1.07**
16.21	2	1	4	5	4	1	3	4	1	3	**2.8**	**1.56**
17.22	1	1	2	1	3	3	2	2	2	3	**2**	**0.78**
20.32		2	3	2		2	2	2	2	2	**2.1**	**0.38**
20.33		1	2	2		1	2		1	1	**1.4**	**0.55**

Table 5.2: Responses to fifteen Likert questions.

Of greater interest than the average of all Likert responses is to compare the questions receiving the lowest and the highest scores. The question receiving the least positive score was 16.20, where the average answer was 3.3 (standard deviation=1.1, median=3), or just below average satisfaction (note that the standard deviation for this question was lower than average, meaning there was more consistency among evaluators on this question). This question, which asked evaluators whether they were able to initiate a new project on their own (refer to **Table 5.1** for the text of the question), addressed what was undoubtedly the most time consuming and the most difficult step in the entire evaluation; thus, it was arguably the most difficult question and should naturally receive the lowest score. At Step 16 which is show below in **Figure 5.3**, respondents were asked, in essence, to create an entirely new Toolkit project: they were directed to an image gallery, to texts, and to a TOI list. They were told to load the TOI list (Step 16d), and to load, initialize, chunk, and section the text (Step 16e). Prior to Step 16, they had been shown examples of TOIs and texts in a sample project, but had never carried out the procedures in Step 16d. We would not have been surprised had the score been between 4 and 5; in sum, it is a very positive sign that respondents found this step close to manageable (3 on the Likert scale) after so little exposure to the Toolkit.

Having seen that the least positive average score was for the question following the most complex and time-consuming step, it should be no surprise that the questions receiving the most positive scores followed relatively easy steps, with one exception. There were four questions receiving average responses at the most positive end of the scale, or above Likert 1.5: one at 1.4 (Q/A 20.33); two at 1.2 (Q/A 11.12 and 11.13); and one at 1 (Q/A 6.6). Of these, the most informative for the evaluation was Q/A 20.33, for it addresses the core functionality of the Toolkit. This question came at the end of a series of steps in which participants reviewed texts to find terms describing three images from the *NCMA Handbook*. They were essentially asked to rate their satisfaction with the entire process of extracting metadata terms from texts. The fact that the average response ranks among the most positive, and ranks nearly the same as the question following what is arguably the easiest step (Step 6, to close the current project and reopen it), reflects significant overall satisfaction.



PART II: Creating a new project using images from the North Carolina Museum of Art

Step 16: In Part II of the evaluation, you will create your own NCMA project.

a) Go to the Image Gallery link on the evaluation page and browse among these 42 images for a few minutes. Your goal in Part II will be to create some metadata for the following three images:

- 52.9.43, Jan Jansz. den Uyl, Banquet Piece
- 57.11.2, Attributed to Benvenuto Cellini, Neptune
- 52.9.92, Jan Brueghel the Elder, Harbor Scene with St. Paul's Departure from Caesarea

b) Leave the image gallery and go back to the main evaluation page to download to your PC the Text Passages Parts A-C, and the TOI list.

c) Create a new project called **NCMA:user#** (# is your evaluator number), using the project name as the password.

d) Upload the TOI list.

NOTE: In its current implementation, the ToolKit cannot handle texts above a certain length. We have prepared 3 texts of the appropriate length that contain the entries for the 42 NCMA images in the Image Gallery.

e) Upload each text into the ToolKit, initialize it, and run the noun phrase chunker. When all three texts say "ready" in the status field, run the TOI-Finder. These four steps will take approximately 5 minutes. The TOI-Finder looks for TOIs in the entire project, i.e., all texts at once. Finally, run the sectioner on each text.

Figure 5.3: Step 16 of the CLiMB Toolkit Evaluation Survey.

As a final means of summarizing the quantitative responses, we note that questions 9.10 (shown above in **Figure 5.2**), 11.15, and 20.32, are identical queries about overall user satisfaction with the Toolkit (these items appear in boldface in **Tables 5.1** and **5.2**). We repeated this question in order to gauge whether continued exposure to the Toolkit had a negative impact on user satisfaction. Note that there is no change in user satisfaction between items 9.10 and 11.15; the average response at this point is 2 ("Pretty good"). At step 20.32, the average response is only modestly less positive (2.13 on a 5-point scale). In sum, user satisfaction starts out relatively high and remains stable throughout the process of creating a new project, and selecting metadata for three images.

5.32 Group differences on Likert questions

For the purpose of analyzing group differences in the Likert data, we grouped the ten respondents into two categories: image professionals (N=5) and others (N=5). There were group differences on questions 9.9, 9.10, and 16.21, but on all other questions, both groups had the same or approximately the same average scores and standard deviations.

On question 9.9, which asked the evaluators about the difficulty of entering a new TOI, the respondents found the task to be "Fairly Easy" on average (Score = 2.33). However, by group, the image professionals found it "manageable to a bit difficult" (Score = 3.25), in contrast the mixed group of art librarians and others, who found it "easy" to "not hard" (Score = 1.6). For question 9.10, regarding their opinion so far of the "look and feel of the CLiMB Toolkit", the average score overall was 2 ("Pretty Good"). By group, the image professionals had an average score of 2.4, or midway between "Pretty Good" and "Fine", whereas the mixed group of art librarians and others had an average score of 1.5, or midway between "Pretty Good" and "Great". Finally, on question 16.21 which asked respondents to gauge how easy or difficult it was to understand that the TOI-Finder applies to a *project* whereas the sectioner applies to a specific *text*, the average score overall was 2.8, or closer to "possible with some thought" (3) than to "fairly easy" (2). The group breakdown, however, was 2.2 for the image professionals, and 3.4 for the mixed group of art librarians and others.

In sum, image professionals had a more negative response than art librarians on questions 9.9 and 9.10. Because question 9.10 immediately followed 9.9, with no intervening steps to carry out, it is likely that the two questions are causally related for both groups; that is, those who found entering a TOI relatively more difficult (9.9) would likely have a less positive overall response to the Toolkit (9.10), and conversely, those who found entering a new TOI relatively easy would likely have a more positive response to the Toolkit. Assuming there is such a dependency, then the main difference between the two groups is on questions 9.9 and 16.21. In the former case, the image professionals were less positive about the relevant task; in the latter case, the image professionals were more positive. It is not clear why there would be these group differences on these two questions, and only these two, but because both questions pertain to the relation between TOIs and texts, it's possible that there may be some underlying consistency.

5.33 Metadata selection task

Three questions (18.25, 19.28, 20.31) required the ten evaluators to perform a metadata selection task. This task provided the "task success" portion of the dual *user satisfaction/task success* approach to system evaluation [Walker *et al.*, in submission; Walker *et al.*, 1997]. Participants were directed to an

Jan Jansz den Uyl
Banquet Piece
(NCMA 52.9.43)

Jan Brueghel the Elder
Harbor Scene with St. Paul's Departure from Caesarea (detail)
(NCMA 52.9.92)

Benvenuto Cellini
(attributed)
Neptune
(NCMA 57.11.2)

Figure 5.4: Three images from the NCMA Collection.

Step 18: Go to the ToolKit and view the text for **NCMA-A-plain.txt**. You will be asked to select some descriptive metadata terms, so set your desired display options in the Control Panel frame in the manner that you would find most convenient for looking at descriptive terms in the text. Go to the text frame and look for the section on "Banquet Piece." You might need to use the "Find" feature of the browser; if you do, make sure you are in the correct frame by clicking anywhere in the left-most frame to make sure the browser's search will run there. Do not select a noun phrase in the text – your browser might shift its focus to the "Detail Panel."

Q/A 18.23: The paragraphs I found for the section on "Banquet Piece" are:

Q/A 18.24: The display options I selected were:

Q/A 18.25: From the paragraphs I listed above (for 18.23), I have selected the following descriptive terms for the "Banquet Piece" record:

Figure 5.5: Step 18 and subsequent questions from the CLiMB Toolkit Evaluation Survey.

image gallery containing forty-two images from the North Carolina Museum of Art with in-structions to view three images in particular (**Figure 5.4**)—Jan Jansz den Uyl's *Banquet Piece*, Jan Brueghel the Elder's *Harbor Scene with St. Paul's Departure from Caesarea*, and a statue of *Neptune* attributed to Benvenuto Cellini. Participants were then asked to use the Toolkit to locate relevant passages in the *North Carolina Museum of Art Handbook*; to choose text display options; and finally to select descriptive terms for a catalog record. The passages were brief (225, 254, and 304 words, respectively).

Because all participants completed question 18.25, the report will focus on those responses in considering the metadata selection task. **Figure 5.5** below shows step 18 of the survey, along with the questions pertaining to it. Question 18.25 in particular asked participants to select terms to be used in a

catalog record describing den Uyl's *Banquet Piece*. The metadata terms were harvested from the following passage from the *NCMA Handbook* (p. 100).

> Jan Jansz. den Uyl Dutch, 1595/96-1640. Banquet Piece, about 1635. Oil on panel. 31 3/8 x 37 in. (79.7 x 94.0 cm) Purchased with funds from the State of North Carolina, 1952 (52.9)
>
> This remarkable Dutch still-life painting, somber yet sumptuous, represents a seventeenth-century type known as a "banquet piece." The various surfaces and textures of luxury tablewares made of pewter, gold, silver, and glass, as well as the foodstuffs and white linen tablecloth, are meticulously rendered. At the same time, symbolically charged elements such as the empty glass, burned-down candle and lute at the far left hint at deeper meanings. The painting has been interpreted both as a vanitas image, that is, one that refers to the rapid passage of life and the emptiness of worldly possessions and pleasures, and as a representation of the five senses. However, it is equally likely that Dutch concerns about overindulgence and ostentatious display also are addressed in the painting.
>
> Jan Jansz. den Uyl was well known and successful during his lifetime—the Flemish master Peter Paul Rubens owned three of his pictures—was largely forgotten after his death until the middle of this century. In most of his known works, Den Uyl made a visual pun on his name: an owl (in Dutch uyl) is perched upon the gold container in the center of the composition.

Ninety-six selections were made, yielding a total of thirty-five terms. **Table 5.3** summarizes the results. The selected terms were then ranked in descending order based on the number of times chosen, as illustrated in **Table 5.4**. For this ranking, variant forms were taken into account, reducing the number of terms selected from 35 to 27 by considering the following terms to be variants: *Vanitas / vanitas image*; *Dutch still life painting / still life / still life painting*; *linen tablecloth / tablecloth / white linen tablecloth*; *luxury tablewares / tablewares*; and *pun/visual pun*. Other terms, such as '*Dutch*' and '*glass*' could not be considered variants as they appeared multiple times within the text passage, both alone and in combination with other terms.

The most agreed upon terms are in rows 1 through 5 of **Table 5.4**, comprising all terms selected at least six and at most nine times. Of the remaining twenty two terms, four terms were selected five times, five terms were selected four times, two terms were selected three times, two terms were selected twice, and the remaining eight terms were chosen only once. Using the Getty Art & Architecture Thesaurus, three of the five most commonly agreed upon terms derive from the Objects Facet: *vanitas, still lifes*, and *candles*. In addition, *vanitas* and *still lifes* appear in the Visual and Verbal Communication hierarchy of the Objects Facet; *candles* appears in the Furnishings and Equipment hierarchy of the same facet. The other terms—*glass* and *pewter*—are both found in the Materials hierarchy of the Materials Facet.

	Term	Times chosen	Chosen by evaluator #	Variations
1	31 3/8 x 37 in.	1	8	
2	banquet piece	2	5, 10	
3	burned down candle	5	2, 4, 5 ,8, 10	*candle*
4	candle	2	3,4	*burned down candle*
5	Dutch	5	1, 2, 3, 4, 9	
6	Dutch still life painting	3	5, 6, 8	*still life, still life painting*
7	empty glass	5	2, 5, 7, 8, 10	
8	five senses	4	5, 6, 8, 10	
9	foodstuffs	2	2,8	
10	glass	6	2, 3, 4, 5, 7, 10	
11	gold	1	2	
12	gold container	1	1	
13	linen tablecloth	3	4 7, 10,	*white linen tablecloth*
14	lute	5	3, 4, 5, 7, 8	
15	luxury	1	3	*luxury tablewares*
16	luxury tablewares	2	8, 10	*luxury, tablewares*
17	musical instrument	1	3	
18	oil on panel	1	8	
19	ostentatious display	3	5, 8, 10	
20	overindulgence	2	5, 8	
21	owl	4	1, 4, 5, 10	
22	painting	4	1, 2, 3, 9	
23	pewter	6	2, 3, 4, 5, 7, 10	
24	pun	1	10	*visual pun*
25	seventeenth century	4	2, 3, 5, 6,	
26	silver	1	2	
27	still life	3	2, 3, 9	*Dutch still life painting, still life painting*
28	still life painting	2	4, 10	*Dutch still life painting, still life*
29	tablecloth	1	3	
30	tablewares	2	2, 5	*luxury tablewares*
31	vanitas	8	2, 3, 4, 5, 6, 7, 9, 10	
32	vanitas image	1	8	
33	visual pun	2	1, 10	*pun*
34	white linen tablecloth	1	8	*linen tablecloth*
35	worldly pleasures	1	10	
	Total	96		

Table 5.3: Metadata selections for question 18.25.

	Term	Times chosen
1	*vanitas, vanitas image*	9
2	*Dutch still life painting, still life, still life painting*	8
3	*burned down candle, candle*	7
4	*glass*	6
5	*pewter*	6
6	*Dutch*	5
7	*empty glass*	5
8	*linen tablecloth, tablecloth, white linen tablecloth*	5
9	*lute*	5
10	*five senses*	4
11	*luxury tablewares, tablewares*	4
12	*owl*	4
13	*painting*	4
14	*seventeenth century*	4
15	*ostentatious display*	3
16	*pun, visual pun*	3
17	*banquet piece*	2
18	*foodstuffs*	2
19	*overindulgence*	2
20	*31 3/8 x 37 in.*	1
21	*gold*	1
22	*gold container*	1
23	*luxury*	1
24	*musical instrument*	1
25	*oil on panel*	1
26	*silver*	1
27	*worldly pleasures*	1
	Total	96

Table 5.4: Ranked metadata selections for question 18.25.

CLiMB is designed to be schema independent. Metadata selected using the CLiMB Toolkit can potentially be inserted into records using MARC (653 or 650 subject fields), VRA Core 3.0 (subject category), Dublin Core (subject element), or other schemas. The expectation is that image catalogers will select as many or as few terms as is applicable to their local practice. With the den Uyl painting, for example, the cataloger could limit his or her selection of terms to the first five rows of **Table 5.4**. **Figure 5.6** is an illustration of a VRA Core 3.0 work record for den Uyl's *Banquet Piece*, enhanced with the free text terms harvested from the *NCMA Handbook*. For the purposes of this example these terms are labeled "CLiMB Subject".

Record Type = work
Type = painting
Title = Banquet Piece
Measurements. Dimensions = 79.7 x 94 cm
Material = oil on panel
Creator = Uyl, Jan Jansz. den
Date. Creation = ca. 1635
Location. Current Repository = Raleigh (NC, USA), North Carolina Museum of Art
ID Number. Current Repository = 52.9.43
Style/Period = Dutch
Style/Period = Seventeenth century
CLiMB Subject = Dutch still life painting
CLiMB Subject = Vanitas
CLiMB Subject = burned down candle
CLiMB Subject = glass
CLiMB Subject = pewter

Figure 5.6: VRA Core 3.0 record enhanced with CLiMB generated subject terms.

Evaluation participants performed the metadata selection task with two more text passages, one about Jan Brueghel the Elder's *Harbor Scene with St. Paul's Departure from Caesarea* (19.28), and the other about a bronze 16th century statue of *Neptune* attributed to Benvenuto Cellini (20.31). Eight people completed question 19.28 (one participant ended the questionnaire at 18.25 and another elected not to provide any descriptive terms from the passage). Seventy-one selections were made yielding a total of twenty-nine terms. The most popular selection was *landscape*, chosen by seven participants; *fishermen* and *sailors* were each chosen by six. Named entities or personal names were deemed important access points as indicated by the fact that *St. Paul*, *Herod the Great*, and *Josephus*, were each selected by six, five, and five evaluators respectively.

Eight participants also completed question 20.31. Fifty-six selections were made yielding a total of thirty-three terms. For this passage, there was less agreement overall among the participants as to which terms would provide useful metadata. The majority of the terms, sixteen, were only selected by one person, and fourteen terms were chosen twice. The most commonly selected terms were: *hippocamps* and its synonym *sea horse monsters*, selected a total of eight times; and *Neptune*, with four selections. Human variation in language tasks can occur for many reasons, including differential motivation and knowledge; however, research on human summarization of texts provides quantitative differences between texts on the accessibility of the semantic content [Nenkova and Passonneau, 2004].

The metadata selection task's success revealed a high rate of satisfaction, as the average response to 20.33 (user satisfaction with the entire process of extracting metadata terms from texts) ranked among the most positive responses. This question was very informative for the evaluation, and complements the task data illustrated in **Tables 5.3** and **5.4**, for it addresses the core functionality of the Toolkit (see discussion of question 20.33 in Section 5.31). Interestingly, the number of terms

selected by a given participant had no impact on user satisfaction, as indicated by **Table 5.5**. The four participants who expressed the most satisfaction with process of selecting descriptive metadata ("Very pleased") chose 49, 23, 14, and 13 terms, respectively; those at the next level ("Satisfied") selected 31, 28, and 12 terms, respectively. Three respondents did not complete question 20.33. That the high user satisfaction rate was independent of the number of terms selected possibly indicates that the users felt the metadata terms were of relatively high quality.

Evaluator	Response to 20.33	# of terms selected
10	1. Very pleased	49
2	1. Very pleased	23
6	1. Very pleased	14
9	1. Very pleased	13
4	2. Satisfied	31
3	2. Satisfied	28
7	2. Satisfied	12
5	0. N/A	26
1	0. N/A	16
8	0. N/A	13

Table 5.5: User satisfaction correlated with number of terms selected.

5.34 Comments by Evaluation Participants

Overall the participants were pleased with the functionality of the Toolkit regarding metadata selection (see the discussion of question 20.33 in Section 5.1 and 5.33 above). Choosing texts to associate with an image collection, however, was a major concern that arose in the additional comments section of the questionnaire. Likewise, four of the nine discussion point categories from the Roundtable Discussion pertain to issues about texts, such as how to select texts that have relevant metadata, and how to select texts that are suitable for the Toolkit; this is discussed in more detail in Section 5.42 below.

The text passages selected for the CLiMB Toolkit Evaluation survey were criticized by some of the participants, particularly the section on Brueghel's *Harbor Scene*, which focused as much on the artist as on the work of art. This raised the question of text selection generally. Many image professionals work in small or one person operations, often separate from the library. They questioned how they would be able to access or even identify relevant titles under these circumstances.

CLiMB evaluators were also quite concerned about the availability of texts in electronic format, which is a necessary prerequisite for using the Toolkit. Again, many face limitations of staffing and funding, and felt they lacked the resources to convert printed books into electronic format.

5.4 Roundtable Discussion

The format followed for the Roundtable Discussion at the CLiMB Toolkit Evaluation was designed to elicit spontaneous comments from the participants on their experiences with the Toolkit, rather than on *a priori* questions posed by the CLiMB team. That is, we did not want to direct the discussion; instead, we sought to use it to discover issues that we might not be aware of from among discussion participants whose expertise we hoped to tap into. CLiMB team members (Becky Passonneau, Robbie Blitz, Dave Elson, Vera Horvath, and Mark Weber) contributed to the discussion in the following ways: by introducing an open-ended discussion point pertaining to the user-centered design of new technology; by answering questions from the evaluation participants; by offering general comments aimed at keeping the discussion lively and focused.

5.41 Discussion Format

The evaluation leader, Becky Passonneau, initiated the discussion by posting on the whiteboard the following open-ended statement[1]:

> *Not to design is to design.*

Design plays a role in the Toolkit with respect to whether the components of the Toolkit that are visible to users can be interpreted easily, and whether the Toolkit helps users accomplish the task of identifying subject access terms. Here we mention here two examples to illustrate the general idea that any part of the human-constructed environment can potentially hinder rather than enhance our ability to function.

The first example of a poor design, or failure to design, illustrates how the appearance of a tool can be a hindrance. In Butler Library, we find examples of faucet handles that look as if they should be turned, when it fact they have to be pressed from above in order to release water. When a faucet handle has spokes, the spokes constitute a perceived *affordance* (cf. [Gibson, 1977; Norman, 1988]) for the hand. Most people assume that handles like these should be gripped and rotated. In cases where the spokes are merely ornamental, many users will first discover the lack of functionality of the spokes by trial-and-error.

A second example that came up in conversation with evaluator Henry Pisciotta involved the need to hire a design firm at Penn State to analyze traffic flow within buildings on campus. Over the years, navigating around campus had become more confusing, and a design firm was engaged to map out convenient paths and provide visually accessible signs and graphics. In this case, the doors, hallways, staircases, and so on, were not necessarily confusing in and of themselves. Rather, the complexity of the relationships among potential paths and destinations required the insertion of a new intervening level of design. In this evaluation, we are interested in user reactions to both levels of design in the Toolkit.

Several situations for which "in depth" interviews are appropriate, according to an NSF handbook on evaluations [Frechtling & Sharp, 1997], are:

a) complex subject matter;
b) detailed information sought;
c) busy, high-status respondents;
d) highly sensitive subject matter.

The CLiMB Toolkit evaluation falls into situations a) and b) above; in addition, we believe the following two categories constitute additional sources of complexity that therefore require user-centered design methods:

---

[1]Adapted from a frequently quoted phrase, *Not to decide is to decide.* This phrase itself was apparently adapted from the writings of the theologian Harvey G. Cox:

> Somewhere deep down we know that in the final analysis we do decide things and that even our decisions to let someone else decide are really our decisions, however pusillanimous. (Harvey G. Cox, *On Not Leaving It to the Snake,* 1967. Quoted as "Not to decide is to decide" in *Peter's Quotations* by Laurence J. Peter.)

e) lack of a standard practice among image professionals in providing subject access to images
f) rapidly evolving conditions, including the emergence of many new digital image collections, and an emerging user community

We applied listening and questioning strategies from Patton's "in depth interview" method [Patton, 1990] either to deepen the discussion as interesting points emerged, or to refocus the discussion on a new topic when the discussion seemed to become too diffuse. The NSF Handbook [Frechtling & Sharp, 1997] likewise reiterates the importance of this:

> Interviewers seek to encourage free and open responses, and there may be a tradeoff between comprehensive coverage of topics and in-depth exploration of a more limited set of questions. In-depth interviews also encourage capturing of respondents' perceptions in their own words, a very desirable strategy in qualitative data collection. This allows the evaluator to present the meaningfulness of the experience from the respondent's perspective.

5.42 Issues

Many of the key issues that emerged from the Roundtable Discussion related to a fundamental fact, namely that the CLiMB project involves a collaboration among two distinct communities: on the one hand, professionals from the visual resources and library arenas, who deal with digital image collections, and on the other hand, computational linguists, who deal with the computer science techniques that might be applied to problems in cataloging. The concerns, expertise and assumptions of the two communities are often distinct. We believe the Toolkit evaluation, as well as our October 2003 formative evaluation of manually selected metadata terms, will help lead to a "common ground." "Common ground" has a commonsense meaning, but it is also a term with a technical meaning in linguistic studies of discourse [Clark & Marshall, 1981; Clark, 1996]. Because the goal of CLiMB is to extract metadata from text automatically, texts and their types constitute important objects of reference within this common ground. The importance of defining a common ground of what texts the CLiMB Toolkit should be designed to process, and how the properties of texts differ, was an important issue that emerged from the discussion.

6. Future Work

The foundation from CLiMB-1 forms the basis for a continuing project on using computational linguistic techniques for improving image access. The focus of CLiMB has been on scholarly texts and highly refined vocabulary. Techniques for filtering the possible range of terms associated with images through thesaural links, such as the AAT also forms part of the current project. Achievements in CLiMB-1 have included criteria for collection selection. If there is one important summary of what we learned from the CLiMB-1, it is that the nature of the text is more important than the image, since it is the text that serves as the basis for information extraction for access.

In order to fully test the hypotheses in CLiMB-1, our results show that CLiMB-2 will need to include:

- testbeds of collections which conform to the criteria defined by CLiMB-1;
- partners to build the basis for testing the CLiMB toolkit—e.g., museums and other libraries with qualifying collections;
- the ability to refine and modify the CLiMB Toolkit to be responsive to ongoing input from users;
- the use of specialists in interface design to enable the Toolkit to be tested with a full working interface.

Therefore, we envision a follow-up project that builds on the achievements of CLiMB-1. Since the PI, Judith Klavans, is now a Research Professor at the College of Information Sciences (CLIS) at the University of Maryland, she will be able to bring in faculty to the project who have expertise in access, computational linguistics, and interface design. At the same time, the Libraries at the University of Maryland, headed by Dr. Charles Lowry, will form an additional partner. Furthermore, with the richness of museums and libraries in the Washington D.C. area, combined with former connections with New York City cultural institutions, the CLiMB user group can be expanded to encompass a wider range. This will enable new input on the CLiMB tools and on CLiMB use. Klavans will maintain partners from CLiMB-1 in order to ensure continuity.

Finally, the output of CLiMB, as demonstrated by the data shown in Section 4 of this report, has proven that computational linguistic techniques have very promising potential. The next step is to further test this potential and to develop refinements so the results can be put to use in the difficult and challenging task of addressing the image cataloging bottleneck.

7. Selected References

Blitz, Robbie, Angela Giral, Judith Klavans, Rebecca Passonneau (2004). "Project CliMB: Using Computational Linguistic Techniques to Harvest Image Descriptors." Presented to the Visual Resources Association, Portland, Oregon, 2004.

Choi, Freddy Y. Y. (2000). "Advances in domain independent linear text segmentation." *Proceedings of the North American Chapter of the Association of Computational Linguistics* (NAACL). Seattle, Washington, 2000.

Clark, H.H. (1996). *Using Language*. Cambridge University Press, Cambridge, 1996.

Clark, H. H. and C. R. Marshall (1981). "Definite reference and mutual knowledge." In Joshi, Webber, and Sag, eds., *Elements of Discourse Understanding*. Cambridge University Press, Cambridge.

Cohen, William W. (1996). "Learning with set-valued features." *Proceedings of the Thirteenth National Conference on Artificial Intelligence*. Portland, Oregon, 1996.

Davis, Peter T., David K. Elson, Judith L. Klavans. (2003) "Methods for Precise Named Entity Matching in Digital Collections." *Third ACM/IEEE Joint Conference on Digital Libraries* (JCDL), 2003.

Day, David, John Aberdeen, Lynette Hirschman, Robyn Kozierok, Patricia Robinson and Marc Vilain. (1997). "Mixed-Initiative Development of Language Processing Systems." *Fifth Conference on Applied Natural Language Processing*. Association for Computational Linguistics, Washington D.C., 1997.

Evans, David K., Judith L. Klavans, and Nina Wacholder (2000). "Document Processing with LinkIT." *Proceedings of the RIAO Conference*. Centre de Hautes Etudes Internationales d'Informatique Documentaire (CID) and the Center for the Advanced Study of Information Systems (CASIS), 2000.

Fellbaum, Christiane (1998). *WordNet, An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Finch, Steve and Andrei Mikheev (1997). "A Workbench for Finding Structure in Texts." *Proceedings of the Fifth Conference of Applied Natural Language Processing* (ANLP). Washington D.C., 1997.

Frechtling, J. and L. Sharp (Eds.) (1997). *User-Friendly Handbook for Mixed Method Evaluations*. National Science Foundation, Virginia.

Gibson, J. J. (1977). *The Ecological Approach to Visual Perception*. Boston, Mass.: Houghton Mifflin.

Hearst, Marti A. (1994). "Multi-Paragraph Segmentation of Expository Text." *Proceedings of the 32nd Annual Meeting of the Association of Computational Linguistics* (ACL). Las Cruces, New Mexico, 1994.

Kan, Min-Yen, Judith L. Klavans, and Kathleen R. McKeown (1998). "Linear segmentation and segment relevance." *Sixth Annual Workshop on Very Large Corpora* (WVLC6). Montreal, Quebec, Canada, 1998.

Klavans, Judith L., Kathleen McKeown, Min-Yen Kan, and Susan Lee (1998). "Resources for the evaluation of summarization techniques." *Proceedings of the 1st International Conference on Language Resources and Evaluation*. Grenada, Spain, 1998.

Likert, R. (1932). *A Technique for the Measurement of Attitudes*. New York: Archives of Psychology.

Francis, W. Nelson and Henry Kucera (1982). *Frequency Analysis of English Usage: Lexicon and Grammar*. Houghton Mifflin, Boston, MA.

Miller, George (1995). "WordNet: a lexical database for English." *Communications of the ACM* (Association for Computing Machinery), vol. 11, 1995.

Nenkova, Ani and Rebecca Passonneau. (2004). "Evaluation content selection in summarization: the Pyramid Method." *Joint annual meeting of Human Language Technology and the North American Association of Computational Linguistics (HLT/NAACL)*, Boston, MA.

Norman, D.A. (1988). *The Psychology of Everyday Things*. New York: Basic Books.

Patton, M.Q. (1990). *Qualitative Evaluation and Research Methods*. Sage, Newbury Park, 2nd Edition, 1990.

Salton, Gerard, editor (1971). *The SMART Retrieval System: Experiments in Automatic Document Processing*. Prentice Hall: Englewood Cliffs, NJ.

Utiyama, Masao, and Hitoshi Isahara (2001). "A statistical model for domain-independent text segmentation." *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics* (ACL/EACL). Toulouse, France, 2001.

Wacholder, Nina (1998). "Simplex NPs sorted by head: a method for identifying significant topics within a document." Workshop on the Computational Treatment of Nominals. *Proceedings of the Joint 17th International Conference on Computational Linguistics—36th Annual Meeting of the Association for Computational Linguistics* (COLING-ACL'98). Montreal, Quebec, Canada, 1998.

Walker, M. A., D. Litman, C. Kamm, and A. Abella (1997). "PARADISE: A Framework for evaluating Spoken Dialogue Agents." In *Proceedings of ACL '97* (Madrid, Spain, July 1997), MIT Press.

Walker, Marilyn, R. Passonneau, J. Aberdeen, J. Boland, E. Bratt, J. Garofolo, L. Hirschman, A. Le, S. Lee, S. Narayanan, K. Papineni, B. Pellom, J. Polifroni, A. Potamianos, P. Prabhu, A. Rudnicky, G. Sanders, S. Seneff, D. Stallard, S. Whittaker. *Cross-Site Evaluation in DARPA Communicator: The June 2000 Data Collection*. In Submission, 2003.

Appendices: CLiMB Supporting Material

A. Additional Information on CLiMB

- **CLiMB home page:** http://www.columbia.edu/cu/libraries/inside/projects/climb/
  (The CLiMB home page includes links to "People," "Tools & Prototypes," "Presentations & Publications," and "Collections," as well as to the original grant proposal and press release.)
- **CRIA home page:** http://www.columbia.edu/cu/cria/
- **First year report:** http://www.columbia.edu/cu/libraries/inside/projects/climb/reports/CLiMB.1year.report.pdf
- **CLiMB Collections information:** http://www.columbia.edu/cu/libraries/inside/projects/climb/collections.html
- **CLiMB Evaluation Questionnaire:** http://www.columbia.edu/cu/libraries/inside/projects/climb/eval-questionnaire.pdf

B. Collections and Related Material

Greene & Greene—Architectural drawings from the Avery Art and Architecture Library's Greene & Greene Collection

1. Bosley, Edward R. *Greene & Greene*. London: Phaidon, 2000.
2. Current, William R. *Greene & Greene: Architects in the Residential Style*. Fort Worth, TX: Amon Carter Museum of Western Art, 1974.
3. Makinson, Randell. *Greene & Greene: Architecture as a Fine Art*. Salt Lake City, UT: Pergrine Smith, 1977.
4. Makinson, Randell. *Greene & Greene: The Passion and the Legacy*. Salt Lake City, UT: Gibbs Smith, 1998.
5. Smith, Bruce. *Greene & Greene Masterworks*. San Francisco: Chronicle Books, 1998.
6. Strand, Janann. *A Greene & Greene Guide*. Pasadena, CA: G. Dahlstrom, 1974.

Finding Aid to the Greene & Greene Architectural Papers and Records Collection at Avery Library: http://www.columbia.edu/cu/lweb/eresources/archives/avery/greene/.

Greene & Greene Virtual Archives: http://www.usc.edu/dept/architecture/greeneandgreene/. This is a collaborative visual resource for Greene & Greene materials that includes the collection at Avery Library.

Chinese Paper Gods—Woodblock prints from the C.V. Starr East Asian Library's Anne S. Goodrich Collection

1. Day, Clarence Burton. *Chinese Peasant Cults: Being a Study of Chinese Paper Gods*. Taipei: Ch'eng Wen Publishing Co., 1974.
2. Goodrich, Anne Swann. *Peking Paper Gods: A Look at Home Worship*. Nettetal: Steyler Verlag, 1991.

3. Laing, Ellen Johnston. *Art and Aesthetics in Chinese Popular Prints: Selections from the Muban Foundation Collection.* Ann Arbor, MI: Center for Chinese Studies, University of Michigan, 2002.

South Asian Temples—Digital images from the Digital South Asia Library, selections from the American Institute of Indian Studies (AIIS) Photo Archives.

Digital South Asia Library: http://dsal.uchicago.edu.

North Carolina Museum of Art—Digital images of works in the NCMA collection.

1. Nagy, Rebecca Martin, Ed. *North Carolina Museum of Art: Handbook of the Collections.* Raleigh, NC: The Museum of Art, Hudson Hills Press, 1998.

Authority Sources / Structured Vocabularies

The Getty *Art & Architecture Thesaurus On Line*: http://www.getty.edu/research/tools/vocabulary/aat/.

CLiMB has also employed the following two books as references on this topic:

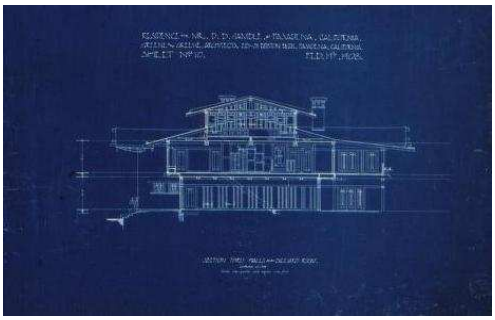
Baca, Murtha (editor). *Introduction Metadata: Pathways to Digital Information.* Los Angeles: Getty Information Institute, 2002.

Baca, Murtha (editor). *Introduction to Art Image Access: Issues, Tools, Standards, Strategies.* Los Angeles: J. Paul Getty Museum Publications, 2002.

C. Sample Catalog Records

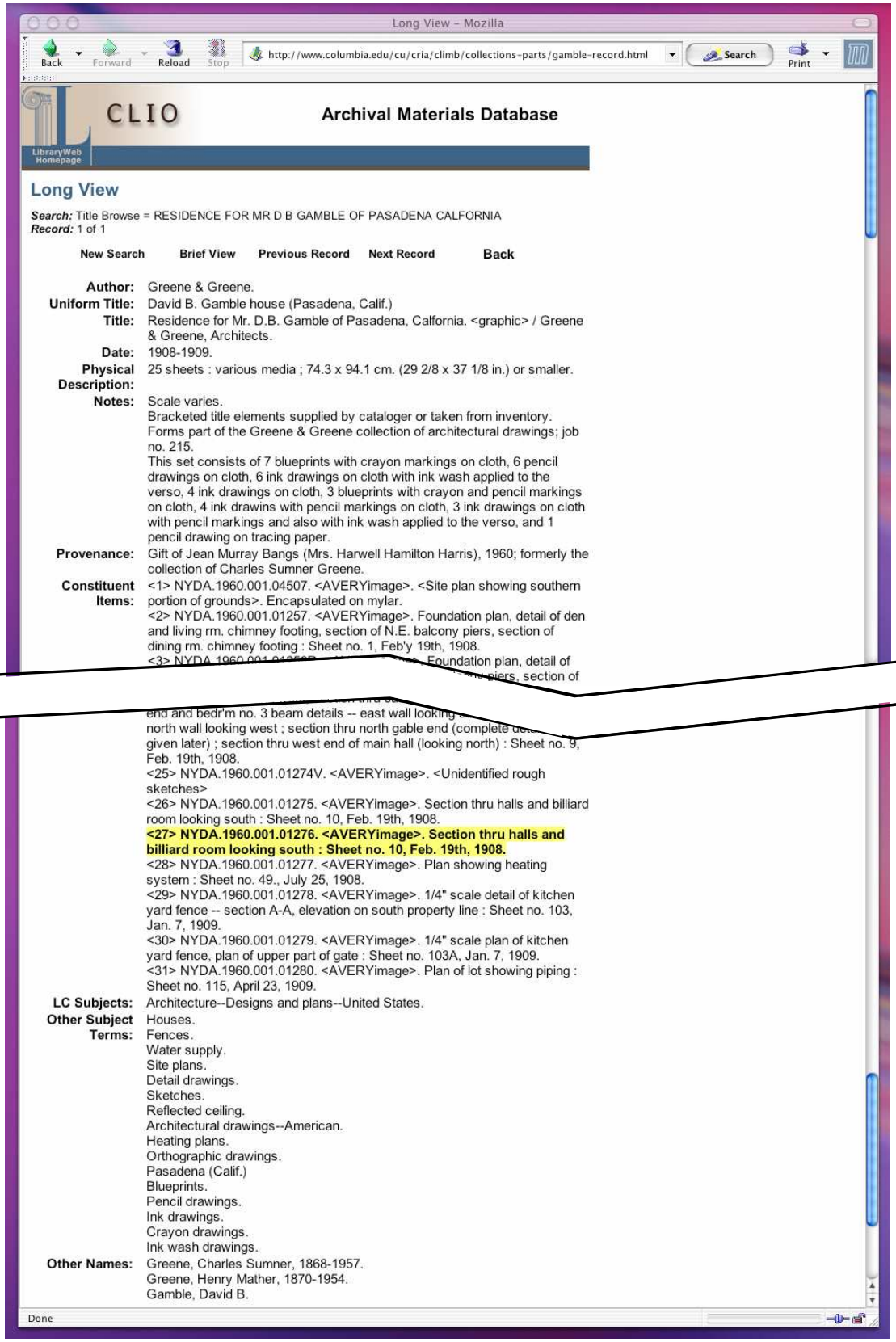
Greene & Greene Architectural Images

Greene & Greene Architectural Records and Papers Collection
Avery Architectural and Fine Arts Library, Columbia University



Residence for Mr. D. B. Gamble of Pasadena, California
Section through halls and billiard room looking south
Avery Architectural and Fine Arts Library,
Columbia University Libraries

Columbia University Catalog record for the Gamble house drawing:



Long View – Mozilla

http://www.columbia.edu/cu/cria/climb/collections-parts/gamble-record.html

CLIO

Archival Materials Database

LibraryWeb Homepage

Long View

Search: Title Browse = RESIDENCE FOR MR D B GAMBLE OF PASADENA CALFORNIA
Record: 1 of 1

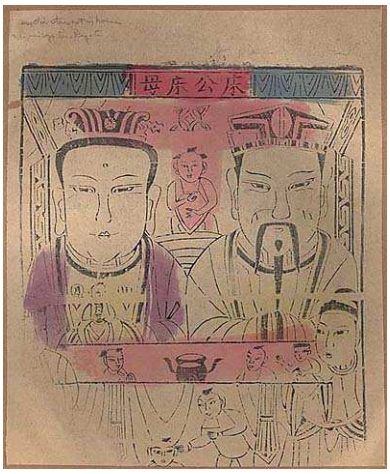
New Search **Brief View** **Previous Record** **Next Record** **Back**

Author: Greene & Greene.
Uniform Title: David B. Gamble house (Pasadena, Calif.)
Title: Residence for Mr. D.B. Gamble of Pasadena, Calfornia. <graphic> / Greene & Greene, Architects.
Date: 1908-1909.
Physical Description: 25 sheets : various media ; 74.3 x 94.1 cm. (29 2/8 x 37 1/8 in.) or smaller.
Notes: Scale varies.
Bracketed title elements supplied by cataloger or taken from inventory.
Forms part of the Greene & Greene collection of architectural drawings; job no. 215.
This set consists of 7 blueprints with crayon markings on cloth, 6 pencil drawings on cloth, 6 ink drawings on cloth with ink wash applied to the verso, 4 ink drawings on cloth, 3 blueprints with crayon and pencil markings on cloth, 4 ink drawins with pencil markings on cloth, 3 ink drawings on cloth with pencil markings and also with ink wash applied to the verso, and 1 pencil drawing on tracing paper.
Provenance: Gift of Jean Murray Bangs (Mrs. Harwell Hamilton Harris), 1960; formerly the collection of Charles Sumner Greene.
Constituent Items: <1> NYDA.1960.001.04507. <AVERYimage>. <Site plan showing southern portion of grounds>. Encapsulated on mylar.
<2> NYDA.1960.001.01257. <AVERYimage>. Foundation plan, detail of den and living rm. chimney footing, section of N.E. balcony piers, section of dining rm. chimney footing : Sheet no. 1, Feb'y 19th, 1908.
<3> NYDA.1960.001.01258. <AVERYimage>. Foundation plan, detail of [illegible] piers, section of [illegible] end and bedr'm no. 3 beam details -- east wall looking [illegible] north wall looking west ; section thru north gable end (complete det[illegible] given later) ; section thru west end of main hall (looking north) : Sheet no. 9, Feb. 19th, 1908.
<25> NYDA.1960.001.01274V. <AVERYimage>. <Unidentified rough sketches>
<26> NYDA.1960.001.01275. <AVERYimage>. Section thru halls and billiard room looking south : Sheet no. 10, Feb. 19th, 1908.
<27> NYDA.1960.001.01276. <AVERYimage>. Section thru halls and billiard room looking south : Sheet no. 10, Feb. 19th, 1908.
<28> NYDA.1960.001.01277. <AVERYimage>. Plan showing heating system : Sheet no. 49., July 25, 1908.
<29> NYDA.1960.001.01278. <AVERYimage>. 1/4" scale detail of kitchen yard fence -- section A-A, elevation on south property line : Sheet no. 103, Jan. 7, 1909.
<30> NYDA.1960.001.01279. <AVERYimage>. 1/4" scale plan of kitchen yard fence, plan of upper part of gate : Sheet no. 103A, Jan. 7, 1909.
<31> NYDA.1960.001.01280. <AVERYimage>. Plan of lot showing piping : Sheet no. 115, April 23, 1909.
LC Subjects: Architecture--Designs and plans--United States.
Other Subject Terms: Houses.
Fences.
Water supply.
Site plans.
Detail drawings.
Sketches.
Reflected ceiling.
Architectural drawings--American.
Heating plans.
Orthographic drawings.
Pasadena (Calif.)
Blueprints.
Pencil drawings.
Ink drawings.
Crayon drawings.
Ink wash drawings.
Other Names: Greene, Charles Sumner, 1868-1957.
Greene, Henry Mather, 1870-1954.
Gamble, David B.

Done

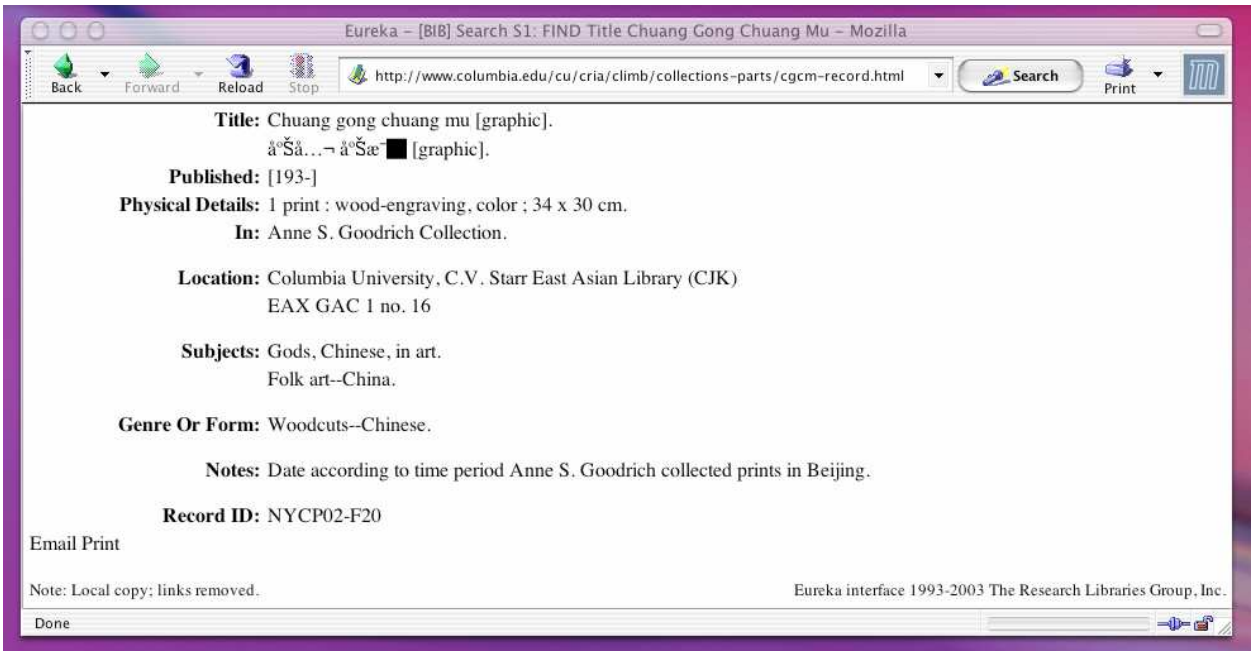
Chinese Paper Gods Collection

The Anne S. Goodrich Chinese Paper Gods Collection
C.V. Starr East Asian Library, Columbia University



Chuang gong chuang mu
Wood-engraving, color
C.V. Starr East Asian Library,
Columbia University Libraries

Catalog Record for *Chuang gong chuang mu*:



Eureka – [BIB] Search S1: FIND Title Chuang Gong Chuang Mu – Mozilla

Back Forward Reload Stop http://www.columbia.edu/cu/cria/climb/collections-parts/cgcm-record.html Search Print

Title: Chuang gong chuang mu [graphic].
å°Šå…¬å°Šæ¯ [graphic].
Published: [193-]
Physical Details: 1 print : wood-engraving, color ; 34 x 30 cm.
In: Anne S. Goodrich Collection.

Location: Columbia University, C.V. Starr East Asian Library (CJK)
EAX GAC 1 no. 16

Subjects: Gods, Chinese, in art.
Folk art--China.

Genre Or Form: Woodcuts--Chinese.

Notes: Date according to time period Anne S. Goodrich collected prints in Beijing.

Record ID: NYCP02-F20

Email Print

Note: Local copy; links removed.

Eureka interface 1993-2003 The Research Libraries Group, Inc.

Done

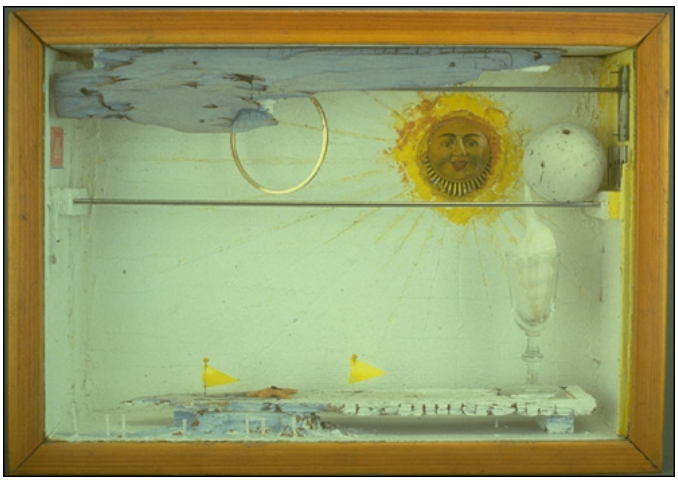
North Carolina Museum of Art

Digital Images from the Collection
Raleigh, NC

Georgia O'Keeffe (American, 1887-1986)
Cebolla Church, 1945
Oil on canvas
20 1/16 x 36 1/4 in.
(51.1 x 92.0 cm.) [72.18.1]

Sample catalog record for *Cebolla Church* in MARC format:

100 O'Keeffe, Georgia, ≠d 1887 -1986.
245 Cebolla church ≠ h [slide] / ≠ c Georgia O'Keeffe.
260 ≠c2003
300 1 slide : ≠ b col.
500 Object date: 1945.
500 Oil on canvas.
500 20 x 36 in.
535 North Carolina Museum of Art ≠ b Raleigh, N.C.
650 Painting, American ≠ y 20th century.
650 Women artist ≠ z United States
650 Church buildings in art.



Joseph Cornell (American, 1903-1972)
Suzy's Sun (For Judy Tyler), 1957
Mixed media construction
10 3/4 x 15 x 4 in.
(27.3 x 38.1 x 10.2 cm.) [78.1.1]

Sample catalog record for *Suzy's Sun (For Judy Tyler)* in VRA Core 3.0 format:

Record Type=work
Type=shadow box
Title=Suzy's Sun (For Judy Tyler)
Measurements.Dimensions=10 3/4 x 15 x 4 in. (27.3 x 38.1 x 10.2 cm.)
Material.Medium=mixed media
Creator.Personal name=Cornell, Joseph
Creator.Role=artist
Date.Creation=1957
Location.Current Repository=Raleigh (NC, USA), North Carolina Museum of Art
ID Number.Current Accession=78.1.1
Subject=assemblages (sculpture)

D. CLiMB Presentations and Talks: May 2002–June 2004

- CLiMB Presentation to Rutgers, School of Communication, Information and Library Studies, Art Librarianship Course, June 24, 2004. Slides available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/rutgers.ppt.

- Second Annual External Advisory Board Meeting, June 18th and 19th, 2004.

- CLiMB Presentation for the Art Libraries Society of North America, April 19, 2004. Slides available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/ARLIS_CLiMB2.ppt.

- CLiMB Presentation for the Visual Resources Association, March 11, 2004. Slides, live demonstration, and PowerPoint screenshots of live demonstration, and introductory text available at http://www.columbia.edu/cu/cria/climb/presentations.html.

- CLiMB Presentation for Formative Evaluation Guests, October 2003. Slides available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- CLiMB Presentation for the Columbia University Libraries Representative Committee of Librarians (RCL) Brown Bag program, Tuesday, October 14, 2003. Slides, live demonstration, and PowerPoint screenshots of live demonstration available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- CLiMB Presentation to the Cornell University Library Metadata Working Group, Friday, October 3, 2003. Slides, live demonstration, and PowerPoint screenshots of live demonstration available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- CLiMB as a CRIA Project: Presentation for Columbia University Libraries, September 16, 2003. Slides available at httphttp://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- Inaugural External Advisory Board Meeting, June 12th and 13th, 2003.

- Third ACM/IEEE Joint Conference on Digital Libraries (JCDL), May 28th, 2003. Presentation of paper on "Methods for Precise Named Entity Matching in Digital Collections." The paper was authored by Peter Davis, David Elson, and Judith Klavans. Available on line at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- Presentation on CLiMB for the Professional Staff at Avery Fine Art and Architectural Library, Columbia University. Given by Roberta Blitz and Angela Giral, May 15th, 2003.

- Coalition for Networked Information, April 28th, 2003. Presentation on the CLiMB project at CNI. Live demo and presentation slides available at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

- CLiMB talks on "Computational Linguistics and Digital Libraries" delivered by Judith Klavans:
  - Hong Kong University Library, Hong Kong, February 14th, 2003.
  - Tsinghua University, Department of Computer Science, Beijing, China, February 28th, 2003.

- National Institute of Informatics, Information Science, Tokyo, March 5th, 2003.
- Toyo Bunko Oriental Library, Tokyo, March 5th, 2003.
- Waseda University Library, Waseda, Japan, March 5th, 2003.
- Kyoto University, Department of Computer Science, Kyoto, Japan, March 6th, 2003.

Abstract: As digital libraries have grown, so has the need for developing more effective ways to access collections. This talk will present an overview of the CLiMB project (Computational Linguistics for Metadata Building), funded by the Mellon Foundation and currently underway at Columbia University. The goal of the project is to use computational linguistic techniques to extract metadata relevant to image collections, and thus to improve cataloging access. This research addresses the access bottleneck by applying the latest natural language processing techniques to the problem of identifying descriptive metadata. Our goal is to load our results into a database for image search, although we have not yet reached this phase of the project. This talk will report on research in CLiMB's first phase. In addition, the talk will provide an overview of selected digital library projects at Columbia, in terms of collections, access and technology.

- Mellon Technical Meeting, November 21st, 2003.
  Presentation given to Mellon Foundation participants on the technical and methodological aspects of the CLiMB project. Slides for this "Report on Technical Progress to the Mellon Foundation" can be found at http://www.columbia.edu/cu/libraries/inside/projects/climb/climb_orig/presentations.html.

E. CLiMB Glossary of Terms

Because CLiMB is an inherently interdisciplinary project, many of the terms we employ are either specialized for use in certain fields, or they have been developed by the CLiMB project internally in order to address particular conceptual needs. With the hope of making the CLiMB project accessible to people in fields as diverse as those represented in our project teams, this short glossary provides definitions and explanations for some of the terms that this report and the project generally tend to rely upon.

associated text. A term used by the CLiMB project for texts that are potentially related to the items in a given image collection. For instance, in the first year of testing, much of the work focused on Edward Bosley's book *Greene & Greene*, which is a text associated with the Greene & Greene Collection in Columbia's Avery Library.

associational context. The extent of the text surrounding an occurrence of a Target Object Identifier or its variants (or any other term located by the CLiMB toolset) that is deemed likely to yield relevant associated words and concepts. Associational context may, for instance, include a certain number of words before and after the occurrence of a TOI. The exact limit would be rule-based but flexible for different cases.

authority list. A set of relevant terms, such as names of places and subject vocabularies, that is provided a priori to CLiMB software by outside experts. The authority list for Greene & Greene architectural projects, for instance, comes from the master list of project names decided upon by the catalogers at Columbia's Avery Library.

Brown Corpus. A 1,000,000-word corpus of edited English prose compiled by W. Nelson Francis and Henry Kucera at Brown University in 1964. Designed to represent common English usage, it draws from sources such as newspapers, humanities texts, and fiction.

catalog record, item level. A catalog record that describes a specific item in a collection. In the catalog record for the Greene & Greene Collection drawing shown above in Appendix C, the highlighted text near the bottom of the record represents item level cataloging: "<27> NYDA. 1960.001.01276. <AVERYimage>. Section thru halls and billiard room looking south : Sheet no. 10, Feb. 19th, 1908." These item level records are subordinated to records at the project level.

catalog record, project level. A catalog record that describes part of a collection based upon the "project" it refers to. CLiMB uses this term mostly with reference to the Greene & Greene collection. The catalog record in Appendix C provides project level information near the top of the screen shot; for instance, the "Uniform Title" is "David B. Gamble House (Pasadena, Calif.)." Thus, the record encompasses those constituent items that relate to this project.

cataloging. Describing works of art, architectural projects, visual representations of works of art or architecture, or bibliographic materials. Cataloging most often involves the creation of a systematic record that provides a basic classification and description for the object being cataloged. Frequently the terms used to catalog an object come from standardized vocabularies.

co-occurrence. A relationship among words or phrases in a text based upon near proximity. Two words or noun phrases are co-occurrent when they appear close to each other in a text on at least one occasion.

co-reference. A relationship among words or phrases in a text determined by whether they describe the same object. Two words or noun phrases are co-referents when they both refer to the same thing: for example, "our nation's first President" and "George Washington," or "the Thorsen house" and "the house designed for William Thorsen."

DTD. Document Type Definition. A DTD is the grammar that describes the structure of an XML file, and tells the computer how that particular file should be read. The CLiMB DTD, for example, indicates that for a given text each chapter is made up of sections, that each section is made up of paragraphs, that each paragraph is made up of sentences, and so forth.

external vocabulary (also *domain dictionary* or *domain vocabulary*). A set of nouns and noun phrases that describe objects relevant to the subject at hand as a whole. For the Greene & Greene collection, where the domain is architecture, external vocabulary terms may include "porte cochere" and "inlaid brick," but not terms specific to Greene projects such as "the Blacker entry." At present the CLiMB suite of tools will locate external vocabulary terms in text from a Getty Structured Vocabulary, the *Art and Architecture Thesaurus*.

gold standard. A benchmark against which to test a software tool. For instance, in running the CLiMB toolset over a text, we might use experts to develop a list of terms and phrases to represent a standard against which to measure the results of automatic metadata extraction; another option would be to employ existing catalog records as standards by which to measure the effectiveness of the toolset in replicating or enhancing existing record information.

MARC. MAchine-Readable Catalog format. MARC represents a standard for describing bibliographic items (or objects in a collection) in catalog records. It is used by the Library of Congress, and aids in the exchange of data among information systems.

metadata. Data describing other data. For example, the size of a Web page, and the word that conveys the gist of its content, are metadata. In the CLiMB project, metadata usually refers to words and phrases in text that describe the items in a given image collection.

metadata schema. A set of rules for structuring metadata information such that it can incorporate specific elements of that information in encoded form.

MMF. Master Metadata File. According to the Columbia University Libraries' Web pages, "Columbia's Master Metadata File (MMF) is a locally developed metadata repository built around a MARC-based relational database schema. As of Sept. 2002, it holds over 75,000 metadata records for digital items & collections held locally or accessed remotely. The schema was designed to be able to represent multiple versions, collections, aggregations such as pages in a book, and hierarchies of digital objects. Information may be imported and exported in several formats. The database also may be used as an intermediate architectural component and may be queried interactively."

NP chunker. Noun Phrase chunker. A software tool that automatically locates noun phrases in a text.

POS tagger. Part of Speech tagger. A software tool that automatically locates (or "tags") the parts of speech in a text.

pinyin (also see *Wade-Giles*)**.** A transliteration system for the romanization of Chinese written characters that has been in use since the 1950s. This method of transferring written Chinese

into a Latin alphabet has already been the standard for the United States Government for more than two decades; it is also the standard used by the United Nations and most of the world's media. Pinyin is largely coming to replace Wade-Giles, an older system that is now thought to provide a much less accurate representation of Chinese phonemes. A common instance of this replacement is the substitution of "Beijing" (pinyin) for "Peking" (Wade-Giles).

precision (also see *recall*)**.** In computational linguistics, a characteristic of the results obtained by running a software tool over a text in order to identify certain terms or types of terms. Precision is a measure of the accuracy with which the tool identifies the correct terms. For example, if a tool is designed to identify noun phrases, and does so with a precision of 90%, this means that 90% of the terms identified by the tool are in fact noun phrases. Precision does not account for whether the tool has located all of the desired terms (this is expressed by 'recall').

precision versus recall. Often, results in computational linguistics are expressed by a combined score for precision and recall. Generally, as one rises, the other tends to fall. There are several methods for obtaining satisfactory overall scores. At present, the CLiMB suite of tools begins by searching for high-recall but low-precision matches in a given text, and then seeks to increase the precision of those matches with further computation.

recall (also see *precision*)**.** In computational linguistics, a characteristic of the results obtained by running a software tool over a text in order to identify certain terms or types of terms. Recall is a measure of the tool's ability to locate all of the desired terms. For example, if a tool is designed to identify noun phrases, and does so with a recall of 90%, this means that the tool has located 90% of all possible noun phrases. Recall does not account for whether the tool has identified the terms accurately (this is expressed by 'precision').

segmenter. A software tool that automatically breaks a text up into smaller parts, or segments, based upon content or topic. "Segmentation" thus refers to a technique for dividing texts into discrete topical segments pertaining to specific themes.

SQL. Structured Query Language". A standard language used for the expression of database queries. SQL allows users to search for and retrieve information from databases.

TEI-XML / CLiMB TEI-XML. Text Encoding Initiative Extensible Markup Language. TEI-XML is XML that adheres to the standard for encoding texts in electronic form that is issued by the TEI Consortium (http://www.tei-c.org/). CLiMB currently uses XML that has overlap with the TEI standard. Our eventual goal is to achieve full compliance with TEI guidelines; in the report we refer to this as "CLiMB TEI-XML."

term, domain specific (or *subject specific*). A term or phrase that is specific to a given subject area or discipline. For instance, the "domain" for Greene & Greene projects is architecture, and would include words that are specific to the field, such as "porte cochere" or "pergola."

term, project specific. A term or phrase that is specific to a particular project or object; often a TOI, but also words associated with a project that are not domain specific. For instance, the city "Pasadena" is a project specific term for the "Freeman A. Ford house," because it is the project's location.

TF/IDF. Term Frequency / Inverse Document Frequency. An equation for measuring the relative frequency of words appearing in different sets of data. CLiMB has applied this equation

to Chapter 5 of Edward R. Bosley's *Greene & Greene* in order to compare frequency of words in that chapter to the Brown Corpus, which is a model representing standard English usage.

TOI / Target Object Identifier. A noun phrase in a text that directly refers to a discrete target object. For the target object "the David B. Gamble house" (Greene & Greene collection), TOIs may include "the Gamble house," "the house," and "his residence in Pasadena," but not "the Gamble entry" or "the site." However, TOIs vary from collection to collection. Each of CLiMB's three collections uses a different kind of term for TOIs: for Greene & Greene, TOIs are terms referring to architectural projects; for the Chinese Paper Gods, TOIs are the names of the gods depicted; for the South Asian Temples Images, TOIs are place names for temple sites.

Wade-Giles (also see *pinyin*)**.** An older transliteration system for the romanization of Chinese written characters. This method of transferring written Chinese into a Latin alphabet is mostly being replaced by pinyin, a more modern system that is thought to provide a superior representation of Chinese phonemes. A common instance of this replacement is the substitution of "Beijing" (pinyin) for "Peking" (Wade-Giles).

XML (also see *TEI-XML*)**.** Extensible Markup Language. A version of SGML (Standard Generalized Markup Language), which is a standard for defining the structure of different types of electronic documents. XML allows organizations to customize their own markup languages for structuring data.

F. CLiMB People

CLiMB Project Team Members

Project Leader / Principal Investigator: Judith Klavans

Coordinating Committee

- Judith Klavans, Director, Center for Research on Information Access (klavans@cs.columbia.edu)
- Patricia Renfro, Deputy University Librarian (pr339@columbia.edu)
- Stephen Davis, Director, Libraries Digital Program (daviss@columbia.edu)
- Bob Wolven, Director of Bibliographic Control & Library Systems (wolven@columbia.edu)
- Angela Giral, Director, Avery Architectural and Fine Arts Library (giral@columbia.edu)

Curatorial Group

- Judith Klavans
- Angela Giral
- Stephen Davis
- Bob Wolven
- Roberta Blitz, Digital Collections / Art Research Librarian (rlb179@columbia.edu)
- Bob Scott, Head, Electronic Text Service (scottr@columbia.edu)
- Amy Heinrich, Director, Starr East Asian Library (heinrich@columbia.edu)
- David Magier, Director, Area Studies (magier@columbia.edu)

Technical Group

- Judith Klavans
- Stephen Davis
- Roberta Blitz
- Rebecca Passonneau, Research Scientist, Columbia Computer Science (becky@cs.columbia.edu)
- Peter Davis, Graduate Research Assistant (ptd7@cs.columbia.edu)
- Vera Horvath, Graduate Research Assistant (vhorvath@cs.columbia.edu)
- David Elson, Programmer (delson@cs.columbia.edu)

Evaluation Group

- Stephen Davis
- Roberta Blitz
- Rebecca Passonneau
- Angela Giral
- David Elson
- Vera Horvath

CLiMB Project Coordinator: Mark Weber (mcweber@cs.columbia.edu)

CLiMB External Advisory Board Members

Alfred Aho	**Columbia University** Chair, Department of Computer Science
Caroline Arms	**Library of Congress** Coordinator, National Digital Library Program
Murtha Baca	**The Getty Research Institute** Head, Standards and Digital Resource Management Program
Michael Buckland	**University of California, Berkeley** Professor, School of Information Management & Systems
Jeff Cohen	**Bryn Mawr College** Director of Digital Media / Visual Resources
Greg Crane	**Tufts University** Editor-in-Chief, Perseus Project
Marilyn Deegan	**Oxford University** Digital Resources Manager, Refugee Studies Centre
David Fenske	**Drexel University** Dean, College of Information Science and Technology
Carl Lagoze	**Cornell University** Senior Research Associate, Information Science
Clifford Lynch	**Coalition for Networked Information** Executive Director
Merrilee Proffitt	**The Research Libraries Group** Program Officer
John Unsworth	**University of Illinois at Urbana-Champaign** Dean and Professor, Graduate School of Library and Information Science
Nina Wacholder	**Rutgers University** Assistant Professor, School of Communication, Information & Library Studies
Clara Yu	**Middlebury College** Director, National Institute for Technology

CLiMB Toolkit Evaluation Participants

Sherman Clarke	**Bobst Library, New York University** Head of Original Cataloging
Marcia Focht	**Art History Department, Binghamton University** Curator of Visual Resources
Eileen Fry	**Indiana University, Fine Arts Slide Library** Slide Librarian
Jeanne M. Keefe	**Architecture Library, Rensselaer Polytechnic** Visual Resources Librarian **Visual Resources Association** Vice President
Emerson Morgan	**ARTstor** Metadata Analyst
Henry Pisciotta	**Pennsylvania State University** Arts & Architecture Librarian
Trish Rose	**Union Catalog of Art Images (UCIA) Project, UC San Diego** Image Metadata Librarian
Lena Shelton	**California Digital Libraries** Bibliographic Analyst
Dustin Wees	**ARTstor** Data Editor
Susan Williams	**Knight Visual Resources Collection, Cornell University** Acting Assistant Director / Systems Administrator

Vice President for Information Services &

University Librarian

Columbia University

535 West 114th Street
New York, NY 10027

http://www.columbia.edu/cu/libraries/inside/projects/climb/