8-26-2008

Computational linguistics for metadata building: Aggregating text processing technologies for enhanced image access

Judith Klavans
University of Maryland, College Park

Carolyn Sheffield
University of Maryland, College Park

Eileen Abels
Drexel University

Joan E. Beaudoin
Drexel University, joan.beaudoin@wayne.edu

Laura Jenemann

See next page for additional authors

Computational Linguistics for Metadata Building: Aggregating Text Processing Technologies for Enhanced Image Access

Judith Klavans[1,2], Carolyn Sheffield[2], Eileen Abels[3], Joan Beaudoin[3], Laura Jenemann, Jimmy Lin[1,2], Tom Lippincott[4], Rebecca Passonneau[4], Tandeep Sidhu[1], Dagobert Soergel[2], Tae Yano[5]

[1]Laboratory for Computational Linguistics and Information Processing (CLIP)
at the Institute for Advanced Computer Studies (UMIACS), University of Maryland, College Park, MD
[2]The iSchool at the University of Maryland, College Park, MD
[3]College of Information Science and Technology, Drexel University, Philadelphia PA
[4]Center for Computational Learning Systems, Columbia University, New York NY
[5]Department of Computer Science, Carnegie Mellon University, Pittsburgh PA
E-mail: jklavans@umd.edu, eileen.abels@ischool.drexel.edu, jeb56@drexel.edu, lj27@drexel.edu, jimmylin@umd.edu, tom@cs.columbia.edu, becky@ccls.columbia.edu, csheffie@umd.edu, tsidhu@umd.edu, dsoergel@umd.edu, taey@cs.cmu.edu

Abstract

We present a system which applies text mining using computational linguistic techniques to automatically extract, categorize, disambiguate and filter metadata for image access. Candidate subject terms are identified through standard approaches; novel semantic categorization using machine learning and disambiguation using both WordNet and a domain specific thesaurus are applied. The resulting metadata can be manually edited by image catalogers or filtered by semi-automatic rules. We describe the implementation of this workbench created for, and evaluated by, image catalogers. We discuss the system's current functionality, developed under the Computational Linguistics for Metadata Building (CLiMB) research project. The CLiMB Toolkit has been tested with several collections, including: Art Images for College Teaching (AICT), ARTStor, the National Gallery of Art (NGA), the Senate Museum, and from collaborative projects such as the Landscape Architecture Image Resource (LAIR) and the field guides of the Vernacular Architecture Group (VAG).

1. Project Goals

Creating access to ever-growing collections of digital images in scholarly environments has become increasingly difficult. Studies indicate that current cataloging practices are insufficient for accommodating this volume of visual materials, particularly for diverse user needs. The goal of the CLiMB project is to leverage text already written about images for automatically identifying, categorizing, filtering and selecting high quality descriptive metadata for image access.

Typically, in libraries and museums, cataloging is performed manually with minimal tombstone cataloging, i.e. the basic set of information (e.g. name of work, creator, date). However, what is usually lacking are rich descriptive terms (e.g. for Picasso's Guernica, "screaming horse", "the frozen women", "fauns" and "minotaurs")[1]. In addition, many legacy records lack subject entries altogether. The literature on end users' image searching practices, though sparse, indicates that this level of subject description may be insufficient for some user groups, including both general users and domain experts with knowledge of specialized vocabularies. Furthermore, the lack of subject-oriented description precludes searching and image analysis across topic area (e.g. searching for works with "minotaurs" as a theme).

Our hypothesis is that automatic and semi-automatic techniques may help fill the existing metadata gap by facilitating the assignment of subject terms. In particular, we are interested in the impact of computational linguistic technologies in extracting relevant access points from pre-selected texts. The CLiMB Toolkit applies Natural Language Processing (NLP), categorization, and disambiguation techniques over texts about images to identify, filter, and normalize high-quality subject metadata

2. The CLiMB Toolkit

Figure 2 shows a screen shot of the CLiMB Toolkit user interface for an image and text from the National Gallery of Art online collection[2]. Note that the center top panel contains the image, so catalogers can examine items as they work. The center panel contains the input text, with proper and common nouns highlighted. Terms under consideration are displayed below the full text with thesaural information accessible in the right-hand panel. Under this is the term the user has selected for consideration. The right-hand panel gives thesaural information. For normalizing terms, we use the Getty Vocabularies[3]: the Art and Architecture Thesaurus (AAT), the Thesaurus for Geographic Names (TGN), and the Union List of Artist Names (ULAN). In this example, two senses for the word "landscape" are displayed on the right. Note that the top portion of the panel displays possible matches in the AAT, followed by the middle portion which shows the chosen definition for the selected term, and finally, the bottom panel in which the entire hierarchy is displayed for the user to view and used to identify any related terms.

To extract terms from these relevant segments, we use off-the-shelf software to perform traditional NLP

[1] Taken from the exhibition notes from the Picasso exhibit at the National Gallery of Victoria, published by www.thornton.com

[2] www.nga.gov

[3] http://www.getty.edu/research/conducting_research/vocabularies/

techniques. In the current Toolkit, the Stanford tagger (Toutanova and Manning, 2000; Toutanova et al., 2003) is used since it is Java compliant and currently outperforms other taggers. We have used the open source Lucene toolbox to index. Internally developed noun phrase and proper noun identification rules have been applied. As part of categorization, we have applied a machine learning technique trained over text in the art and architecture domain to select a functional semantic category (Passonneau, 2008). Finally, we explore several disambiguation techniques, which we continue to refine and test with our user groups (Sidhu, 2007). Finally, candidate terms are proposed to catalogers for selection and export into an image database.

Currently, CLiMB focuses on nouns and noun phrases. Recent literature on image indexing indicates, however, that other parts of speech may be valuable in retrieving images. In a study of image professionals, (graphic designers, advertising staff, etc.), Jorgensen (2005) found that "while nouns account for the largest percentage of term type in image searches (just over 50%), adjectives account for 18% of the total term usage, verbs 10%, proper nouns 5%, concept 8%, byline 2%, visual content 2%, and date 1%. Of course, these results are highly dependent on the users and their image needs, but it does give some indication of the relative importances of the term types being searched."

3. Related Research

Broad domain users (as opposed to specialists) require access using broader non-specialist terms. Choi and Rasmussen (2003) studied the image-searching behaviors of faculty and graduate students in the domain of American history and found that generalists submitted more subject-oriented queries than known author and title searches. Currently, much cataloging is geared towards the specialist. On the other end of the spectrum is pure indexing of textual material in the physical domain of an image, such as that done by google (Palmer n.d.). Although such approaches are valuable for initial image access, the resulting high recall can make for a frustrating browsing experience for the end user.

On the other hand, the subjective nature of images inherently complicates the generation of accurate and thorough descriptions. Berinstein (1999) points out that even the guidelines provided by the Shatford-Panofsky matrix on what to describe are fluid and may be difficult to apply. Shatford (1994), building on Panofsky (1962), proposed a method for identifying image attributes, which includes analysis of the generic and specific events, objects, and names that a picture is "of" and the more abstract symbols and moods that a picture is "about". Panofsky describes the pre-iconographic, iconographic, and iconologic levels of meaning found in Renaissance art images. Shatford's generic and specific levels correspond to Panofsky's pre-iconographic and iconographic levels, respectively, and encompass the more objective and straightforward subject matter depicted in an image. The iconologic level (Shatford's about) addresses the more symbolic, interpretive, subjective meanings of an image. To aid user access, catalogers are encouraged to consider both general and specific terms for describing the objective content of an image as well as to include the more subjective iconologic, symbolic, or interpretive meanings. Iconologic terms may be the most difficult for catalogers to assign but occur often in texts describing images.

4. Current Cataloging Approaches

In the CLiMB workflow studies, we examined existing cataloging practices and gathered cataloger perspectives on current challenges in image indexing. Understanding the component processes in current practice has enabled the development of the CLiMB workbench to be easily integrated into existing standards, systems, and practices. Furthermore, by determining which challenges are general to the field and which arise in conjunction with specific collections, we were able to identify additional needs which our research may address. In architecture collections, for example, text may describe a building or architectural site as a whole while the corresponding image typically provides only a detailed view of the work. Part-whole relationships such as these present specific linguistic challenges for associating segments of text with one or more images. This research is not the topic of this paper, and will be described in a forthcoming article.

5. CLiMB Architecture: Systems and Methods

The CLiMB architecture is shown in Figure 1. The data flow for CLiMB starts at the upper left which shows the input to the system:

1. an image,
2. minimal metadata (e.g. image, name, creator)
3. text.

This input is pre-processed, using external technologies, to identify coherent segments of text and associate those segments with relevant images. Input texts are marked up using TEI lite (Text Encoding Initiative) to identify topical divisions (chapters, sections, etc.). These divisions, or segments, are then mapped to corresponding images through the identification of plate and figure numbers. For art historical survey texts, such as Jansen (2004) and Gardner (2001), the automation of text-image association produces reliable results. CLiMB has investigated the application of linguistic technologies to semi-automatically classify, or categorize, text segments according to their semantic relationship to the image(s) which they describe Passonneau, et al (2007).

Through our partnership with the Getty Research Institute, we have been given access to three resources:

- The Art & Architecture Thesaurus (AAT), a structured vocabulary for describing art objects, architecture, and other cultural or archival materials. The AAT's structure is comprised of seven major facets (Associated Concepts, Physical Attributes, Styles and Periods, Agents, Activities, Materials, and Objects) from which multiple hierarchies descend. In total, AAT has 31,000 such records. Within the AAT, there are 1,400 homonyms, i.e., terms that can lead to several AAT records that may have multiple

meanings only one of which may apply in a given context.

- The Union List of Artist Names (ULAN), a name authority that includes the given names of artists, as well as any known pseudonyms, variant spellings, and name changes (e.g., married names). The structure of this resource is similar to the Agents facet of the AAT in that it contains Person and Corporate Body as its primary facets.
- The Thesaurus of Geographic Names (TGN), an authority for place names, including place names as they appear in English as well as in other languages, historical names, and names in natural order and inverted order.

These vocabularies are well-established and widely-used multi-faceted thesauri for the cataloging and indexing of art, architecture, artifactual, and archival materials. Each of these resources specifies which variation of a given concept or name is the preferred term, enabling consistent cataloging across collections. We have utilized these resources to link terms derived from testbed texts to standardized, controlled terms, thus helping users expand their information space. The Getty resources are used to select the particular homograph of a term.

5.1 Disambiguation

We have tested three approaches to disambiguation in our domain, using the AAT as our baseline thesaurus (Sidhu, 2007). However, it is clear that we need to utilize additional terminological resources since many common terms—and senses of ambiguous terms--are missing from the specialist thesaurus. The challenge of using domain-specific vocabularies combined with general vocabularies, and the impact on disambiguation, is a little-studied topic. We have observed that terms with many senses in the AAT may have just one sense in a general dictionary, and that some terms with many senses in a general resource are simply missing altogether in the AAT. The impact of these observations on disambiguation has yet to be established.

In order to test our disambiguation technique, we first annotated a text to use for evaluation. Following standard procedure in word sense disambiguation tasks (Palmer et al., 2006), two labelers manually mapped 601 subject terms to the AAT. Inter-annotator agreement for this task was encouragingly high, at 91%, providing a notional upper bound for automatic system performance (Gale et al., 1992). We have used SenseRelate (Banerjee and Pederson, 2003; Patwardhan et al., 2003) for disambiguating AAT senses. SenseRelate uses word sense definitions from WordNet 2.1, a large lexical database of English nouns, verbs, adjectives, and adverbs.[4]

Results from our evaluations (discussed in Sidhu et al, 2007) show that mapping to WordNet first and then to the AAT causes errors. As a general resource, WordNet is domain independent and thus offers wider, more comprehensive coverage. However, the lack of domain specificity also creates overhead as there are many irrelevant senses to choose from and the correct sense needed for art and architecture discourse may not be available. Similarly, Iyer and Keefe (2004) report on an exploratory study on the use of WordNet to clarify concepts for searching architectural visual resources. Twenty participants were shown images which they were asked to locate using natural language or WordNet terms. Although 70% of participants stated that WordNet clarified the terms or the images, 30% reported problems with conceptualizing the image, and 55% had terminology problems. To address these types of problems, we are exploring the option of re-implementing concepts behind SenseRelate to directly map terms to the AAT. Additionally, in Future Work we will test approaches for employing hybrid techniques (including machine learning) for disambiguation. This will enable us to explore the trade-off in precision between different configurations of resource calling.

5.1.1. Catalog Record Creation: Select

As shown in Figure 2, a cataloger is presented with the image to be cataloged, the text segment associated with the image, and a number of index terms suggested by the Toolkit. The user decides which of the terms proposed by the CLiMB system should be included in the image's record.

5.2 Testbed Collections

We are currently working with five image-text sets and one image collection for which we are conducting experiments with dispersed texts located online. Table 1 illustrates the relationship between the associated texts and the image collections which we use to test our system.

Feedback from catalogers indicates that one thesaural resource is insufficient for cataloging a range of art historical and architecture images. The Getty resources are extensive but, as with any resource, are not entirely comprehensive. Our goal is to expand our capabilities for disambiguating domain-specific terminology by cross-searching multiple, established thesauri in the art and architecture domain. Resources currently under consideration include Iconclass [5] and the Library of Congress' Thesaurus for Graphic Materials (TGM) I and II[6,7].

6. Conclusion and Future Work

The CLiMB project techniques exceed simple keyword extraction and indexing by:

- applying novel semantic categorization to text segments,
- identifying and filtering linguistically coherent phrases,
- associating terms with a thesaurus, and
- applying disambiguation algorithms to these terms.

Although each of these techniques has been used in other projects, they have not been combined and tested in the art

[4] http://wordnet.princeton.edu/

[5] http://www.iconclass.nl/

[6] http://www.loc.gov/rr/print/tgm1/

[7] http://www.loc.gov/rr/print/tgm2/

and architecture domains for improving digital library access. Our future work will consist of three foci:

- Integration of functional semantic categorization with disambiguation
- Improvement of disambiguation
- Testing the system and its components with users to drive improvements

We also hope to incorporate the output of CLiMB text data mining with a social tagging approach to image labeling, such as that of steve.museum to examine terminological comparisons and their impact on image access.

Acknowledgements

We acknowledge input on the project from Dr. Murtha Baca of the Getty Vocabularies Institute. We also appreciate ongoing input from both our internal and external advisory boards, the members for which are listed on our web pages. Finally, many users whom we cannot name have helped in evaluations along the way. The project has been funded by the Andrew W. Mellon Foundation to the Center for Research on Information Access at Columbia University. Later support was provided to the College of Information Studies and the University of Maryland Institute for Advanced Computer Science at the University of Maryland, College Park, MD.

References

Banerjee, S., Pedersen, T. (2003) Extended Gloss Overlaps as a Measure of Semantic Relatedness. In: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, pp. 805–810.

Berinstein, P. (1999). Do you see what I see? Image indexing principles for the rest of us. Online, 23(2), 85-87.

Choi, Y., Rasmussen, E. M. (2003). Searching for images: The analysis of users' queries for image retrieval in American History. Journal of the American Society for Information Science and Technology, 54.6: 498-511.

Gale, W., K. W. Church, and D. Yarowsky. (1992) “Estimating upper and lower bounds on the performance of word-sense disambiguation programs.” Proceedings of the 30th conference on Association for Computational Linguistics. pp. 249-256.

Gardner, Helen. (2001) Art through the Ages. 11th edition. Edited by Fred S. Kleiner, Christin J. Mamiya, Richard G. Tansey. Harcourt College Publishers, Fort Worth, Texas.

Iyer, H., & Keefe, J. M. (2004). WordNet and keyword searching in art and architectural image databases. VRA Bulletin, 30, pp. 22-27.

Janson, H. W. (2004) History of Art: the Western tradition. Revised 6th edition. Upper Saddle River, NJ: Pearson/Prentice-Hall.

Jorgensen Corinne, and Peter Jorgensen. (2005) Image querying by image professionals. Journal of the American Society for Information Science and Technology. Hoboken: Oct 2005. Vol. 56, Iss. 12; pg. 1346.

Klavans, Judith L. (2006) Computational Linguistics for Metadata Building (CLiMB). In Procedings of the OntoImage Workshop, G. Greffenstette, ed. Language Resources and Evaluation Conference (LREC), Genova, Italy.

Palmer, Justin (n.d.) Optimizing Your Images for Google Image Search - 4 Image SEO Tips. http://ezinearticles.com.

Palmer, M., Ng, H.T., Dang, H.T. (2006) Evaluation. In: Edmonds, P., Agirre, E. (eds.): Word Sense Disambiguation: Algorithms, Applications, and Trends. Text, Speech, and Language Technology Series, Kluwer Academic Publishers, Netherlands.

Panofsky, E. (1962). Studies in Iconology: Humanistic Themes in the Art of the Renaissance. Harper & Row, New York .

Passonneau Rebecca, Tae Yano, Judith Klavans, Rachael Bradley, Carolyn Sheffield, Eileen Abels, Laura Jenemann (2007) Selecting and Categorizing Textual Descriptions of Images in the Context of an Image Indexer’s Toolkit. Technical Report. Columbia University.

Passonneau Rebecca J., Tae Yano, Tom Lippincott, and Judith Klavans. (2008). Functional Semantic Categories for Art History Text: Human Labeling and Preliminary Machine Learning. International Workshop on Metadata Mining for Image Understanding; VISAPP International Conference on Computer Vision Theory and Applications. MMIU 2008. Funchal, Madeira - Portugal.

Patwardhan, S., Banerjee, S., Pedersen, T. (2003) Using Measures of Semantic Relatedness for Word Sense Disambiguation. In: Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics, Mexico City.

Shatford-Layne, S. (1994): “Some issues in the indexing of images.” Journal of the American Society for Information Science 45.8. pp. 583-588.

Sidhu, T., Klavans, J.L., Lin, J. (2007) Concept Disambiguation for Improved Subject Access Using Multiple Knowledge Sources. In: Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTech 2007), 45th Annual Meeting of the Association for Computational Linguistics. Prague, Czech Republic

Toutanova, K., and C. D. Manning. (2000) Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. EMNLP/VLC 2000. pp. 63–70.

Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. (2003) Feature-rich part-of-speech tagging with a cyclic dependency network. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, vol. 1. Edmonton, Canada: Association for Computational Linguistics. pp.

173-180.

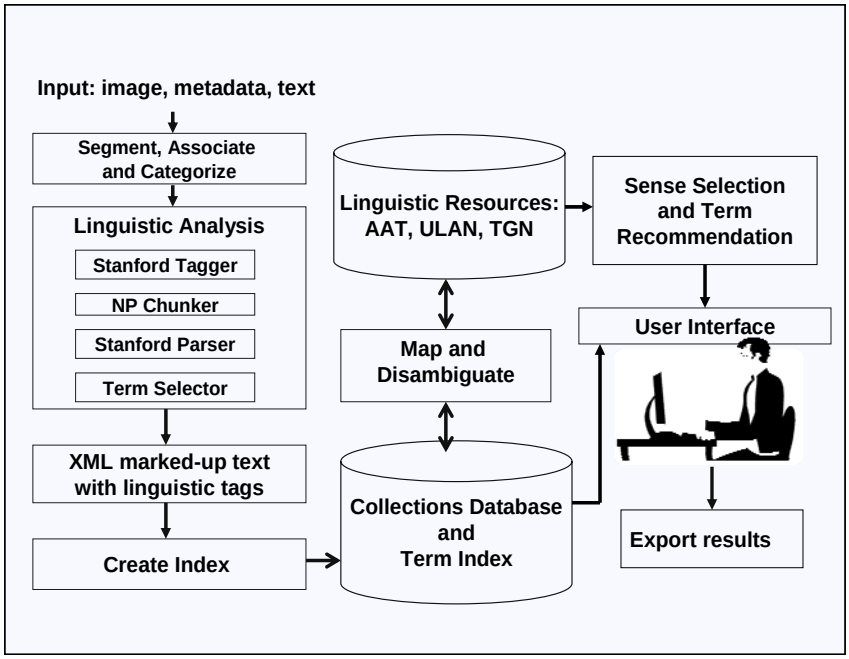

Figure 1: CLiMB Architecture

Image Collection	**Text**	**Image/Text Relationship**
National Gallery of Art (NGA) Online Collection	Narratives associated with images on the NGA website	Integrated
U.S. Senate Museum	U.S. Senate Catalogue of Fine Arts	Integrated
The Vernacular Architecture Forum (VAF)	VAF Field Guides:	Integrated
The Society of Architectural Historians (SAH): World Architectural Survey and the American Architectural Survey	*Buildings Across Time: An Introduction to World Architecture* by Marian Moffett, et al.	External
Landscape Architecture Image Resource (LAIR)	*Landscape Design: A Cultural and Architectural History* by Elizabeth Barlow Rogers	External
Art History Survey Collection (AHSC), ARTstor	Disparate texts located online	External

Table 1: Sources of Image and Testbed Text Collections

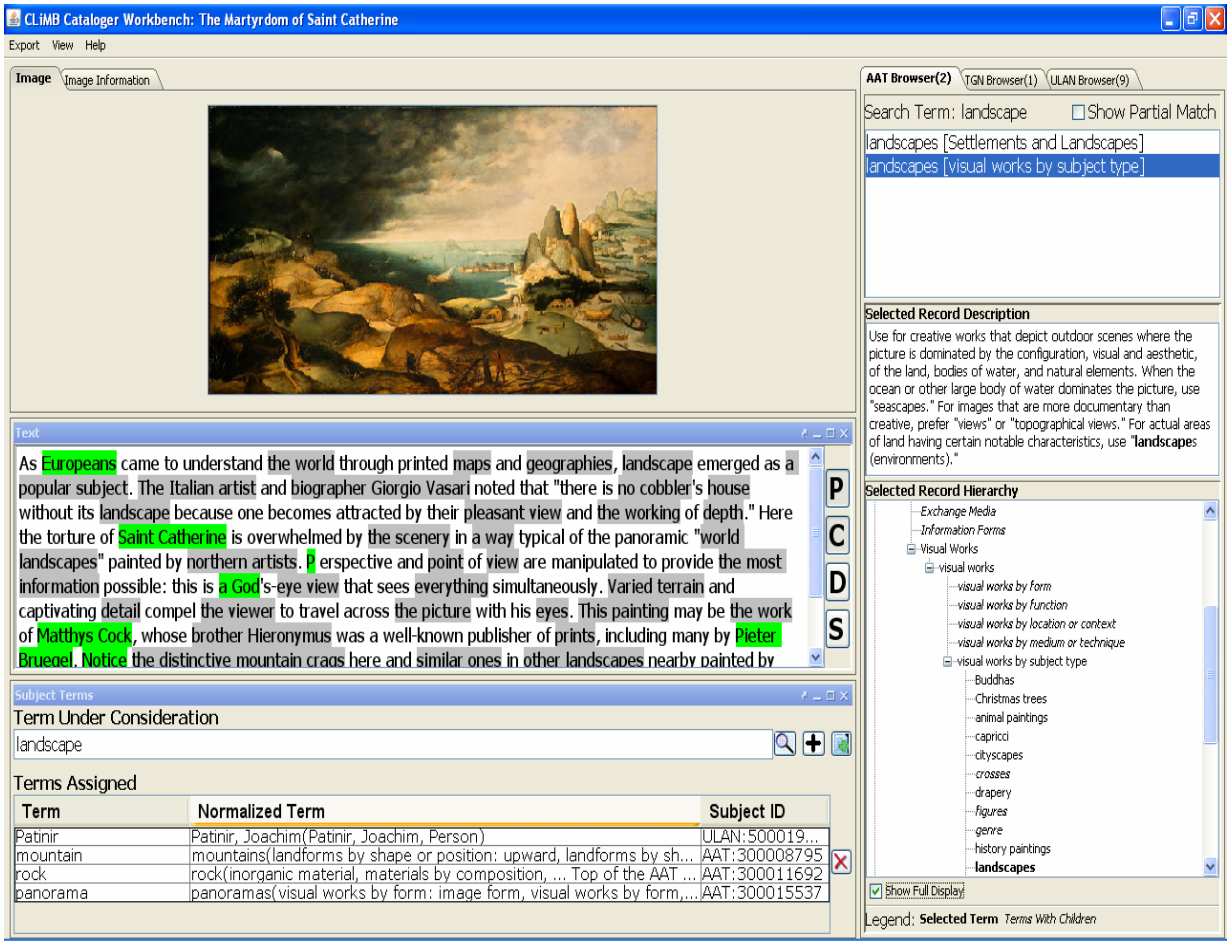

Figure 2: CLiMB User Interface for the term “landscape”