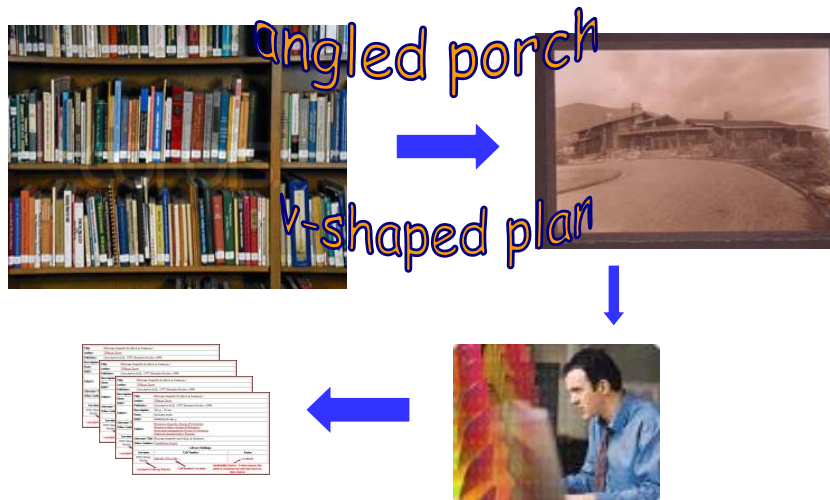


CLiMB ToolKit: A Case Study of Iterative Evaluation in a Multidisciplinary Project

Rebecca Passonneau, Roberta
Blitz,

David Elson, Angela Giral
Columbia University

Judith Klavans
University of Maryland



Motivation

- Fast-growing collections of digital images
- Image search using keywords
- High cost of manual indexing/cataloging
- Potential for automated mining from texts about images

Types of Resources Available

NLP Tools/Knowledge

- POS taggers
- Chunkers
- NamedEntity recognizers
- WordNet
- ML toolkits

Art or Library Knowledge Sources

- Getty Art & Architecture Thesaurus (AAT)
- Library of Congress name and subject list
- Library of Congress Thesaurus of Graphic Materials

Iterative Evaluation Process

1. Formative Evaluation: How to optimize use of NLP/Thesaural resources
 - Conducted after creating a development environment to extract potential terms from texts
 - Participants: heterogeneous users
2. User Study: How to investigate a proposed work process before it exists
 - Conducted after creating CLiMB ToolKit (Image cataloger's workbench)
 - Participants: catalogers and image professionals

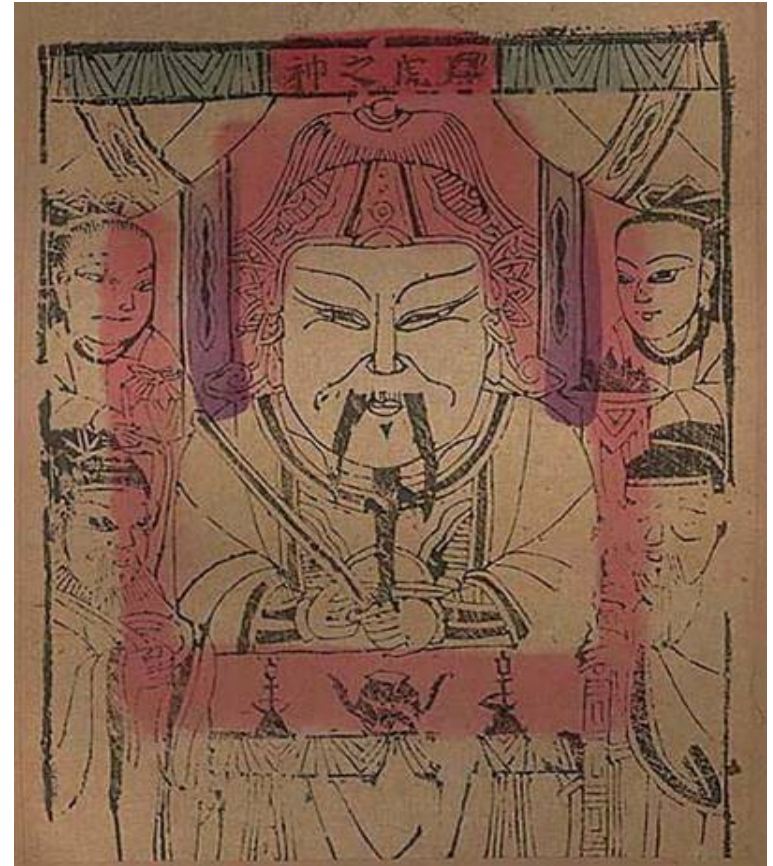
Text Collection Sets (TCS): Criteria

- Image Collection
 - Substantial collection of related images in digital form
 - Authoritative list of images (E.g., database UIDs, referred to as Target Object Identifiers - TOIs)
 - Associated electronic texts
- Text(s)
 - Discussion of many items depicted in the images
 - Authoritative discussion of image content

Formative Evaluation:

TCS1: Chinese Paper Gods

- **Image Collection:**
Anne S. Goodrich Collection of Chinese Paper Gods, C.V. Starr East Asian Library
- **Texts:**
Goodrich, Anne S. *Peking Paper Gods: A Look at Home Worship*. Nettetal: Steyler Verlag, 1991.



Formative Evaluation:

TCS2: Greene & Greene

- **Image Collection:**

Greene & Greene Collection of Architectural Records and Papers, Avery Architectural and Fine Arts Library (G&G)

- **Text Collection:**

1) Bosley, Edward R. *Greene & Greene*. London:Phaidon, 2000.

2) Makinson, Randell L. *Greene & Greene*. Salt Lake City : Peregrine Smith, c1977-1979.

3) Current, William R. *Greene & Greene: Architects in the Residential Style*. Fort Worth: Amon Carter Museum of Western Art [1974].



Formative Evaluation: Design

- Two-part survey using two of four conditions
 - User Scenario
 - Image
 - Free Text
 - CLiMB Checklist
- Thirteen participants who completed the survey
 - Librarians, art historians, computer scientists, computational linguists
- Partly crossed design

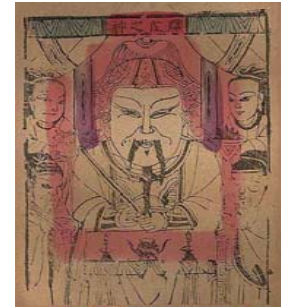
Two Non-Text Conditions

- User Scenario

In this task, the survey item contained one of two hypothetical user scenarios. Respondents were asked to list keywords and phrases that could be used “to search for relevant images in an image database.”

1. *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?*

- 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.*”

Two Text Conditions

- **Free Text:**

This task contained a passage from one of the texts associated with TCS1 or TCS2.

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.”*

1. *Please circle 10 words or phrases as your top choices.*
2. *Please underline 10 as your second tier choices.”*

- **CLiMB Checklist:** Respondents were given a long list of words and phrases (117 TCS1 entries; 194 TS2 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.”*

- _____ garden pergola
- _____ dark green tile
- _____ ridge beams

Overview of Responses

- **User Scenario:** fewest terms proposed, very general terms (*home, exterior*)
- **Image:** About 10 terms on average, most for Survey 3, least for Survey 1 (**brick, driveway**)
- **Free Text:** very specific terms, some similarity to CLiMB terms (**garden pergola, dark green tile**)
- **CLiMB Checklist:** Significant overlap of terms selected by many humans, and terms with high CLiMB weights (**plaster frieze, ridge beams**)

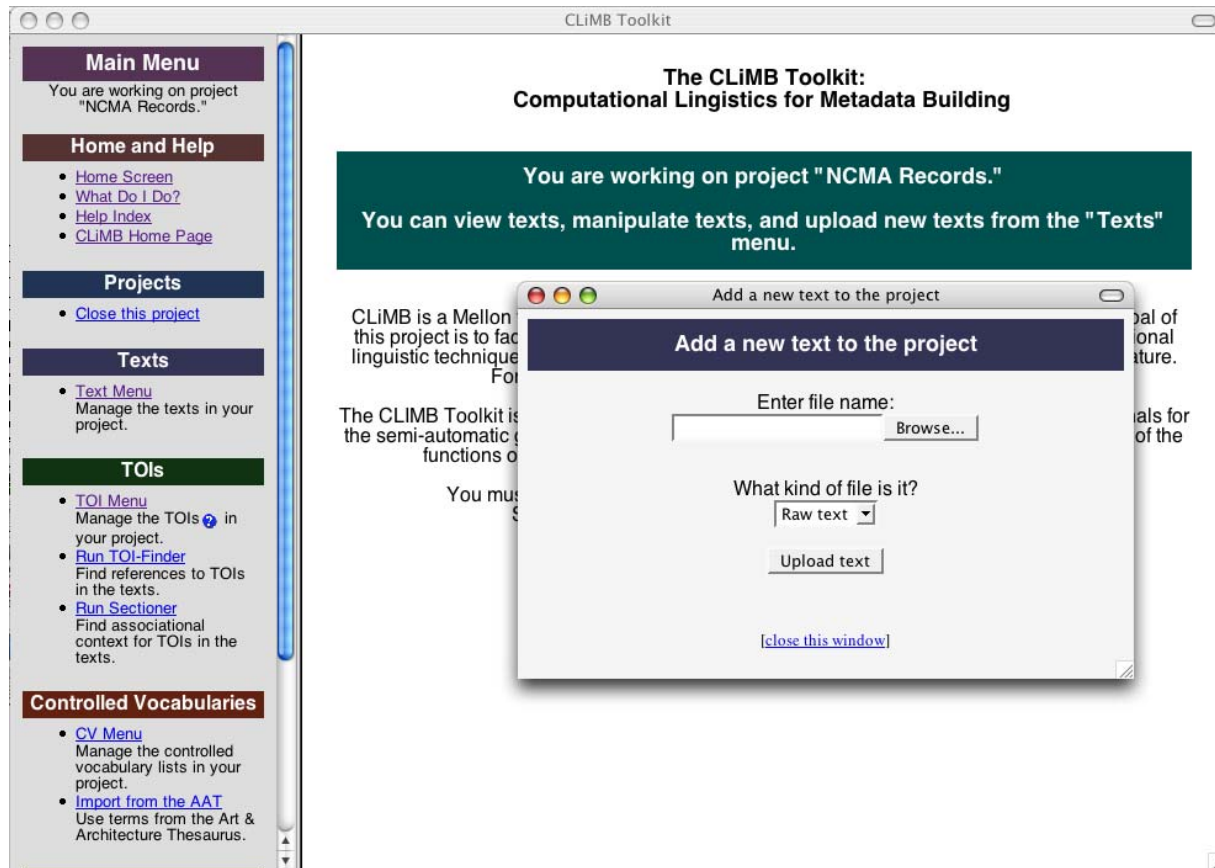
Comparison

Free Text Terms and CLiMB Checklist

RESULT: Significant overlap of high ranking terms by humans with high ranking CLiMB terms

INTERPRETATION: ToolKit will assist catalogers better if it proposes terms

CLiMB ToolKit



CLiMB ToolKit Functionality

- 1) Loading and initialization of raw (ASCII) text
- 2) After initialization, text could be processed by a noun phrase chunker (termed “chunking”)
- 3) An image TOI list could be loaded or manually created
- 4) TOI Finder could be run to locate references to TOIs in the loaded texts
- 5) Texts that had been processed by the TOI Finder could also be sectioned into associational contexts correlated with specific TOIs
- 6) Lists of Controlled Vocabulary could be loaded—including in this feature users were provided access to the Getty Art & Architecture Thesaurus (AAT), and the capability of selecting specific subsets from the AAT
- 7) A Noun Phrase detail frame was available, e.g., to illustrate intersections of text phrases with AAT.

User Study TCS

Tight relation of text to image

TCS 3: NCMA

- **Image Collection:** North Carolina Museum of Art website, Highlights of the Collection
- (<http://www.ncartmuseum.org/collections/highlights.shtml>)
- **Texts:** *North Carolina Museum of Art: Handbook of the Collection (NCMA Handbook)*

Georgia O'Keeffe (American,
1887-1986)
Cebolla Church, 1945



User Study Design

- Two Interleaved User Activities:
 - User text processing and metadata tasks using sample of current TCS
 - Questions on a 1 to 5 scale, with 1 most positive, 3 neutral, 5 most negative
- Ten librarians, image professionals and metadata professionals

User Tasks

- Questionnaire stepped users through each of the 7 ToolKit functions
- Immediately after, users were instructed to create an entire *project* from scratch by loading designated images and texts, and creating metadata for three images

Scaled Questions: Examples

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

Sample Results on Scaled Qs

Question	Group Mean	Std Dev
11.12	1.2	0.33
11.13	1.2	0.33
11.15	2.0	0.78
15.18	1.9	0.33
16.20*	3.3	0.33

Sample Metadata Results

Rank	Terms (optional terms)	N Respondents
1	<i>vanitas (image)</i>	9
2	<i>(Dutch) still life (painting)</i>	8
3	<i>(burned down) candle</i>	7
4	<i>glass</i>	6
5	<i>pewter</i>	6
6	Dutch	5
7	empty glass	5



Jansz den Uyl
Banquet Piece (NCMA
52.9.43)

Results

- Task Success
 - All users completed the metadata task
 - Many terms selected by many users
- User Satisfaction
 - Starts out high, remains high
 - Question with lowest “score” was average (3.3) after hardest step (at 16.2)