CLiMB-1 Evaluations

Feedback and Testing at Critical Stages

• Formative Evaluation, October 2003:
  *Defining Goals for CLiMB Image Cataloging Toolkit — Solicit expert advice*

• Prototype Toolkit Evaluation, March 2004:
  *Iterative Design — Elicit cataloger feedback during development*
Formative Evaluation: 4-part Questionnaire

How *many* terms, and *what types* of terms, do various experts (librarians, image professionals, computer scientists) suggest for images:

A. Given a sample search request? (Scenario)
B. When they are shown an image? (Image)
C. When they have texts about images? (Text)
D. When they have a list of candidate terms from CLiMB tools? (Terms)
Questionnaire: Scenario

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?
Questionnaire: Image
The Fleishhackers specifically requested that the Greenes design for them an English-style house, with a thatched roof, if possible, and they included the landscaping and gardens in the commission.

The main residence, designed primarily for summer use, was constructed around a dog-leg plan under a hipped roof, the main axis parallel to the front elevation and the rear terrace. The Fleishhackers did not care for the Greenes legendary wood interiors, but requested instead a white-tinted plaster finish throughout.

The material of choice for the Fleishhacker exterior was the new and innovative product (more correctly a process) called Gunite, a fine cement that was applied with a patented pneumatic gun to a prepared surface (another indication that the Greenes were not set in their ways). Though the Fleishhackers had asked for a thatched roof, they got something more permanent and practical. A sawed-shingle roof was applied instead, built up in thick, undulating rows.
Questionnaire: Terms

___Arts and Crafts  ___garden walkway

___cladding

___extensive water gardens

___interior fittings

___modest frontal elevation
Survey Responses: Overview

- **Scenario:** fewest terms proposed, very general terms (*home, exterior*)
- **Image:** About 10 terms on average, still somewhat general (*brick, driveway*)
- **Text:** many terms; very specific terms; similarity to CLiMB terms (*pergola, thatched roof*)
- **Terms:** Significant overlap of terms selected by many humans, and terms with high CLiMB weights (*plaster frieze, ridge beams*)
Analyze Responses for Terms

• Create consensus ranking of terms by aggregating all checklist responses

• Compare with CLiMB Toolkit weighting of terms
Conclusion

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

INTERPRETATION: CLiMB Toolkit will assist catalogers best if it proposes terms
Toolkit Evaluation Questions

• Can catalogers understand the Toolkit?
• Can catalogers accomplish Toolkit steps?
  – Load texts
  – Load lists of image identifiers (TOIs: Target Object Identifiers)
  – View candidate terms
  – Select terms
• How quickly can catalogers work on their own?
Two Metadata Tasks

• North Carolina Museum Collections
  – Image gallery
  – Catalogue of the Collections

Banquet Piece, Jan Jansz. den Uyl

• Greene & Greene Collection of Architectural Images
  – Photographs
  – Scholarly books
Two Evaluation Metrics

• Task Success:
  – Can catalogers complete the task of selecting subject access metadata?
  – Are the selected terms high quality?

• User Satisfaction:
  – Do catalogers like the Toolkit?
  – Does satisfaction remain constant over use?
Task Success

• **Completion:**
  – All completed the North Carolina task
  – Some completed both tasks

• **Performance:** *Banquet Piece* example
  – 35 terms
  – 5 of the terms were selected by a majority of respondents
Banquet Piece

VRA CORE 3.0 record enhanced with the 5 *majority vote* terms

Record Type = work
Type = painting
Title = Banquet Piece
Measurements. Dimensions = 79.7 x 94 cm

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
On a Scale of 1 to 5
(Where 5 is excellent)

Fifteen scaled questions:

• **Result:**
  – Average of all *scaled* questions is 4
  – Very few negative scores (1 or 2)

• **Interpretation:**
  – Respondents are satisfied with all aspects of Toolkit functionality
Does User Satisfaction Stay High?

Same question about overall satisfaction appears 3 times in the questionnaire

- First time — after introduction: Avg.=4
- Second time — after first metadata selection task: Avg.=4
- Third time — after completing first collection (3 North Carolina images): Avg.=4
Learning: Easy

- **Context:** Users were shown a single example to illustrate all CLiMB functions.
- **Just before question:** Users were then pointed to some data and with *no further guidance* were told to load, process and view the texts.
- **Note:** All users succeeded.
- **Asked about difficulty,** average response was 3 (3 is neither positive nor negative).
Core Functionality: High Satisfaction

Final scaled question: I was ____ with the process of selecting descriptive metadata

5. Extremely satisfied . . . . . 1. Very dissatisfied

• RESULT:
  – Average: 4.6, or very high
  – Same score as easiest question

• INTERPRETATION:
  – Users feel satisfied overall
  – Users believe Toolkit assists metadata selection
Conclusions

• Toolkit functions were easy to understand
• Toolkit actions were easy to perform
• Toolkit facilitates creation of subject access metadata that would otherwise be omitted
CLiMB-2 Evaluations

• Continued evaluation of effectiveness of toolkit as it evolves: Image professionals
• Evaluation of usefulness in work setting: Image catalogers in work setting
• Evaluation of effect of CLiMB generated terms on searching and browsing: End users of all types
• Suggestions?