

Computational Linguistics for Metadata Building: Enhancing image access through linguistic techniques



Judith Klavans, Jimmy Lin, Rebecca Passonneau, Eileen Abels, Carolyn Sheffield, Dagobert Soergel, Tandeep Sidhu

1 Identify relevant text for an image and suggest terms

a Minimal metadata for image, no descriptive terms.

b Excerpt of descriptive text from Gardner (v. 11, pl. 3-33), suggested CLIMB terms highlighted in yellow

The famous painted limestone bust of Akhenaten's queen, Nefertiti (fig. 3-33), exhibits an almost unerring sensitivity and delicacy of curving contour. The piece was found in the workshop of the queen's official sculptor, Thutmose, and is a deliberately unfinished model very likely by the master's own hand. The left eye sockets still lack the initial eyeballs, making the portrait a kind of before-and-after demonstration piece. With this elegant bust, Thutmose may have been alluding to a heavy flower on its slender stalk by exaggerating the weight of the crowned head and the length of the almost serpentine neck.

- a) Many catalog records contain minimal subject-oriented access points
- b) CLIMB extracts information from scholarly texts for subject fields

2 Conceptual analysis and term extraction

Conceptual Analysis: One Paragraph, Three Semantic Categories

[Historical Context] Of the great projects built by Akhenaten hardly anything remains . . . Through his choice of masters, he fostered a new style.

[Implementation] Known as the Amarna style, it can be seen at its best in [Image Content] a sunk relief portrait of Akhenaten and his family.

Key
Image Content: Objective or allegorical content
Historical Context: Context of creation
Implementation: Style or technique

Problems in image indexing and access

- too many images available with only limited access points
- subject-oriented cataloging is expensive, time-consuming, and requires domain expertise

Solution

- apply computational linguistic techniques over text about images to identify potential subject terms
- populate existing catalog records with these subject terms for enhanced access

Methodology

- create a cataloger's toolkit for mining scholarly art historical texts for subject-oriented metadata
- identify relevant text segments and conceptual function
- disambiguate terms and phrases by mapping to the appropriate sense in the integrated thesaurus
- use mined terms to guide cataloger's process of creating subject terms
- test the efficacy and effectiveness from the perspectives of both catalogers and end users

www.umiacs.umd.edu/~climb

3 Mapping terms to thesauri

Search Term: commission
 AAT Record Displayer
 ID#: 300024870 Record Type: C
 Name: commission

Terms:
 commissions (preferred,C,PLD)
 commission (CU,AD)

Catalogers: use thesaurus for controlled vocabulary
End users: expand or narrow a search through hierarchies

4 Export terms to catalog record

Main Entry: Category or Artist
 Egypt (New Kingdom)

Title or Description
 Queen Nefertiti. (Front view), ca. 1350 B.C. (Dynasty XVIII)

[Location] - AICT Photo ID #
 Aegyphtech Museum, Berlin #EN010

CLIMB Terms
 busts [figure], art id: 300011286
 Akhenaten queens [monarch], art id: 300025483

Text References
 Adams 3: pl. 5:22 [alt]
 Adams 4: pl. 5:22 [alt]
 Adams AAT: pl. 4:42 [alt]
 Adams AAT2: pl. 3:38 [alt]
 Gardner 11: pl. 3-33
 Gardner 12: pl. 3-33