

COMPUTATIONAL LINGUISTICS FOR METADATA BUILDING: ENHANCING SUBJECT ACCESS THROUGH LINGUISTIC TECHNIQUES

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

1 Typical Record: Image with minimal metadata



Problems in image indexing and access

- Many images available with limited access points
- Subject-oriented cataloging is expensive, time-consuming, and requires domain expertise

Solution

- Apply computational linguistic techniques over text about images
- Identify potential subject terms
- Populate existing catalog records

2 Texts contain rich subject terms

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



The famous painted limestone bust of Akhenaton's queen, Nefertiti (fig. 3-33), exhibits a similar expression of entranced musing and an almost mannered 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 socket still lacks the inlaid eyeball, 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...



Computational Linguistics Toolkit

- Available for download
- Being tested with catalogers

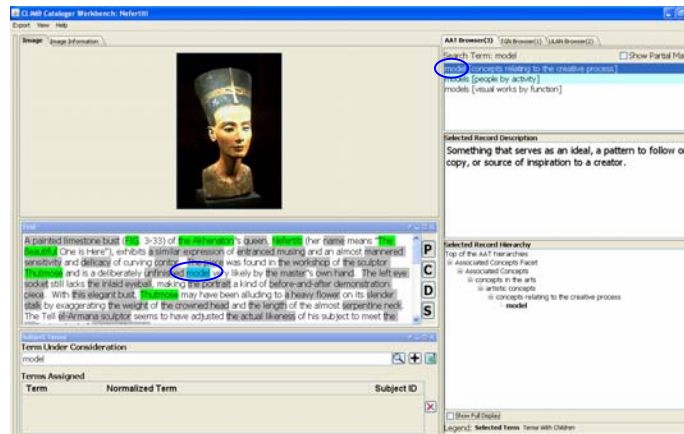
Benefits for Catalogers

- Provides candidate terms for catalogers to judge
- Links to thesauri (e.g. AAT, ULAN, TGN)
- Exports automatically to multiple metadata formats

Methodology

- Mine scholarly texts for subject-oriented metadata
- Identify relevant text segments and conceptual function
- Disambiguate terms and phrases by mapping to integrated thesauri (AAT, ULAN, TGN)
- Gather user feedback and perform iterative improvements

3 Mapping extracted terms to thesauri



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

4 Export terms to catalog record

