Art Images Online: Leveraging Social Tagging and Language for Browsing
Judith Klavans¹, Jen Golbeck¹, Susan Chun², Rob Stein³, Ed Bachta¹, Irene Eleta¹,
Raul Guerra¹, Rebecca LaPlante¹
http://umiacs.umd.edu/research/t3/
University of Maryland¹, Independent Museum Consultant², Indianapolis Museum of Art³

One of the most significant benefits and challenges of using crowd-sourcing is that it enables the collection of large sets of user-generated tags. However, the first view of this data usually sends both the cataloger (museum or library) and computer scientist running for help: the number of misspellings, the ambiguity of terms, multiple languages, the use of partial phrases and overall level of specificity (sometimes very general and sometimes highly specific) creates a sense of chaos not found in a careful catalog. At the same time, the richness of the descriptive data creates depth usually unable to be input into a traditional catalog record.

In this session, we present approaches to solving these problems. We report on results of research for the IMLS Leadership project “T³: Text, Tagging and Trust to Improve Image Access for Museums and Libraries”. We focus on two issues: (1) we present the computational linguistic processing, and morphological and semantic analysis techniques used to analyze the large steve.museum tag set. Our computational techniques enable us to examine tags and phrases of importance for browsing, and to discriminate useful multi-word phrases with high descriptive value. (2) we present results on a comparison of social tagging patterns in two languages, while seeking exploitable strengths for providing multilingual support in digital libraries and museums.

Computational Linguistic Processing of User-Contributed Tags and Associated Text

We have examined and evaluated computational linguistic tools to create a tag processing pipeline. Architecturally this pipeline is a software module at the back end of a tagging website like Steve in Action (www.steve.museum). The website’s front end gathers user tags, and then the pipeline processes a stream of tokens and the associated metadata.

Among the processes required to normalize the variability in tags are lemmatization (i.e. finding the base form of words, such as book/books or sleep/sleeps/sleeping); part of speech (POS) tagging (e.g. book-Noun; books-PluralNoun; sleep-Verb; sleeping-VERBing); named entity identification (e.g. New York, Gauguin); and multi-word phrase analysis to determine if the phrase is a fixed item (e.g. phone cord vs. standing woman). In order to evaluate tools for each of these complex steps, we created a hand-labeled tagset consisting of 850 user-tags (500 one word terms, 200 two word terms, 100 three word terms, 50 four word terms). We ran three lemmatizers, and evaluated against the manual “gold standard”, selecting NLTK (www.nltk.org). For POS tagging, we ran three tools, and selected the Stanford MaxEnt Tagger which achieved 79% accuracy (nlp.stanford.edu/software). This is a very high result since there is no context on which to select POS; also, tags have many misspellings and foreign words. The tools selected are all available for use in other applications.

To address some of the Part of Speech ambiguity of one-word tags, we have looked at calculating the log likelihood ratio of the collocation of one-word tags that co-occur together in the same tag cloud. The reasoning for this is that bigrams (two-word phrases) tend to reduce some of the POS and sense ambiguity of one-word tags. We observed, for example, that one-word tags (one-word tags ‘green’ and ‘poplar’) co-occur, and they are also collocations according to the Google N-Gramm corpus, which can be used to determine cohesion of bigrams.

We also report on research to obtain useful phrases from an images’s handbook text using phrase finders (Chunkers) and Named Entity Recognizers. Our results have shown that very few images have handbook text and this text talks more about the context of the image (i.e. the painter) rather than the work of art.
Leveraging Multilingual Tags

The language barrier in digital libraries and museums is more present than ever due to the language diversity of Internet users. Also, collections are composed of images of art originating from different cultures whose peoples might want to access using their own language, but often those images are annotated only in the languages of the institutions that display them. If we are able to translate tags from one language to another or to enable the input of tags in multiple languages through localized user interfaces, it may enrich the annotations.

Once we have social tags in multiple languages, they could enable multilingual search with the help of machine translation, where the user searches in a language different from the language of the query.

We report on the results of a tagging experiment over a collection of 33 images of paintings of five types for which tags in Spanish and English were collected. We analyzed these social tags to measure the agreement between speakers of the same language and speakers of two different languages; then we compared tags and tag categories to draw conclusions about tagging behaviors across languages. Our results indicate that the most frequent category was “general person or thing”, followed by “visual elements”, which was especially frequent the non-representational type of paintings, and by “non-subject matter” (i.e. painter names, nationalities and styles). Interestingly, the category “emotions and abstract ideas” was very common; for instance, terms such as tranquility, fear, and justice. Other categories included “general action or state”, “specific person or thing” (i.e. biblical names), and “general place”. These results are consistent with findings over images of representational works in Klavans et al. 2011.

Regarding the similarity of tags across languages, we found that often (45%) the most popular term in one language was an exact translation of the most popular term in the other. Also, in 36.36% of images, the most popular term in one language matched an exact translation among the second or subsequent most popular terms in the other. In line with the categorization results, these frequent terms that favored consensus across languages fell in the categories “general person or thing” (elephant, woman), “visual elements” (green, blue), “non-subject matter” (abstract), “specific persons or things” (peacock, Jesus). Despite “emotions and abstract ideas” being a popular category, the frequent terms were rarely found in it.

One unexpected finding uncovered that American participants used on average twice as many terms for tagging an image as the Spanish participants (6.78 vs. 3.41); the reasons for this difference remain unclear. Future research could explore if the consensus level correlates with the distance between cultures and languages in the tagging community. For instance, if instead of Spanish we add Chinese tags for the same paintings, would that impact the level of consensus?

Selected References

Project website: http://umiacs.umd.edu/research/t3/