

Cataloguing by Crowd Workshop: Grindstone Island: July 2005

Issues in Term Analysis July 23, 2005

What do we want to know about terms? (analysis)

- Are descriptions necessarily multi-word instead of single term?
- How much do people switch language within sessions?

- When do miss-spellings have value?
- What language are the terms in?
 - Can we algorithmically identify foreign language terms?
 - With some hints from user profile (like preferred language)
 - Could you pop-up a 'language question' box if you can't identify the language in use?
- What is saturation for new term assignment? A number of views after which nothing new is added?
 - How can you use this to guide how long works remain within the system and how many users they are shown to?
- Is there a difference between terms used for searching and those used for description?
 - Guess the word that describes the image?
 - How much synonymy is there?
- Is there anything that is added to existing colour analysis algorithms that results from human viewing?
 - Which are the best colour recognition tools?
- Can direction of users guide their terms?
 - Compare terms where areas prompted and not (with tools like grid/quadrants, magnifying glass)
 - Compare terms where facets / guide terms given or not?
- What parts of speech are supplied?
 - Nouns, verbs, adjectives?
 - How does this relate to stemming / normalization / synonymy in retrieval?
- Do black and white images produce different results than colour ones?
 - Compare terms from colour / black and white images
 - Could you process images to guide viewing?
- How does image quality affect results?
 - How much resolution or detail is needed to get ok results?
 - How big?
- Does sequence influence term relevance?
 - Are there 'better' terms at the beginning or end of a session?
 - How does presence /absence of label copy influence terms?
 - How does 'speed' influence terms?
 - Are the same terms applied to multiple images of the same object?

- BW vs colour of the same work?
- Do multiple images produce better work?
- Does juxtaposition elicit different results? Are there relationships between works that could be recorded? Word-net (semantic network) searches / relationships?
 - Show works that have the same terms so that differences are highlighted?
 -

What do we want to do to our terms? (processing)

- Search them within the tool?
- Compare them to other vocabularies?
 - Do these terms appear there?
 - How many of our terms don't appear there?
 - If they don't appear, should they?
 - Where might they fall?
 - Could we "Hierarchicalize" terms that are otherwise just lists?
 - What about other local lists of terms?
 - How do we share what we learn / use what others already know?
- What kinds of terms are proposed that relate to 'specialist' aspects of works?
 - Artist, date, style
 - How accurate is this?
 - Could it be used or does it all have to be 'thrown out'?
- Percentile review of terms assigned to the same work
 - Thresholds that define legitimacy?
 - Way to eliminate use-less terms?
- How can we 'weight' terms?
 - Saturation in new terms BUT what about which ones are *more* appropriate
- Compare terms assigned by different users
 - Professional vs amateur
 - First-time user vs long-term user
 - Specialist vs general public
 - Users of different ages?
- Can we segment terms by characteristics + use this to cluster context-specific or context-appropriate terminology?
 - Age
 - Object
- How does an individual's work change over time?
 - By time
 - By # of terms
 - By accuracy
 - By session
 - By weighting
 - In relation to defined thresholds

- How often is a work ignored?
-
- How do we group terms by type of user so that they can be processed in a different manner?

What do we want to do with our terms? (deployment)

- Is it any different than all of the other terms we have?
- Do we segregate them from other terms?
- Are there cases where you would just use folksonomic terms or would they always be combined with other data? Do the terms have stand-alone value?
 - Is this a politically defined question?