

**Research Questions: A Summary of Issues Raised in Discussion
July 23, 2005**

Assumptions

- Differentiate between prototype that is a research environment and a deployed application built as a result of research.
- What are the research questions and how are they addressed in the front / back end?

**Research Questions
(What is immediate and what is long term?)**

1. General Utility

1.1. What external sources also support this sort of research?

- Web logs
 - Terms used in existing search systems

1.2. Is this approach useful?

- Does it gather terms that help people find things?

1.3. How do we share what we learn / use what others already know?

2. Prototype Deployment = What short-term questions do we need to resolve to move ahead?

2.1. How do you compare across implementations?

- Does the experiment have different iterations / versions/ questions in different contexts?
- How can results compared across methods of research?
- If the same system / different collections might produce different results, what are the variables?

2.2. Language and multi-linguality?

- When and where is it important to enable language choice?
- How much of the supplied terminology is conflicting / confusing in a multi-lingual environment?

2.3. Do you get better results if images are presented in meaningful clusters?

- What are meaningful clusters?

2.4. What level of 'set' is required to create meaningful clusters of works for user analysis?

- How broad are the classes?
- How large / small do they have to be?

- How granular are the distinctions?

2.5. What level of monitoring / logging of user interactions and preferences is required? What data do you need to analyse the terms?

- Session configurations (are these external to system and time-defined)

3. User Interface = How the tool works? Does the tool work? How the tool should look? What effect do changes have? What works?

3.1. How do you design the interface to gather data that supports the 'research program'?

- What variables make sense to test in which combinations for which communities?
- Modes of use?

3.2. Is the random presentation of images more effective than the clustered presentation of related groups of images more efficient?

- More related terms
- More consistency across terms
- More terms

3.3. What influence does the flavour or personality of the user interface have?

- Research vs Gaming feel
- What are the effects of the 'gaming' functions?
- How do the terms gathered in game-like interfaces relate to the terms gathered in more 'seriously presented' interfaces?

3.4. As more user choice is defined, how do we keep them focused on the task at hand?

- When does the 'user cultivation' aspect get in the way of the term-collection motivation?
- Where's the balance?
- Could an enhanced art-viewing function impede cataloguing?
- How crowded is the interface ?

3.5. Is Speed a useful goal?

- Does increased functionality impede output / throughput?
- How does speed affect stickiness?
- Is a lean interface not as engaging?

3.6. Does inter-Cataloguer discussion improve terms?

- Chat about this image

- Use of the tool in a social situation

3.7. Does knowing about the work of art help the provision of terms?

- Do you show the cataloging information or not?

3.8. Do facets (categories of terms) improve term collection?

- Could the tool lead provision of terms by suggesting aspects of image to describe?
- What is the effect if facets are prompted at the outset vs user assigned after terms have been suggested?
- Does prompting for facets make terminology supplied more comprehensive?
- Do facets themselves improve the value of the terms (term in context)?

3.9. What interface aspects are influential in which aspects of use?

- Size of input box in relation to term

4. Term Analysis

4.1. General

4.1.1. Are the terms we get from user sessions living up to our expectations?

What are 'better results'?

- More terms
- More stickiness
- User completes task
- User comes back to task again
- More foreign language results
- More age-appropriate results
- More users (numbers of people adding terms)
- More effective retrieval in different environments
- More "related to" terms used in real search environments that don't now find results in on-line systems (terms against null results sets)
- What is "more"? Strict numerical?
- Do longer-term users produce "better terms" than casual users?

4.1.2. How are terms collected processed and analysed?

- Related to existing taxonomies / other controlled vocabularies
- Clustered in facets
- Statistically analysed
- Does any of this influence data collection strategies?
 - How can one support the other?
- How much of this is locally defined?
- Needs test data to seed discussion

- 4.1.3. How do you compare the results of data gathered from community cataloguing with data from other kinds of keyword harvesting methods (data mining of texts associated with works, paid cataloguing...)?
- 4.1.4. Can lists of null results in existing search environments serve as a proxy for “real users’ terms”?
 - How much of this kind of data is conditioned by users’ perceptions of these other environments?
 - Could we get more generic search terms from somewhere (like google images)?
- 4.1.5. Is Volume a goal? (Lots of users and terms)
 - Or do smaller, more specialized groups contributing more focused terms provide a more useful result?
 - Are these mutually exclusive?

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

- 4.2.1. Is there a difference in quality/results when a user is shown different object types?
 - Three vs two dimensional objects
 - Western vs Non-Western (in North American environment)
 - Different materials, Different time periods
 - Culturally ‘alike’ for user vs culturally ‘different’
- 4.2.2. If user is allowed to search for works to catalogue, should search use terms applied by others?
 - Does this help find more, or just create an ‘echo chamber’?
- 4.2.3. What are the differences between users driven to the tool from different places?
 - General portal
 - Museum pages
 - Project pages
 - Are these different enough?
- 4.2.4. Are descriptions necessarily multi-word instead of single term?
 - How much do people switch language within sessions?
- 4.2.5. When do miss-spellings have value?
- 4.2.6. What language are the terms in?
- 4.2.7. 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?
- 4.2.8.** 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?
- 4.2.9. Is there a difference between terms used for searching and those used for description?
 - Could you test this with a game that asks users to guess the word that describes the image?
 - How much synonymy is there in terms used in either context?
- 4.2.10.** Is there anything that is added to existing colour analysis algorithms that results from human viewing?
 - Which are the best colour recognition tools and how could they be deployed to make term provision more efficient?
- 4.2.11.** 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?
- 4.2.12. What parts of speech are supplied?
 - Nouns, verbs, adjectives?
 - How does this relate to stemming / normalization / synonymy in retrieval?
- 4.2.13. 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?
- 4.2.14. How does image quality affect results?
 - How much resolution or detail is needed to get ok results?
 - How big an image is needed?
 - How large should it be displayed?
- 4.2.15. Does sequence influence term relevance?
 - Are there 'better' terms at the beginning or end of a session?
- 4.2.16. How does presence /absence of label copy influence supplied terms?
- 4.2.17. How does 'speed' influence terms?
- 4.2.18. Are the same terms applied to multiple images of the same object?

- BW vs colour of the same work?
- Do multiple images produce better terms?

4.2.19. 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?

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

4.3.1. Search them within the tool?

4.3.2. 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?
- Used to expand or develop those vocabularies
- Could we "Hierarchialize" terms that are otherwise just lists?
- What about other local lists of terms?

4.3.3. 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'?

4.3.4. Percentile review of terms assigned to the same work

- What are the thresholds that define legitimacy?
- What are the ways to eliminate use-less terms?

4.3.5. How can we 'weight' terms?

- We may reach saturation in new terms BUT what about which ones are *more* appropriate than others? Should we keep accepting terms so that we can learn which are used most?

4.3.6. How do we compare terms assigned by different users?

- Professional vs amateur
- First-time user vs long-term user
- Specialist vs general public
- Users of different ages?

4.3.7. Can we segment terms by characteristics + use this to cluster context-specific or context-appropriate terminology?

- Possible clusters could be by age, or object type, for example

4.3.8. How does an individual's work change?

- By time
- By # of terms
- By accuracy
- By session
- By weighting
- In relation to defined thresholds

4.3.9. How often is a work ignored?

4.3.10. How do we group terms by type of user so that they can be processed in a different manner?

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

4.4.1. Is it any different than all of the other terms we have?

4.4.2. Do we segregate them from other terms?

4.4.3. 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?

5. People / Community = medium term questions about implementation and sustainability

5.1. How do you build community to do this folksonomic work?

- What constitutes community?
- What external rewards (outside the system but between the volunteer and the museum) could provide motivation? Are necessary for motivation?
- What kinds of rewards work, for whom?
 - Are complex interface functions motivating for 'power users'?
 - Numerical scores / points / community ratings [per session, per time period, per peer group]
 - "product" reward (image, poster, discount)

5.2. What is Stickiness?

- Associated with the tool
- Associated with the museum
- Associated with the activity
- When does this leave the on-line environment and become more of a directed museum experience?
- What contexts are the tool used in?

- If stickiness is a goal, what is the user 'sticking' to? (e.g. image sets that draw from across multiple institutions)

5.3. How does this relate to other parts of the museum?

- What do museum educators think of the environment and the user engagement it might foster?
- Where does the shop fit in?

5.4. How does use differ in multi-institutional context vs single institution environment?

- How do you share specialist users?
- What are the technical infrastructure implications?
- What are the social/Political implications?
- Is largest possible collection most attractive (satisfying) for the user?
- Is better cataloguing created by users drawn from largest possible group of potential users?
- What are the trade-offs for a museum in a multi-institutional environment
 - Loose direct contact with users?
 - Get better data?
- Does creating sets create better cataloging, and therefore is better cataloging more possible within a multi-institutional environment?
- Could multi-institutional context provide access to broader user-base with different cultural intelligence?

5.5. Who are the searchers who use the terms supplied?

- Users of web site – do they find things that they want?
- Demographics

5.6. Who are the likely groups of cataloguers?

- Focus groups of existing users? Why are they doing it?
- Demographics
- Contexts of use
- Ways to enhance experience that are context-driven

5.7. Do people need to edit the terms that they submit?

- Feel like they have control

5.8. Identity – What does it mean to be known to the system as an individual user? What does being able to catalogue anonymously do to results?

- What kinds of characteristics need to be recorded about users in order to support individualized features and data analysis?
- Is anonymity a value [requirement] for some group(s) of participants?
- Do anonymous users need to have full control of all functions in the system?
 - How much of this is required (anon users are differentiated)?

- How do anonymous users effect the ability to weight users' contributions?
- Does data analysis requires some level of user identification?

5.9. Will using the same system with different collections with differing levels of community involvement produce different results?

5.10. Can you achieve an international audience through multi-lingual access points?

6. subsequent questions

6.1. about community development – how do you create community among steve users?

- 6.1.1. can you share a tag stream
- 6.1.2. can you share the things you've tagged with others
- 6.1.3. can you subscribe to others tags (find all things other people say are 'happy')

6.2. how does steve related to the social bookmarking sites, where the site is neutral and the reward is given to the user and the user's group?