Public and Professional Vocabularies: Comparing user tagging with museum documents and documentation

J. Trant + D. Bearman
Archives & Museum Informatics

research question

Can social tagging and folksonomy improve online access to art collections?
J. Trant + D. Bearman

A MODEL OF TAGGING

Research Questions

- What kinds of terms are assigned?
- Who assigns and how many terms?
- How does users' tagging differ?
- Are terms applicable to works of art?

Vocabulary Sources

- Art and Architecture Thesaurus (AAT)
- Union List of Artists Names (ULAN)
- Searches of Art Image Databases
- Museum Records

Research Process

Friday, September 19, 2008
Research Questions

Who tagged What?

- users who assigns how many terms?
- folksonomy what kinds of terms are assigned?
- vocabulary sources are terms found in...
  - Art and Architecture Thesaurus (AAT)
  - Union List of Artist Names (ULAN)
- works of art
- museum records
- searches of art image databases

What was tagged?

J. Trant, February 2008
based on Triant 2006
Multi-Institutional Tagger: Works by Institution

1,784 works of art

- The Metropolitan Museum of Art: 14%
- Tate Modern: 3%
- Skirball: 9%
- San Francisco Museum of Modern Art: 9%
- Rubin: 6%
- Minneapolis Institute of Arts: 14%
- Boston Museum of Fine Arts: 13%
- Cleveland Museum of Art: 7%
- Denver Art Museum: 1%
- Indianapolis Museum of Art: 13%
- Los Angeles County Museum of Art: 11%

Research Questions: Tagging
Characterizing Tagging: Works Tagged

Friday, September 19, 2008
Research Questions: Tagging

Characterizing Tagging

Research Questions: Tagging

Characterizing Tagging: Basic Descriptive Statistics

Multi-Institution Tagger

<table>
<thead>
<tr>
<th># of works</th>
<th># of tags</th>
<th># of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>distinct</td>
<td>total</td>
</tr>
<tr>
<td>1,784</td>
<td>36,981</td>
<td>1,621</td>
</tr>
</tbody>
</table>

Single-Institution Tagger

<table>
<thead>
<tr>
<th># of works</th>
<th># of tags</th>
<th># of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>total</td>
<td>distinct</td>
<td>total</td>
</tr>
<tr>
<td>251</td>
<td>56,399</td>
<td>690</td>
</tr>
</tbody>
</table>

Vocabulary sources

Folksonomy

Users

What kinds of terms are assigned?

Are terms found in...

Who assigns how many terms?

Works of art

Research questions:

Social tagging and folksonomy analysis

Terms assigned 1

Terms assigned 2

Museum records

Art and Architecture Thesaurus (AAT)

Union List of Artists Names (ULAN)

Searches of art image databases

How does users’ tagging differ?

J. Trant, February 2008 based on Trant 2006

Are terms applicable to...
Characterizing Tagging: Basic Descriptive Statistics

- Total tags
- Distinct tags

Multi-institution: 22.5
Single institution: 67.5

Characterizing Tagging: Compare Registered | Unregistered

Multi-Institutional Tagger: Tags Per User: Compare Registered / Unregistered
- 2,017 users; 1,621 Users Tagged; 36,981 terms;
- 530 Registered Users; 15,615 terms; 396 Users with 0 tags not shown

Number of Users

Number of Terms

Friday, September 19, 2008
**Multi-Institutional Tagger: Sessions per Registered User**

- 1 session: 68%
- 2 sessions: 20%
- 3 sessions: 6%
- 4 or more sessions: 6%

**Characterizing Tagging: Sessions per Registered User**

---

**Multi-Institutional Tagger: Tags per Work, Average by Object Type**

- Number of tags per object type divided by number of works of that object type.
- Highlighted values for samples of fewer than 10 works.

**Characterizing Tagging: Tags Per Work by Object Type**
Research Questions: Tagging

Characterizing Tagging: Vocabulary Stabilization?

Multi-institutional Tagger: Terms by Novelty to Work

<table>
<thead>
<tr>
<th>Novel</th>
<th>Not Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>33,032</td>
<td>5,949</td>
</tr>
</tbody>
</table>

Are tags new terms?
Do user tags differ from museum documentation?

Research Questions: Museum Documentation

Museum Documentation: Catalogue Records On-line

Friday, September 19, 2008
Museum Documentation: Catalogue Records Online

<table>
<thead>
<tr>
<th>Museum Data</th>
<th>Definition</th>
<th>Req.</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution</td>
<td>The display version of your museum's name</td>
<td></td>
<td>The Metropolitan Museum of Art</td>
</tr>
<tr>
<td>Image filename</td>
<td>The name of the image file that matches this object</td>
<td></td>
<td>apps854.jpg</td>
</tr>
<tr>
<td>Institution ObjectID</td>
<td>An identifying number assigned to the object in your institution, other than an accession number (e.g. the record number from your collections management system)</td>
<td></td>
<td>11479</td>
</tr>
<tr>
<td>Title/Object Name</td>
<td>The display form of the work's preferred Title or Name</td>
<td></td>
<td>The Gulf Stream</td>
</tr>
<tr>
<td>Date</td>
<td>The display form of the work's date</td>
<td></td>
<td>1899</td>
</tr>
<tr>
<td>Maker/Culture</td>
<td>The display form of the creator's [or creators'] name[s] (and dates if desired)</td>
<td></td>
<td>Winslow Homer, 1836–1910</td>
</tr>
<tr>
<td>Medium</td>
<td>The display form of the material, support and/or techniques used to make the work</td>
<td></td>
<td>oil on canvas</td>
</tr>
<tr>
<td>Dimensions</td>
<td>The display form of the work's dimensions</td>
<td></td>
<td>28 1/8 x 49 1/8 in. (71.4 x 124.8 cm)</td>
</tr>
<tr>
<td>Object Type</td>
<td>The type of work, e.g. painting, sculpture, installation</td>
<td></td>
<td>painting</td>
</tr>
<tr>
<td>Credit Line</td>
<td>The display form of a credit to be shown with the work of art</td>
<td></td>
<td>Catharine Lorillard Wolfe Collection, Wolfe Fund, 1906</td>
</tr>
<tr>
<td>Copyright</td>
<td>The display form of a copyright statement for the work of art</td>
<td></td>
<td>© Jasper Johns/Licensed by VAGA, New York, NY</td>
</tr>
<tr>
<td>Accession Number</td>
<td>The accession number assigned to the work of art in your museum</td>
<td></td>
<td>06.1234</td>
</tr>
<tr>
<td>Notes</td>
<td>Curatorial or other notes about the work</td>
<td></td>
<td>The Gulf Stream was based upon studies made during Homer's two winter trips to the Bahamas in 1884–85 and 1898–99. First exhibited at the Pennsylvania Academy of the Fine Arts in Philadelphia in 1900, the picture was subsequently reworked and &quot;improved&quot; by the artist. Early photographs show changes to the sea and to the back of the ship, making the composition more dramatic and vivid. The painting was shown in the Carnegie Institute in Pittsburgh in 1900–1901, and then at M. Knoedler and Co. in New York, where the artist placed on the picture the record-asking price of $4,000. There were problems selling the work because of either its high price or its unpleasant subject matter. Homer may have reworked the painting again in the face of this criticism in order to add the rigger on the horizon that signals hope and rescue from the perils of the sea.</td>
</tr>
</tbody>
</table>

Research Questions: Museum Documentation

1. Do user tags differ from museum documentation?

Museum Documentation: Catalogue Records On-line

Friday, September 19, 2008

http://www.archimuse.com

Page 10
Research Questions: Museum Documentation

Do user tags differ from museum documentation?

Multi-Institutional Tagger: Tags and Museum Documentation
{36,981 terms: 35,307 unmatched; 1,674 matched}

Unmatched 96%
Matched 14%

Multi-Institutional Tagger: Museum Documentation
{1,605 Full Matches by Field}

Primary Title 24.8%
Object Type 44.9%
Accession Number 0.1%
Copyright 0.6%
Creation Date 1.5%
Creator 7.1%
Materials 21.1%
Multi-Institutional Tagger: Tags to Museum Documentation: "Full Word Partial Field" Matches by Data Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright</td>
<td>0.4%</td>
</tr>
<tr>
<td>Creator</td>
<td>7.6%</td>
</tr>
<tr>
<td>Credit Line</td>
<td>1.4%</td>
</tr>
<tr>
<td>Dimensions</td>
<td>0.9%</td>
</tr>
<tr>
<td>Materials</td>
<td>15.0%</td>
</tr>
<tr>
<td>Primary Title</td>
<td>41.1%</td>
</tr>
<tr>
<td>Object Type</td>
<td>3.9%</td>
</tr>
<tr>
<td>Object Metadata</td>
<td>29.3%</td>
</tr>
<tr>
<td>User Tags</td>
<td>0.1%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

Do user tags differ from museum documentation?

Friday, September 19, 2008
Multi-institutional Tagger: Terms Not Found in Museum Documentation

Do user tags differ from museum documentation?

Research Questions: Extended Museum Documentation

Are other forms of museum documentation more useful?

Genres
- Gallery label text
- Docent tour talking points / tour scripts
- Special exhibition catalogue entries
- Permanent collection guidebook text
- Curatorial analysis or notes
- Scholarly article
- Web site ‘feature’ texts
- Audio tours or podcasts
- Newspaper/popular article
RESEARCH QUESTIONS: EXTENDED MUSEUM DOCUMENTATION

I. Text mining museum extended documentation for the sorts of terms we get in tagging cannot be algorithmic.

II. Extended documentation is not better than online museum documentation.

III. Public genres of documentation are not meaningfully better than scholarly and internal for tag matches.

IV. False results from Inapplicable matches if extended documentation was searched for tags.

Do Museum Professionals find tags useful?
Research Questions: Usefulness

Do Museum Professionals Find Tagging Terminology Useful?

Steve Term Review

If you found this work, using this term in a query, would you be surprised?
Research Questions: Usefulness

Sample Works: Useful Terms

What kinds of words are tags?
ARE TAGS IN MUSEUM VOCABULARY SOURCES?

J. Trant + D. Bearman

NKOS September 2008

Research Questions: AAT/ULAN

Multi-institutional Tagger: Terms Found in AAT

[36,981 Terms; 25,978 Matched Some Field in AAT]

Matched 70.2%

Unmatched 29.8%
Research Questions: AAT/ULAN

Are tags in Museum Vocabulary Sources?

Multi-institutional Tagger: Terms Found in AAT: Distinct Terms
(11,944 distinct terms: 4,448 Matched Some Field in the AAT)

Matched 37.2%

Unmatched 62.8%

Multi-institutional Tagger: Full Matches to AAT by Hierarchy
Research Questions: AAT/ULAN

Are tags in Museum Vocabulary Sources?

Multi-Institutional Tagger: Tags matched to ULAN
[36,831 Tags: 598 Full Matches; 1,243 Partially Match ULAN]

Full Matches
Partial Matches
Unmatched

Are tags in Museum Vocabulary Sources?

ULAN Name: Green Terms are Unambiguous

Multi-Institutional Tagger: Full Matches to ULAN
[Top 502 of 698 Terms – assigned 3 or more times
only eleven (11) of 33 distinct terms are unambiguous artists’ names]
Are tags useful for searching?

Research Questions: Searching

Could tagging improve on-line searching for works of art?

Vocabulary sources
- folksonomy
- users
- what kinds of terms are assigned?
- how does users’ tagging differ?
- Art and Architecture Thesaurus (AAT)
- Union List of Artist Names (ULAN)
- searches of art image databases
- works of art

Terms assigned 1
- museum records

Terms assigned 2
- works of art
Could tagging improve online searching for works of art?

Research Questions: Searching

- New Tags
- Tags could improve search
- Search Terms
- Museum Documentation
- Tags
- Redundant Tags
- Museum Documentation not used in search
- Successful searches
- Unsuccessful searches

Summary

85%+ tags are not found in museum documentation
60%+ don’t match vocabularies
Most tags can’t be mined from other sources
Public Vocabulary is Different. Tagging does Contribute.

Thank you.

jennifer: jtrant@archimuse.com
David: dbear@archimuse.com