

For ICS/LIS seminar, Wed. 20.1.2010

Utilizing Folksonomies to Increase Accessibility to Media Collections

Joan C. Nordbotten

Jan. 20, 2010

Abstract

Currently, information retrieval from digital media collections relies heavily on text annotations. Libraries and museums have invested substantial sums of time and finances in annotating and cataloging their material so that it is retrievable. Unfortunately, the taxonomies used for cataloging do not necessarily include the search terms used by the general public. This creates a language gap that can hinder location of relevant and interesting material for the Internet user.

Folksonomies (Thomas Vander Wal, 2004), which consist of terms used and/or assigned by the general public while visiting a web-site, are currently being advocated to address the language gap between information seekers and professionally constructed object annotations. However, there are still numerous questions concerning utilization of Folksonomies. Just what is a folksonomy? How does it relate to library/museum taxonomies? What is the quality of user tags? Should a folksonomy be edited? How? Will the public 'really' participate? Can a library/museum accept these non-expert annotations? ...

This talk explores various aspects of these issues with the intention of setting a framework for debate.

Outline:

- Digital image collections
 - Image annotations
 - Text based image retrieval – TBIR
 - The language-gap problem
 - Folksonomies
 - Definition & Content
 - Capture & maintenance
 - Comparison to library metadata
 - Are folksonomies useful?
- examples*
[Library of Congress](#)
[The commons on Flickr](#)
Steve.museum

A few Digital Image collections

<i>Source</i>	<i>Owner</i>	<i>Approximate Size (Sept '09)</i>
Library	Library of Congress	1 000 000 images
	NY Public Library	700 000
	UB.UiB	20 000
Museum	Smithsonian Inst	246 000
	Metropolitan Museum Of Art	140 000
	Powerhouse Museum	74 000
Archive	Swedish National Heritage	70 000
		<u>2 250 000</u> *

* All manually annotated !

**Library of Congress: Image annotation**

Title: Fantoft Church, Bergen, Norway ca. 1890 and ca. 1900.
Subjects: Norway--Bergen.
Format: Photochrom prints--Color--1890-1900.
Rights: No known restrictions on publication.
Repository: Library of Congress, ... hdl.loc.gov/loc.pnp/pp.print
Part Of: Landscape and marine views of Norway (DLC)
URL: hdl.loc.gov/loc.pnp/ppmsc.06109
Call #: LOT 13432, no. 007 [item] ...
Notes: Title from the Detroit Publishing Co., ..., 1905.

Notes:

1. Annotation 'form' is a standard based on the Library's archival need.
2. The lack of semantic descriptive terms may hinder search form the general public.

Typical collection search and the language gap

Text based *image* retrieval – TBIR \approx Google search

Requires 'match' between query and annotation terms.

Ex-1: LOC search for "Norwegian stave church"

\Rightarrow No images were found for the search. (Jan.2010) ☹

\Rightarrow *Language gap* between *user query terms*
and *library annotations*

Note:

TBIR is 'universally' used as the image access type for photos and scanned museum artifacts, as it supports **semantic** search, i.e. search using terms describing image content.

Access using a collection folksomony

Ex-2: [Flickr.commons](#) search for "Norwegian stave church"
 \Rightarrow 1-3 of 475 (Sept. '09)



\Rightarrow *No language gap* 😊

Notes:

- Flickr uses source (library, museum, private owner) annotation + tags
- There are only 2 stave church images in Flickr.commons, both from LOC
- There are currently (Jan.2010) 512 images of Norwegian stave churches on Flickr.com

Flickr Folksonomy

Folksonomy: a collection of terms/tags assigned by the public to artifacts, such as: images, books, music, web-sites, etc.



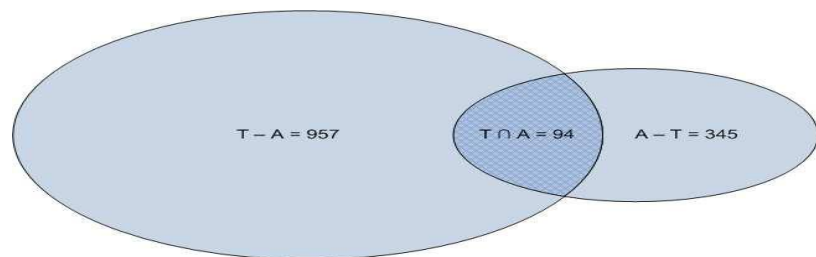
Example: Tags from [Flickr.commons](https://www.flickr.com/photos/14811470@N00/5214811470/)

Tags

- Library of Congress
- Bergen
- Norway
- church
- Fantoft
- stavkirke
- stave church
- photochrom
- color
- colorized
- Hordaland
- Fantoft stavkirke
- Scandinavia
- scenic
- Norwegian architecture
- haze
- hazy
- sky

Annotation vs folksonomy study, 2009. Eirik Bråthens

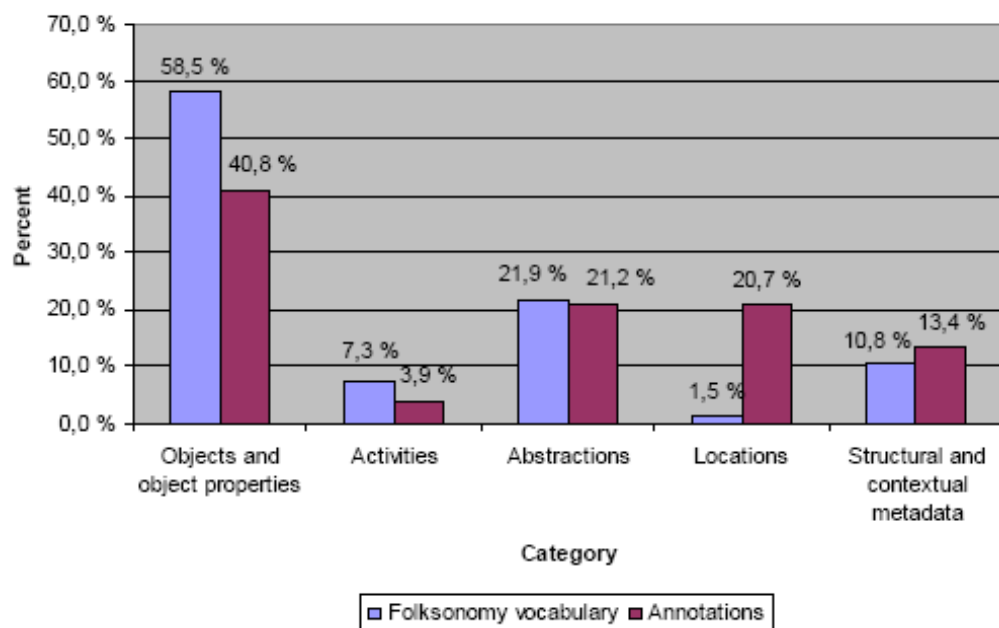
<i>Experiment size</i>	<i># terms (tags) comment</i>	
20 images	439	Univ. library annotations
20 UiB students	1 051	Student taggers
Descriptor overlap	94	91% new terms
Total search terms	1 396	68% new terms



Notes:

- Trant studies showed from 88 to 95% new terms from tags for museum images.
- An MMOA study of 30 images tagged by 35 administrators gave 88% new tags.

Annotation and Tag analysis, Bråthen, 2009



An Analysis of Image Folksonomy Generation, Jan-Erik Bråthen, 2009. Masters thesis.
http://caim.uib.no/publications/jeb_masterThesis.pdf

Collecting tags: **Steve museum tagger** <http://tagger.steve.museum>

LOG OFF THE ART MUSEUM SOCIAL TAGGING PROJECT


SET 2 213 REMAINING

Start Here

ADD ↓

bird by shore
Islamic art
bird

NEXT →



Painter: Abu'l Hasan; Calligrapher: Mir 'Ali
55.121.10.15 recto: Spotted Forktail: Album Leaf, 55.121.10.15 recto: ca. 1610-15
Ink, colors, and gold on paper; H: 15 3/16 in. (38.6 cm) W: 10 3/8 in. (26.3 cm)
Purchase, Rogers Fund and The Kevorkian Foundation Gift, 1955, 55.121.10.15
The Metropolitan Museum of Art

Tags for this work:
Abu'l Hasan, Persian miniature, bird, bird by shore, border, feathers, floral,
flowers, frame, gilt, illustrations, islamic art, long tail, miniature, ornate
border, rock, yellow

Tag collection alternatives

Alternative interfaces:

1. Image only
2. Image + current tags
3. Image + current tags + annotation

Alternative tag specification – accept:

1. Only single term tags
2. Only 1 version of each tag
3. Multiple term tags

Tag management:

1. Count/weight multiple identical tags
2. Edit terms

Combining annotations and tags – ex.

<i>Term</i>	<i>Def. Frequency</i>			<i>Usage frequency</i>
	Annotation	Tag	Sum	
Bergen	2	1	3	400+
Church	1	2	3	100+
Fantoft	1	2	3	90
Norway	3	1	4	150
Norwegian	0	1	1	50
Sky	0	1	1	0
Stave	0	1	1	100+
Stavkirke	0	2	2	50
...				

Folksonomy/ tag management**Proposals**

1. Keep singular/plural forms and compound terms
2. Term weight \approx # of replicated definitions and use
3. Filter tag specifications
4. Delete unused and/or unwanted tags
5. Construct a semantic hierarchy
6. ...

Is tagging useful? *Can be!*

- Tagging is FREE
- Done by the public.
- Indexing + maintenance
- can be automated.
- Increase access AND
- Save time and \$ ☺

References:

- Bråthen, J-E. (2009). An Analysis of Image Folksonomy Generation. Masters thesis, University of Bergen. http://caim.uib.no/publications/jeb_masterThesis.pdf
- Trant, J. (2009, last viewed 25.8.09) Tagging, Folksonomy and Art Museums: Early Experiments and Ongoing Research. Journal of Digital Information 10(1), p1-40. Published at <http://journals.tdl.org/jodi/article/view/270/277>