Faceted navigation of social tagging applications

Louise Spiteri

School of Information Management

Dalhousie University

DALHOUSIE UNIVERSITY
Inspiring Minds

Pervasiveness of social software applications

- Blogs: Wordpress, Tumblr, etc.
- Clipping services: Diigo, Licorize, etc.
- Social bookmarking: Delicious, Digg, etc.
- Social citation: CiteUlike, Zotero, etc.
- Social cataloguing
  - Books: LibraryThing, Goodreads, etc.
  - Films: Flixster, Flickchart, etc.
  - Music: Last.fm, Discogs, etc.

Zeitgeist (with thanks to LibraryThing)

CiteULike: 5,320,389 members; 1800 articles added daily.

LibraryThing: 1,364,328 members; 63, 763,679 book catalogued; 77,191,852 user tags

Wordpress: 365,389 blogs; 359,110 new posts a day

Delicious: Over 200 million bookmarked URLs.

Searching tags

- In most social tagging applications, tags may be searched in two ways
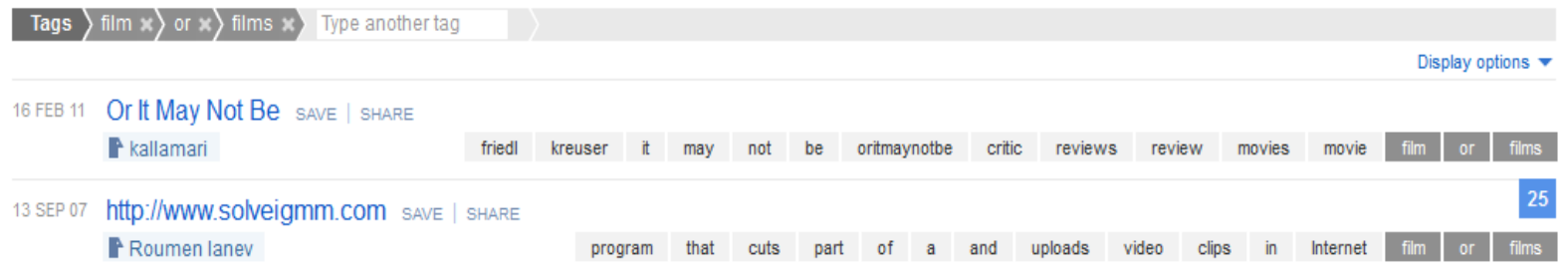 - Keyword search
  - Tag clouds

Keyword search

- The uncontrolled nature of tags poses difficulties for keyword searching; one must account for such variables as
  - Spelling variants: cataloging/cataloguing
  - Singular and plural variants: movie/movies
  - Synonyms: films/movies
  - Full versus abbreviated terms: CFL/Canadian Football League
  - Polysemes: Jaguars (cars)/Jaguars (cats)

ISKO-UK July 2011

Keyword searching

- In a keyword simple search for bookmarks related to reviews for film noir movies, for example, I may need to search for such tags as:
  - film, films, movie, movies, review, reviews, and film noir.

- Searching for compound tags could be problematic, since if a system does not allow for such tags, you need to consider variants such as filmnoir, film_noir, etc.

Boolean search (or not)

- ○ Social tagging systems may not allow for Boolean searches. A search for "film or films", for example, retrieves the following:

Boolean search 2

- If I try searching for variant terms, I get sites that have ALL the tags as variants:

Tag clouds

- Tag clouds are visual representations of social tags, displayed usually in alphabetical order, where the font size for each tag corresponds to the relative frequency of its use.

Limitation of tag clouds

- A few popular topics and their related tags tend to dominate the whole cloud.

- Related tags may lie far apart, and so meaningful associations may be missed

- You cannot search more than one tag at a time, so I need to determine which tag (assuming it is in the cloud) will be of most use to me in a search.

Facets and tags

- In social tagging applications, rather than have flat, undifferentiated lists of tags, it would be more useful to organize these tags into categories (and possibly sub-categories) to allow for more structured browsing.

- In this way, rather than access resources via individual tags, one could browse groups of terms associated with any one facet, or a combination of facets

Example of facets and social tags

Cooking (domain)

Dish Type (Main, Soup, Salad, Side, Dessert),

Ingredient Type (Meat, Vegetables, Grains, Spices)

Cooking Method (Bake, Fry, Grill, Easy)

Cuisine Type (Italian, Indian, French)

Facets already in place

- Facets have been used successfully to facilitate browsing in a number of online systems (e.g., Flamenco, Wine.Com, L. L. Bean).These systems are typically controlled by the site owners, who create and control the choice of facets and how they are populated.

- The tagging process in social tagging applications tends to be autonomous in nature: While I may choose to use other members' tags, I may decide equally to assign my own tags to a resource.

Facets in Wine.com

Shop for Wine ▶

90+ rated under $20	+
Best Wine Deals	+
Fine Wine	+
Type/Varietal	+
Style of Wine	
Region	
Rating Source	+

Shop for Wine by Type

Red Wine	**White Wine**	**Champagne & Sparkling**
Cabernet Sauvignon	Chardonnay	**Rosé**
Syrah / Shiraz	Sauvignon Blanc	**Saké**
Pinot Noir	Riesling	**Dessert & Port**
Bordeaux Blends	Pinot Gris / Pinot Grigio	
Merlot	Bordeaux Blends	
View All	View All	

Goals of the study

To what extent can facets be used to facilitate the organization, display, and retrieval of tags in social tagging applications?

Conduct an analysis of existing and proposed methodologies for the use of facets in social tagging applications, with particular emphasis placed on the extent to which these methodologies address the following questions: (next slide)

Research questions

Choice of facets
- Given the vast scope of the subjects covered in a typical social tagging application, how do you choose facets that can apply generally to all these subjects or domains?

How many facets?
- Given the potentially vast scope of the subjects or domains covered in most social tagging applications, how many facets would suffice to cover site members' needs?

ISKO-UK July 2011

Research questions 2.

Who
- Should the site members choose the facets? The site editors? Researchers?

Quality
- How do you ensure and maintain facets chosen reflect the needs of the site members?

Definition of facets

- Lack of consistency across the 17 studies regarding what constitutes a facet.

- "Classic" definition of facets: "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject." (Taylor, 2000).

Definitions of facets 2

"a set of meaningful labels organized in such a way as to reflect the concepts relevant to a domain.

"a set of attributes expressed as words or phrases."

"orthogonal descriptors (i.e. categories) within a metadata system ... each facet has a name and it addresses a different conceptual dimension or system."

"a set of headings in which the assignment of one heading to a resource limits the assignment to that resource of other headings in the set."

Definitions of facets 4

- Not much emphasis on the mutual exclusivity of facets: If facets are to serve as efficient tools by which to organize tags and provide them with meaningful context, it is important that each facet be defined clearly, represent only attribute, or characteristic of division, and not overlap with any other facet.

- If site members cannot distinguish between the meaning and scope of the facets *subject* and *description*, for example, it is very likely that related tags will be distributed between the two facets.

Research question 1: Choice of facets

- Only 5 of the 17 studies addressed how facets are to be derived.

  - 3 of these studies derived their own sets of facets, but do not explain clearly how this was done, whether their choice of facets was tested by other parties, and how reflective these facets are of the subject domain or clients of interest.

Proposed Flickr facets

- Beaudoin's Flickr facets: adjective, compound, emotion, event, humor, language, living thing, number, person, photographic, place-general, place-name, poetic, rating, thing, time, unknown, and verb.
  - No explanation of how these facets were chosen.

- These facets are unusual in that they represent not only conceptual attributes (e.g., person, thing), which is what one typically associates with facet analysis, but also grammatical constructs (e.g., verb, adjective, compound) and affect (e.g., poetic, humor, rating).

7W approach for Dept. of Defense

- U.S. Department of Defense suggested facets derived in response to 7W approach:

1. *Who* uses the service?
2. *What* does the service do?
3. *On what* does the service act?
4. *To whom* is the service generally directed?
5. *Where* is the service used?
6. *When* is the service used?
7. *Why* is the service used?

FaceTag proposed facets

CRG Facet	FaceTag Facet
Type	Resource type
Property	Language
Material	Themes
Product	Deliverables
Patient	Purposes, Marketing
Agent	People
Time	Date

Proposed music facets

AllMusic:

genres, audio attributes
(prominent instrument),
geographic, and opinions
(subjective data including
mood and themes)

Last.fm:

personal meaning, extra-
musical associations,
superlative expressions,
descriptors, charts, genres,
geographic, audio
Attributes, people, song ID,
time.

Proposed book facets

- Weaver derived the following facets that could be used to organize fiction-related tags: character, plot, subject, setting, and genre.

- Weaver suggests that the catalogue could provide an input form to ask clients to apply tags for a fiction-based information resource to this set of facets

- Weaver's study has promise for the integration of structured tagging in public library catalogues, but unfortunately, the methodology he used in the tagging exercise and by which he derived his list of facets is not explained clearly.

Research question 2: How many facets?

- The studies do not agree on the optimal number of facets to use in social tagging applications.

- Ranganathan's original five categories of *Personality*, *Matter*, *Energy*, *Space*, *Time*, are likely too restrictive for the multidisciplinary nature of social tagging sites, not to mention the ambiguous nature of the Personality category.

How many facets? 2

- The Classification Research Group's expanded list of categories: thing, kind, part, property, material, process, operation, agent, patient, product, by-product, space and time.

- These facets may be a good starting place, since they were devised to represent attributes that are basic to most subject domains but is it reasonable to expect site members to choose from amongst 13 facets in which to place their tags?

Research question 3: Who chooses the facets?

- No agreement found on who should actually derive the facets. In some cases, the facets were derived by the researchers, while some of the studies suggest that site editors be responsible for this function.

- Most of the studies make no tangible recommendations for who should derive the facets used in a social tagging application. The diverse nature of the site members in such applications means that it may be particularly difficult to reach consensus.

Who chooses? 2

- The initial choice of facets may require some expertise in the design of taxonomies, thesauri, or classification systems, as you want to ensure that the facets chosen are homogenous, mutually-exclusive, and representative of the tags used in the system. A set of site editors may also work in this capacity, but they would need to have some experience,

- The use of site members could be challenging, since it would be very hard to control for variables that would naturally exist among them when it comes to subject expertise, knowledge of taxonomic structures, and so forth.

Research question 4: Validity & maintenance of the facets

- None of the studies addresses the need to evaluate the reliability and validity of the chosen facets.

- In applying facets to a social tagging application, a number of areas need to be evaluated, including the choice of facets, the extent to which the facets reflect the needs of the site members, their ease of use, how much they are used, and whether they are used correctly (e.g., location-type tags are placed correctly in the location facet).

Maintenance 2

- None of the studies discusses how the use of facets is to be monitored. The democratic nature of most social tagging applications means that you cannot force site members to use the facets.

- Regardless of how clear the facets may be in their scope, it is always possible, and probably likely, that some site members will place tags in the incorrect facets. How will the use of the facets be monitored? If facets are not used correctly, then their potential benefits may not be realized fully.

Conclusions: Adopt a common definition

- The Taylor definition that facets are "clearly defined, mutually exclusive, and collectively exhaustive aspects, properties or characteristics of a class or specific subject" would work, as it emphasizes the importance of creating mutually-exclusive facets.

- Given the potentially varied content and scope of social tagging applications, however, it is questionable whether the "collectively exhaustive" part of this definition would be very applicable.

Conclusions: Modified 7W approach

- Combining and modify the 7W questions approach with the application of the 13 CRG facets.

- The combined approach could consist of crafting a simple question to correspond to each of the 13 CRG facets, in line with the 7W model; each question could then be applied to a set of tags in order to determine in which facet to place that tag.

Modified 7W 2

- For the space facet, for example, one could ask the question "Where is this person, place, or item?"

- One would need to consider the optimal number of facets to include in a social tagging application, as well as the level of granularity may be so specific that end users may have trouble distinguishing between or among what may appear to be similar facets, such as process and operation, or patient and agent.

Modified 7W 3

- These questions could then be applied to a representative sample of tags to determine their placement in the appropriate facet; this process may help determine which of these facets are used most frequently, which are applicable, and which may not apply.

- Once a list of facets is chosen, the set of questions could be added to the user interface, so that when site members create or re-use a tag, the questions could help them determine in which facet to place the tag.

Questions?