

CHAPTER 3

Tagging and Folksonomies

A major challenge of dealing with digital content -- our own and others -- is organizing it. We want to be able to find the piece of content that we want, and see its relationship to the whole and to other digital content. We might want to be able to reuse this content. Also, importantly, we want other people to be able to understand the organization of your digital content so that they can find and reuse it.

Tags are one of the most popular mechanisms used in contemporary websites for letting users organize digital content. A tag is a label, typically a word or short phrase, that a user can add to a piece of digital content (a photo, a URL, a video, an email). You can then search for digital content with those tags. As we saw in Chapter 2, when tags are embedded in URLs, you can link and embed content related by tags through those URLs.

The term folksonomy was coined to contrast tags with taxonomies, which are formal schemes typically created by communities with strict practices of classifying items. There are few restrictions on the tags you come up with to associate with your content. There are no preset categories or controlled vocabularies from which you must choose. Tags have proliferated; users have taken to them en masse, generating collections -- or clouds -- of tags that help order their own content as well as content throughout the web. We can use these tags to relate content in our mashups, mindful, however, that tags can often be idiosyncratic, ambiguous, and irregular.

For now at least, tags have not led to the anarchy predicted by some taxonomists, and there is more order to how people tag than you might think, created by things like personal and social conventions and the syntax of tags. On the other hand, the proliferation of tagging has certainly not obviated the need for formal classification schemes. There are rich opportunities to bring together user-generated, bottom-up folksonomic tags and controlled vocabularies and taxa.

This chapter will show you how to connect content, to mash things up, with tagging as the glue. Tags allow the aggregation of resources within a system (say, pictures in Flickr -- your own and others) and across websites (Technorati).

This chapter:

- * Illustrates how tags are used in Flickr, del.icio.us, and Technorati
- * Shows how people are using tags to create interesting apps with tags
- * Discusses how people are hacking the tagging system to put more information into Flickr and other websites, specifically *geotagging*, and now, more generally, machine tags
- * Studies some issues around the interoperability of tags across systems, specifically through a study of Technorati.
- * Briefly shows how tagging relates to formal classification systems, using books as an example.

Tagging in Flickr

According to the Flickr FAQ¹: "Tags are like keywords or labels that you add to a photo to make it easier to find later." Tagging is a central way of tying words to pictures. (Think about how search works -- you type in words and phrases.) Tagging is very important for photos since computer vision/automatic scene recognition is in its infancy.

WILL WE HAVE VISUAL SEARCHING INSTEAD OF RELYING ON TAGS?

Note that companies like riya.com are hard at work to bring us visual search.² What might a non-word-based search look like? Draw something that you want to look for and the search engine will bring up pictures that look like what you drew? Or would you present a photo to the search engine, and it would bring up similar photos? The fact that we still have to type words in to a search engine to search for pictures or video or music, shows how dependant we are on words for search and for describing nonverbal objects. That's why tags are so central in Flickr, where the dominant form of data is visual. That's not to say that there aren't interesting experiments in non-textual search such as the "search by sketch" system retrievr (<http://labs.systemone.at/retrievr/>).

End Sidebar

Let's list some practical skills related to tags in Flickr that you will learn in this section:

- * We'll see how tags are used in the Flickr community -- by individuals, by subgroups -- right across Flickr to bind photos together. (It's useful to study tags before creating your own.)
- * We'll see how to tag a picture, and thereby run into issues when we sit down to tag our pictures or those of others.
- * We'll see how to deal with the syntax of tags in Flickr; how to use in multiword tags; and how multiword tags get boiled down to canonical tags.

Tags in Flickr

In Chapter 2, I present an overview of how tags are used in Flickr, specifically as they are manifest in the websites' URL Language. Here, we look deeper at Flickr tags, specifically to look at the social context of tags in Flickr, the syntax and semantics of tags in Flickr, hacks of Flickr tags, and some remixes and mashups that build upon the Flickr tags.

Before I jump to those topics, let me present parts of the URL language concerning tags. For instance, you can see a list of popular tags in Flickr at:

<http://www.flickr.com/photos/tags/>

The URL for the most recent photos in Flickr with associated with a **tag** is:

<http://www.flickr.com/photos/tags/{tag}/>

e.g.,

¹ <http://flickr.com/help/tags/#37>

² <http://riya.com> and <http://www.riya.com/riyaAPI> (for the Riya API)

<http://www.flickr.com/photos/tags/flower/>

Instead of sorting photos by the date uploaded, you can see sort them by descending "interestingness" (a quantitative measure calculated by Flickr of how "interesting" a photo is):

<http://www.flickr.com/photos/tags/{tag}/interesting/>

Finally, for some tags, Flickr identifies distinct clusters of photos, which you can access at:

<http://www.flickr.com/photos/tags/{tag}/clusters/>

e.g.,

<http://www.flickr.com/photos/tags/flower/clusters/>

You can display the popular tags used by a specific user at

<http://www.flickr.com/photos/{user-id}/tags/>

or list all the user's tags at:

<http://www.flickr.com/photos/{user-id}/alltags/>

You can show all photos with a given tag for a specific user at:

<http://www.flickr.com/photos/{user-id}/tags/{tag}/>

How are Tags Used in Practice

So, how do people actually use tags in Flickr? Look around to get a feel for how people have been tagging their photos. It is also helpful to draw upon the observations of seasoned Flickr users with respect to general trends for how tags are used -- or should be used.³

The issue of how tags are used is very complicated. To get a feel for the issues involved, let's look at how people tag photos for July 4. You can imagine a number of different ways of tagging, including the following, which are all used:

- * [july4](http://www.flickr.com/photos/tags/july4/) (<http://www.flickr.com/photos/tags/july4/>)
- * [fourthofjuly](http://www.flickr.com/photos/tags/fourthofjuly) (<http://www.flickr.com/photos/tags/fourthofjuly>)
- * [july4th](http://www.flickr.com/photos/tags/july4th) (<http://www.flickr.com/photos/tags/july4th>)
- * [july04](http://www.flickr.com/photos/tags/july04) (<http://www.flickr.com/photos/tags/july04>)
- * [july4th2007](http://www.flickr.com/photos/tags/july4th2007) (<http://www.flickr.com/photos/tags/july4th2007>)

As an end-user, which tag or tags should you use? It depends. Are you trying to use the most popular one? Flickr offers no guidance about which specific tag to use but attempts to make pictured related to July 4th all findable regardless of the exact tag used. The Flickr clustering algorithm, when applied to some of these specific tags (e.g., <http://www.flickr.com/photos/tags/july4th/clusters/>) groups together pictures with tags aimed at describing the same phenomenon.

³ <http://www.flickr.com/groups/central/discuss/2026/> and <http://www.flickr.com/groups/central/discuss/2730/>

It is significant that you can set a default permission that allows other people (which you can limit to your family or friends or contacts or any registered Flickr user in general) to add tags and notes to your photos -- but there is no provision for letting other people change the title or description of your photo. This suggests that it might be a good idea to let other people tag your photos. Think of scenarios when it would be helpful to let others tag your photos. Consider why it might not be a good idea to let other people change the title or description of a photo.

Creating Your Own Tags

To add a tag to a photo for which you have permission:

1. Go to the Flickr page of the photo.
2. Click on the "Add a tag" link. A textbox will open up and you can enter a single tag or a series of tags separated by spaces. You can also enter phrases by using quote marks around the phrase.
3. You can also choose to add tags by selecting from tags that you already use by hitting the "Choose from your tags" link instead of entering tags in the textbox.

Syntax of Tags in Flickr

The Flickr tagging system is sufficiently well designed that you may never have occasion to think about the syntactical limitations of tags in Flickr. However, let's look at a simple case study. As noted above, you can add phrases as tags, using quotes, e.g., "San Francisco". The tag is displayed as "San Francisco" but internally, it is represented with spaces and punctuation removed and letters turned to lower case -- that is, `sanfrancisco`. You can prove this by going to a picture and trying to enter "San Francisco" and `sanfrancisco` as tags. Flickr will take only one of the tags since it considers them to be the same tag.⁴

Now, why should we care about the exact syntax of a tag? One reason is that tag syntax is going to be different among systems. To understand this fact, it helps to understand at least one system, such as Flickr, and then to figure out the syntax of tagging for these other websites or applications. Also it gives you insight into one issue that will challenge all tagging system: figuring out which tags are the same and which are not.

Potential Weaknesses of Tags

Anyone who has spent much time using tags runs into the idiosyncrasies, inaccuracies, and irregularities often present in tagging. Drawing from an analysis in the Wikipedia, I list some possible causes for these problems:⁵

⁴ <http://www.flickr.com/services/api/misc.tags.html> draws the distinction between the "clean" version of a tag and the "raw" version of the tag.

⁵ <http://en.wikipedia.org/wiki/Folksonomy> as <http://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=145985651>

- * *polysemy*: since words often have multiple meanings, which meaning is supposed to be associated with a tag? (e.g., does a tag **apple** refer to the fruit or to a computer?)
- * *synonymy*: when multiple words can have the same or similar meaning, which tag should you use and how do you find all the tags that mean the same? (e.g., "Independence Day" in the U.S. and "July 4th")
- * *word inflections*: Since words are modified for specific grammatical contexts, which variation do you use for a tag? (e.g., **mouse** vs **mice**)
- * *syntactic constraints*: How should you create tags out phrases when spaces are not allowed? How should you deal with punctuation? How do you deal with non-ASCII words?

In this chapter, I take up the issue of word inflections, specifically the handling of single vs plural forms -- and the syntax of tags, a topic that is not explicitly mentioned, but which presents practical difficulties in making mashups based on tags.

Singular and Plural Forms of Tags in Flickr

Websites often leave ambiguous whether users should use the singular or plural form for tags. When you use these tags, it's helpful to know whether tags created with the single and plural forms are treated as the same tag.

Here I describe a small experiment to figure out how Flickr deals with this issue, one you can adapt for other websites. I tag one of my photos with the tag **mouse** and do a full-text search and a tag search for **mouse**, **mouses**, and **mice**. Table 3-1 records whether the photo is returned in the search.

Table 3-1. Stemming of terms related to "mouse" in Flickr

Search term	Full text or Tag Search	Picture Found
mouse	full text	yes
mouse	tag	yes
mouses	full text	yes
mouses	tag	no
mice	full text	yes
mice	tag	no

Based on these limited observations, I would make the following tentative conclusion about how Flickr handles singular and plural English nouns in tags:

- * Singular and plural forms of English nouns used are considered as different tags.
- * In full text searches, Flickr uses some form of stemming to match singular and plural forms of English nouns. The Flickr stemming process is at least sophisticated enough to recognize that **mouse** and **mice** are related words.

Obviously, we would have to either find official documentation from Flickr or test with many more tags to validate these conclusions.⁶ The point here is not to rigorously test these conclusions but to point out how simple experiments can sometimes reveal interesting aspects about a website such as Flickr.

Hacking the Tagging System: Geotagging and Machine Tags

The Flickr Map (<http://www.flickr.com/map/>), which displays Flickr photos on a map, is the official implementation of what started as a hack. Before the map, there was no official way to store the location information of a picture and display that location information on a map. The ad hoc solution that became widely adopted was to insert geo-related information into the Flickr tags, specifically, the **geotagged** tag along with **geo:lat** and **geo:lon** to indicate latitude and longitude of a photo.

This convention of geotagging worked well in many ways. Hundreds of thousands of Flickr photos were geotagged according to this convention. Tools such as the Google Maps in Flickr arose to use the geotagging data. On the downside, the Flickr user interface became cluttered with tags that were meant for programmatic consumption. There wasn't ideal support for such tags in the Flickr API (for instance, the only reason for the **geotagged** tag to be there was that the API did not allow you to look for tags that began with **geo:lat**).

It was to fix these problems that Flickr introduced machine tags, also known as triple tags. Machine tags are tags with a specific syntax aimed primarily for programmatic consumption and not directly for display to the typical end-user. You can use machine tags to store extra data elements for a given photo. The most important example of such data has so far been the latitude and longitude associated with a photo; so important that Flickr ultimately introduced specialized functionality to handle this data, to prevent people shoehorning it into tags.

Machine tags are meant to support new types of applications along the lines of geotagging by adding functionality to the API that recognizes that machine tags have a different use pattern than standard tags. Also, the UI of Flickr has changed to hide by default machine tags from users.

The syntax of machine tags, which relates the triplets of **namespace**, **predicate**, and **value**, is:

`namespace:predicate=value`

So, for example, **geo:lat=37.866276**, where **geo** is a namespace, **lat** is a predicate, and **37.866276** is a value.

Since machine tags are still in the early stages of uptake in Flickr, a pioneer in the field of letting people place arbitrary data into their systems, I would be surprised to find other web applications that are further along. There are some nascent developments along these fronts in Google Base (which has attributes)⁷ and Amazon S3 (with its item level metadata).⁸ In

⁶ The thread at <http://www.flickr.com/forums/bugs/31668/> includes a Flickr staff member's confirming the use of stemming in titles and descriptions. <http://tech.groups.yahoo.com/group/yws-flickr/message/1913> mentions stemming in the context of tags.

⁷ <http://base.google.com/support/bin/answer.py?answer=27882>

Chapter 6, I discuss how to use machine tags with the Flickr API. In Chapter 16, I return to the topic of Amazon S3.

Interesting Apps using Flickr Tags

A good way to understand how tags are used in Flickr is to study how others have built on top of the tagging system. Here are several to study:

- * Related tag browser (http://www.airtightinteractive.com/projects/related_tag_browser/app/) lets you browse relationships among related tags.
- * findr (<http://www.forestandthetrees.com/findr/findr.html>) lets you display related tags and photos that have been tagged by a combination of related tags.
- * fastr (<http://randomchaos.com/games/fastr/>) is a game in which you guess a tag based on from the photo presented to you.
- * zonetag (<http://zonetag.research.yahoo.com/>) is an example of Flickr tag hacking to insert location data of photos taken by cellphones.
- * tagmaps (<http://tagmaps.research.yahoo.com/>) shows on a map popular tags correlated with geotagged Flickr photos for a region

These two examples show how Flickr calculates relationships among tags by mining information about how tags are being used. You can get a sense of how people use tags.

Tagging in Del.icio.us

del.icio.us is a social-bookmarking application, the first of its kind, and in many ways, still the best. Users use deli.cio.us to keep track of bookmarks, identified by URLs, and to follow other users' bookmarks. Tagging is an important part of del.icio.us, which pioneered tagging in general and has done much to popularize it.

In the discussion of Flickr, I show how tagging enable textual searching and browsing of nonverbal objects such as pictures. Why would tags be useful in del.icio.us for categorizing webpages, whose primary constituent still tends to be text? Tags capture essentials about a webpage that cannot be easily uncovered by full text searching. Useful tags might not even involve any of the words that are actually in the text of the webpage. Tags often describe the relationship between the bookmark and the user (e.g., the tag [toread](#)) rather than anything intrinsic to the webpage itself. Nonetheless, we might get to the point in which computer summarization techniques could automatically generate tags for a given webpage. For instance: tagthe.net (<http://tagthe.net/>) provides such an API.

Note a fundamental difference between tagging in Flickr and del.icio.us. In Flickr, each object being tagged (a photo) has only one set of tags, created by the object's owner and others granted permission to tag the photo. In del.icio.us, each object (a bookmark) being tagged could belong to many users, eaching having their own sets of tags. As Thomas

⁸ <http://docs.amazonwebservices.com/AmazonS3/2006-03-01/BasicsObjects.html> and <http://docs.amazonwebservices.com/AmazonS3/2006-03-01/RESTObjectPUT.html>-- you can stick in user-metadata (name value pair).

Vander Wal explains, "broad" folksonomies, such as that of del.icio.us -- as opposed to the "narrow" folksonomies, such as Flickr's -- enable one to compare how different people tag the same object.⁹ For objects that are tagged by many people, del.icio.us is able to make recommendations of tags to use, based solely on how others have already tagged the object. In Flickr, you can't get such recommendations since there is only one set of tags for any photo.

Chapter 2 documents the URL Language of del.icio.us. In this chapter, I describe more about the mechanics of adding tags and the issues of multiple-word, and multilingual tags.

Mechanics of Adding Tags in del.icio.us

Without the del.icio.us Firefox plugin, you could use the websites upload form:

1. Go to <http://del.icio.us/post/> and enter the URL (e.g., <http://www.rubyonrails.org/>) and hit the **save** button.
2. You will end up on a page that prompts you for the description, notes, and tags. Note that del.icio.us offers recommended tags and lists your tags, which are tags that you have already used in del.icio.us -- if any.

With the del.icio.us Firefox plugin (<http://del.icio.us/help/firefox/extension>), it becomes easier to push a link into del.icio.us. You can also use a bookmarklets to put in pages (<http://del.icio.us/help/buttons>) -- or get IE buttons (<http://del.icio.us/help/ie/extension>).

Dealing with Cases and Multi-word Phrases

In contrast to Flickr, del.icio.us tags are single-word labels. Tags in del.icio.us cannot contain any spaces but they can contain punctuation. The example given in the documentation (<http://del.icio.us/help/tags>) is what to do with a multi-word phrase **San Francisco**, for which suggested tags are: **sf**, **san-francisco**, **SanFrancisco**, and **san.francisco**, "or whatever makes sense to you". Does it matter which of these tags you choose?

Let's gather some facts about how del.icio.us works with search phrases. (There's some documentation at <http://del.icio.us/help/search> -- we can also do a little experiment.) Let's look for "San Francisco" in del.icio.us. If you type "San Francisco" in the search box, selecting the option to search all of del.icio.us, you go to:

```
http://del.icio.us/search/?fr=del_icio_us&p=san+francisco&type=all
```

You can limit the domain of the search (your own bookmarks, all of del.icio.us, or the web). This search "goes through bookmark descriptions, notes, and tags." You can limit the search to tags via a **tag:** prefix (**tag:sanfrancisco**):

```
http://del.icio.us/search/?fr=del_icio_us&p=tag%3Asanfrancisco&type=user
```

What can we learn from this search?

⁹ http://www.personalinfocloud.com/2005/02/explaining_and_.html

- * The case of tags is preserved in how a tag is displayed (i.e., if you enter `SanFrancisco`, it will stay `SanFrancisco`). However, searches for tags are case-insensitive, i.e., if you search for `sanfrancisco` or `SanFrancisco`, you still get the same tags (<http://del.icio.us/tag/SanFrancisco>).
- * On the other hand, punctuation is significant in search as well as in the display. Unlike Flickr, in which punctuation is stripped from the canonical representation of a tag, punctuation does not behave like whitespace.

In del.icio.us because you can't have spaces in tags, there are many variations in dealing with multiword tags. Coming back to the example of `San Francisco` and the variants `sf`, `san-francisco`, `SanFrancisco`, and `san.francisco` -- and contrast the syntax of tags in del.icio.us and Flickr:

- * in del.icio.us: `San Francisco` is not a valid tag because it contains a space. `sf`, `san-francisco`, `SanFrancisco`, and `san.francisco` would be all distinct tags.
- * in Flickr: `San Francisco` is a permissible tag. However, you would not be able to tag the same photo also with any of the following variants (`san-francisco`, `SanFrancisco`, and `san.francisco`) because the punctuation is stripped away to determine the clean version of a tag.

Miscellaneous

<http://tech.groups.yahoo.com/group/ydn-del.icio.us/> is a good place to get answers to developer-type technical questions. You'll often see Joshua Schachter the founder of del.icio.us, actively answering people's questions.

REPRESENTATION OF LATIN-8 CHARACTERS AND UNICODE CHARACTERS MORE GENERALLY IN TAGS

Let's see how tags work for Latin-8 characters first (e.g., the French word français) and then Chinese.

In Flickr

Let's look at <http://flickr.com/photos/tags/fran%C3%A7ais/>. There is no collapsing of français to francais. See photo (<http://flickr.com/photos/raymondye/368644336/>) to see that I can have both a `français` and `francais` tag; invoking the API¹⁰ confirms that the two tags stay distinct.

It seems that Chinese works in a similar way. I don't know much Chinese, but I do know my name in Chinese (余俊雄). I managed to add it as a tag for one of my pictures.¹¹ You can pull up all pictures with that tag:

<http://flickr.com/photos/raymondye/tags/%E4%BD%99%E4%BF%8A%E9%9B%84/>

¹⁰

http://api.flickr.com/services/rest/?method=flickr.tags.getListPhoto&api_key=972aa8be46049e46dbaab3b16aeb3323&photo_id=368644336

¹¹ <http://flickr.com/photos/raymondye/79915850/> and <http://flickr.com/photos/raymondye/tags/%E4%BD%99%E4%BF%8A%E9%9B%84/>

Again, you can confirm that `%E4%BD%99%E4%BF%8A%E9%9B%84` is a URL-encoded UTF-8 representation of my Chinese name. With Python:

```
import urllib
print urllib.unquote('%E4%BD%99%E4%BF%8A%E9%9B%84').decode('utf-8')
u'\u4f59\u4fca\u96c4'
```

And you can see that the Unicode character point 4f59 is indeed 余.¹²

In del.icio.us

I added a URL for the France-Berkeley program¹³.

<http://del.icio.us/tag/fran%C3%A7ais> does come up with many links with the tag `français`, as well as the corresponding full-text search for français.¹⁴

To test Chinese functionality in del.icio.us, I added my picture¹⁵ and, as expected, I can pull up the picture via the tag of my Chinese name¹⁶ and a search works.¹⁷

How do you get `fran%C3%A7ais` from `français`? With a bit of Python programming, you can convince yourself that it's a URL-encoding of the UTF-8 encoding of français.

```
>>> print chr(231)
ç>>> print urllib.urlencode({'q':chr(231).decode('ISO-8859-1').encode('utf-8')})
q=%C3%A7
```

YouTube copes well with Chinese characters too: I can find a video tagged with my Chinese name:

http://www.youtube.com/results?search_query=%E4%BD%99%E4%BF%8A%E9%9B%84.

In rel-tag

The rel-tag specification gives the following example of how to encode tags:¹⁸

```
<a href="http://technorati.com/tag/Sant%C3%A9+et+bien-%C3%AAtre" rel="tag">Santé
et bien-être</a>
```

We can verify that the tag is the URL encoding of the UTF-8 encoding of the tag string. In Python, the following code:

```
import urlencode
s = "Santé et bien-être"
```

¹² http://www.cojak.org/index.php?function=code_lookup&term=4F59 and <http://www.unicode.org/cgi-bin/GetUnihanData.pl?codepoint=4F59>

¹³ <http://del.icio.us/url?url=http://ies.berkeley.edu/xbf/>

¹⁴ http://del.icio.us/search/?fr=del_icio_us&p=fran%C3%A7ais&type=all

¹⁵ <http://del.icio.us/url?url=http://flickr.com/photos/raymondyeec/79915850/>

¹⁶ <http://del.icio.us/tag/%E4%BD%99%E4%BF%8A%E9%9B%84>

¹⁷ http://del.icio.us/search/?fr=del_icio_us&p=%E4%BD%99%E4%BF%8A%E9%9B%84&type=all

¹⁸ http://microformats.org/wiki/rel-tag#Encoding_issues as <http://microformats.org/wiki?title=rel-tag&diff=0&oldid=18625>

```
u = s.decode('iso-8859-1')
print urllib.urlencode({'q':u.encode('utf8')})

returns

q=Sant%C3%A9+et+bien-%C3%AAtre

http://technorati.com/tag/Sant%C3%A9+et+bien-%C3%AAtre

is a search on Santé et bien-être.
```

End sidebar

Gathering Content Through Tags in Technorati

Technorati is a search engine, focused primarily on searching weblogs but also "tagged social media" (specifically, photos in Flickr and videos in YouTube). Technorati is an excellent case study of how a website crawls for tags on the web and then uses those tags to organize digital content. (Think of technorati as a big tag-based mashup.) Let's now look in detail at how Technorati presents tags to users and how it finds the tags in the first place.

Searching Technorati with Tags

The primary emphasis in the Technorati user-interface is on searching by tag. In fact, the default search is a tag search. For instance, a search for the term **mashup** brings you to the page

<http://technorati.com/tag/mashup>

Generally, items for a given tag are at the URL:

<http://technorati.com/tag/{tag}>

where {tag} is the URL-encoded of the UTF-8 encoding of the tag. The items are broken down by:

- * blog posts: <http://technorati.com/posts/tag/{tag}>
- * videos: <http://technorati.com/videos/tag/{tag}>
- * photos: <http://technorati.com/photos/tag/{tag}>
- * weblogs: <http://technorati.com/blogs/tag/{tag}>

Note that you can string tags together with OR to search for multiple tags.

A quick way to get a feel for Technorati is to look at the most popular search listed at:

<http://technorati.com/pop/>

How Technorati Finds Tags on the Web

Technorati derives its tags from a variety of sources, as documented at <http://technorati.com/help/tags.html>:

- * categories embedded in Atom and RSS 2.0 feeds (See Chapter 4 for more on feeds)
- * tags in links using the rel-tag microformat e.g., `tagname` (See Chapter 18 for a complete description)
- * tags from public photos in Flickr
- * tags from public videos in YouTube

Word Inflections and Syntactic Constraints in Technorati Tags

As with Flickr and del.icio.us, singular and plural nouns in tags are not conflated. For example,

<http://technorati.com/tag/mouse>

and

<http://technorati.com/tag/mice>

return different results. Technorati is, however, able to recognize that `mouse` and `mice` are related tags, as are `peripherals` and `animals`. Unlike Flickr, but like del.icio.us, punctuation in Technorati tags are significant in tag-based searches. For example,

<http://technorati.com/tag/san+francisco>

returns different results from

<http://technorati.com/tag/san-francisco>

Tag searches are not case-sensitive in Technorati, though other applications use the `rel-tag` microformat may be case-sensitive. Through `rel-tag`, you should be able to pass in the full range of non-ASCII words as tags. (See the Sidebar on representing on representing non-ASCII characters in tags learn more.)

The next time you want to make a mashup of digital content based on tags, you can model what you do on how Technorati has dealt with making tags from different websites work (interoperate) with one another. Moreover, you can leverage its work by linking directly to Technorati (through its URL Language) or use its API (<http://technorati.com/developers/api/>).

Case study: Using Tags to Mashup Flickr and del.icio.us

In this section, I show how you can use tags in del.icio.us to collect Flickr pictures and make a very simple visual collection. The idea is simple: you can use del.icio.us to gather pictures from Flickr by tagging Flickr URLs in del.icio.us and use a specific del.icio.us tag on all the pictures you want in the same set. Because del.icio.us shows thumbnails of photos from Flickr, you get a very simple album maker using this combination of Flickr and del.icio.us and tagging.

Here's an example:

<http://del.icio.us/rdhyee/set:Berkeley>

In this case, I've tagged a selection of Flickr URLs with the tag `set:Berkeley`.

This mashup is certainly not a replacement for Picasa or iPhoto. You can't sort the pictures, for instance -- though you could imagine adding another tag with a number and writing a Greasemonkey script that would sort the pictures for you (and allow you to edit the ordering). This mashup is a helpful supplement to Flickr, but you might ask, why not just use the Flickr favorites or collections to accomplish this goal? The problem that this little mashup solves is combining your own photos with those of others. Favorites must be other people's pictures; your collections can contain only our own photos.

Other Systems that Use Tagging

Many other applications use tags. If you look at the Wikipedia article on tags, you will see some of the following¹⁹ mentioned:

- * Other social bookmarking sites
- * Other photo sharing sites
- * Video sites such as YouTube
- * Email: GMail and Thunderbird 2.0
- * Blogs: you can generate tag clouds based on categories from your blog (e.g., Ultimate Tag Warrior 3 WordPress plugin²⁰).

Relationship of Tags to Formal Classification Schemes

I don't think that folksonomies will supplant formal subject headings and taxonomies. There's plenty of room to experiment with the interplay between folksonomic and taxonomic approaches. Indeed, how can one combine some of the simplicity of tagging with the careful structures of formal classification schemes? In this section, I sketch a specific example to highlight some of the relevant challenges.

Let's come back to an example I first used in Chapter 1, the book Czesław Miłosz's *New and Collected poems 1931-2001*, specifically the hardcover edition with the ISBN-10 of 006019667X. You can search for the book at the Library of Congress at

<http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi?v3=1&DB=local&CMD=kisn+006019667X&CNT=10+records+per+page>

to learn how the Library of Congress has formally classified the book and its author. The book is assigned to the Library of Congress Subject Heading (LCSH) - `Miłosz, Czesław` `Translations into English`:

<http://catalog.loc.gov/cgi-bin/Pwebrecon.cgi?SC=Subject&SA=Mi%C5%82os%20Czes%C5%82aw%20Translations%20into%20English>

¹⁹ <http://en.wikipedia.org/wiki/Tags>

²⁰ <http://www.neato.co.nz/ultimate-tag-warrior/>

Through this subject heading, which you can access through its corresponding URL, you can get all the books that are classified into the same group. In this specific case, you can reliably find a list of many, if not all, of the English translations of Miłosz's poetry published in the U.S.

Why does this matter? By using the LCSH as a category, I get to leverage the careful and reliable work that the Library of Congress in classifying books. Just because you use tags doesn't mean you have to ignore formal classifications.

The LCSH is not the only formal classification scheme around for books. If you look the same book up at OCLC's worldcat.org:

<http://worldcatlibraries.org/wcpa/isbn/006019667X>

you will find the book listed under the subject of **Miłosz, Czesław**:

<http://worldcatlibraries.org/search?q=su%3AMi%C5%82os%C5%82+Czes%C5%82aw>

The subject headers used by OCLC are based on its FAST project, which aims to simplify yet be upward compatible with LCSH:

<http://www.oclc.org/research/projects/fast/>

To see a sophisticated example of tags can be effectively combined with formal classification, let's look at librarything.com, a social book cataloging site. You can feed a ISBN to librarything with

<http://www.librarything.com/isbn/{isbn}>

which will redirect to a URL with a **work-id** (Different editions of a book, which can have different ISBNs, are collected together under the same **work-id**):

<http://www.librarything.com/work/{librarything-work-id}>

Using our example, the URL

<http://www.librarything.com/isbn/006019667X>

redirects to:

<http://www.librarything.com/work/161671>

where you see tags that users of librarything have applied to the book. At the same time, librarything lists at

<http://www.librarything.com/work-info/{librarything-work-id}>

i.e.,

<http://www.librarything.com/work-info/161671>

how the book has been formally classified (including such metadata as the Library of Congress Call Number, the Dewey Decimal classification) along with the LCSH:

<http://www.librarything.com/subject.php?subject=Mi%B1os%C5%82+Czes%B1aw%09Translations+into+English>

Caution There is an error in character encoding in librarything that causes **Miłosz, Czesław to be incorrectly displayed.**²¹

Summary

In this chapter, we looked at how to use tags to create mashups. We first compared and contrasted how tags are used in Flickr and del.icio.us. Flickr's tagging system, is an example of a narrow folksonomy, enabling textual searches to be done over visual media. As a broad taxonomy, del.icio.us involves many people tagging any given bookmark, creating multiple sets for tags for a bookmark. We considered some factors that reduce the reliability of tags and study specifically the issue of singular vs .plural nouns and the role played by syntactic constraints as the use of spaces, punctuation marks, multiple cases, and non-ASCII characters in Flickr and del.icio.us. We looked at Technorati as an example of a tag-based search engine as a case study of how to use tags to relate disparate digital content. We showed how you can create a very simple mashup of Flickr and del.icio.us using del.icio.us tags to create sets of pictures that intermix your photos and other people's photos in Flickr. We ended this chapter with an example of combining tags with formal classification schemes in the context of books.

²¹ <http://www.librarything.com/talktopic.php?topic=12559#138896>