

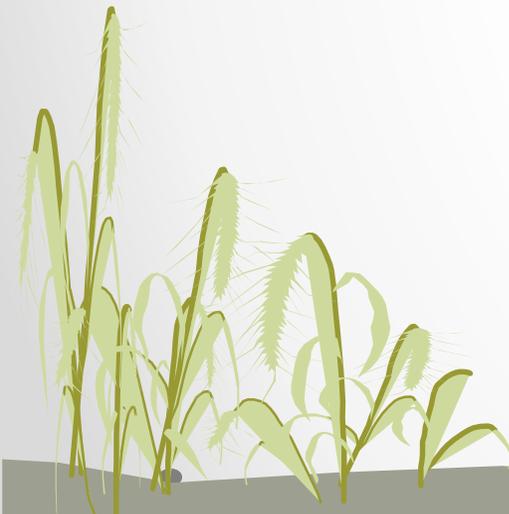


User-Generated Metadata: Boon or Bust for Indexing and Controlled Vocabularies?

Louise Spiteri

School of Information Management

Dalhousie University



The traditional metadata landscape

- Traditionally, client participation in web-based repositories of information has been largely reactive: Clients can search for and select items from these repositories, but have little ability to organize and categorize these items in a way that reflects their own needs and language.
- Digital document repositories such as library catalogues and bibliographic databases index the subject of their contents with keywords or subject headings. Traditionally, such indexing is performed either by an authority, such as a librarian or a professional indexer, or else is derived from the authors of the documents.



User generated metadata

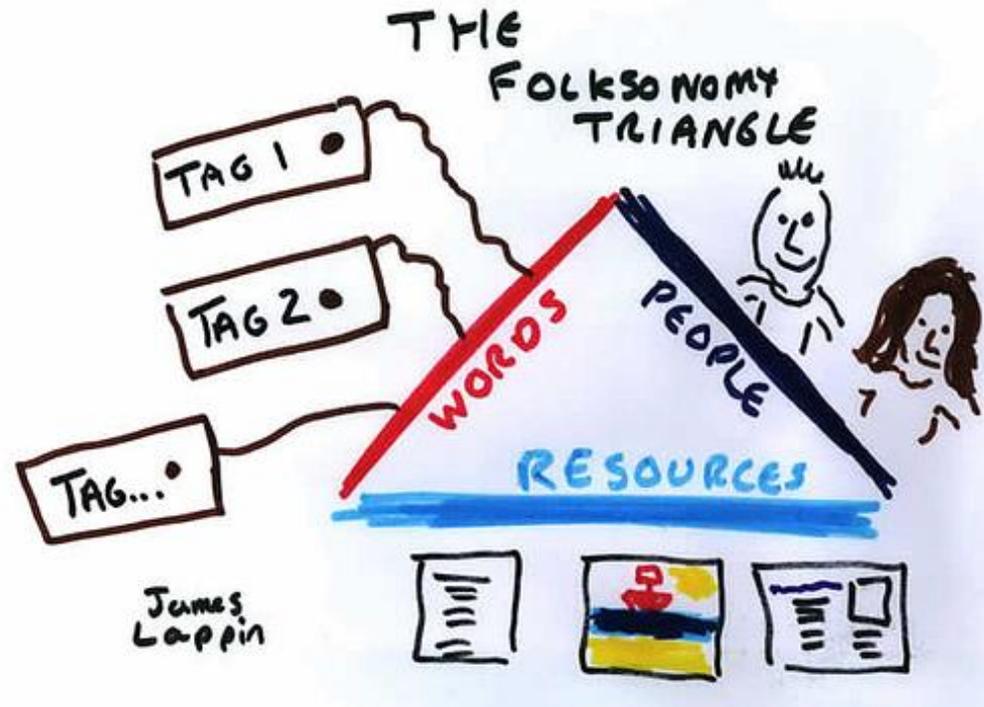
In recent years, significant developments have occurred in the creation of customizable user features in a wide variety of websites.

These features offer users the opportunity to customize and store items of interest to them, such as wish lists or records of items to read, watch, or listen to; collections of photographs; blog posts; wikis, and so forth.

Users can organize and categorize these items by adding their own keywords; further, in many cases, they can add further metadata in the form of ratings and reviews.

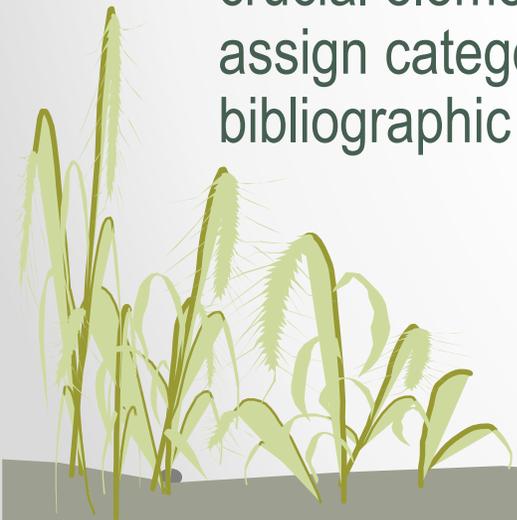


Social tagging and folksonomies



Tagging

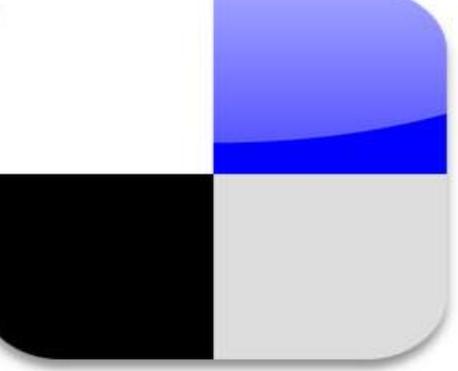
- A tag is a non-hierarchical keyword or term assigned to a piece of information (e.g., a website, digital image, ebook, etc.). Tags are assigned by the creator of the information, or the person viewing it.
- User-generated metadata such as tags and categories go back to the late 1990s with the growth of blogs, where authors assigned categories and tags to individual blog posts. The crucial element here is that this type of tagging is purely individualized; only the author can assign categories and tags, so it's not much different from author-assigned keywords in bibliographic databases.



Social tagging

- The social aspect of assigning tags was popularized in 2004 by social bookmarking sites such as [Delicious](#), [CiteULike](#), and [Connotea](#) (discontinued this year), as well as social image sites like [Flickr](#). The point of these sites is not just to control what information is posted, but to share that information, and its metadata, with a fellow community of users.
- Delicious is often considered the parent of social tagging. Although Delicious has lost some of its popularity recently, since people are using Twitter increasingly to follow sites of interest, it presented a novel and important way of keeping track of, and organizing, links to websites of interests that are independent of any computer; it was, in fact, an instance of cloud computing before that term meant anything.

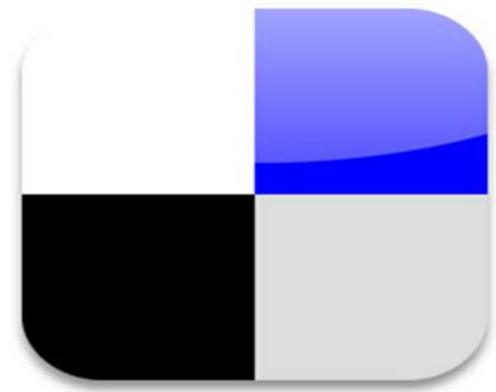




Delicious and social tagging, 1

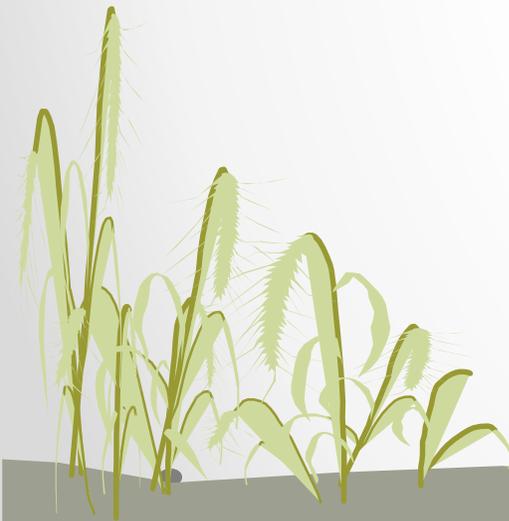
- You can add the URLs of websites of interest to you in a cloud environment; when you do so, the system prompts you to add tags of your choosing (no limit on the number).
- If you choose to make these links public, anyone who follows your account can see all the tags you have assigned, as well as the bundles, or categories, under which these tags are organized.
- One of the innovative features of Delicious is its recommender feature: When you add a URL to your collection, you are provided with a suggestion of tags that others have assigned to the URL.





Delicious and social tagging, 2

- This recommender system leads to the crowdsourcing, or social aspect of tagging.
- In my own blog, for example, I have total control over the tags and categories I create. In Delicious, I can use the wisdom (or folly) of the crowd: The more often I used the recommended tags, the more I am contributing to a relatively standard set of tags, so it's possible to form some kind of standardized vocabulary with a recommender system.

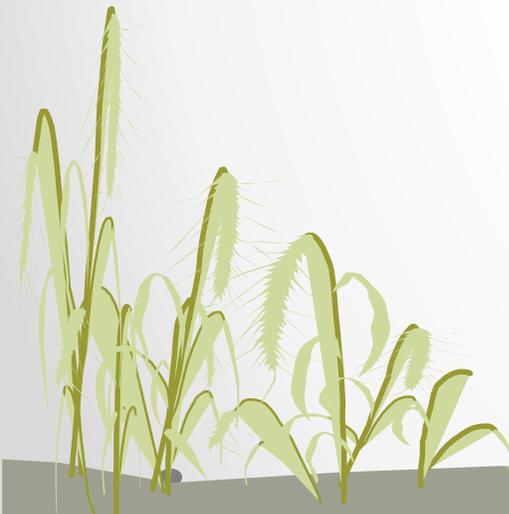


Examples of my Delicious tag bundles and tags

TAG BUNDLES [Edit](#)

- cataloguing
- films
- language
- shopping
- Vegan

shopping	68
imported	49
Canada	44
vegan	31
Films	20
Cataloguing	19
Recipes	17
food	16
Cosmetics	15
reviews	14
Books	14
LIS_Journals	11
libraries	11
environment	10
Nova_Scotia	8
Conferences	8
organic	8



Folksonomies

- Folksonomies is a term used to describe the social aspect of tagging. The term folksonomy was created by Thomas Vander Val in a discussion on an online information architecture site, and represents a merging of the terms folk and taxonomy.
- In a folksonomy the set of terms is a flat namespace; there are no clearly defined relations between and among the terms in the vocabulary, unlike formal taxonomies and classification schemes, where there are multiple kinds of explicit relationships (e.g., broader, narrower, and related terms) between and among terms. Folksonomies are simply the set of terms that a group of users tagged content with; they are not a predetermined set of classification terms or labels.



Popularity of social tagging

- The growing popularity of social tagging can be attributed to:
 - An increasing need to exert control over the mass of digital information that we accumulate on a daily basis
 - A desire to democratize the way in which digital information is described and organized by using categories and terminology that reflect the views and needs of the actual end users, rather than those of an external organization or body.



Perceived need for social tagging

- Perhaps the most important strength of social tagging is that it allows users to organize resources in a way that reflects directly their own vocabulary and needs.
- Social tagging represents a fundamental shift in that it is derived not from professionals or content creators, but from the users of information and documents.
- Folksonomies can adapt very quickly to changes in user needs and vocabulary, and adding new terms to a folksonomy incurs virtually no cost for either the user or the system.



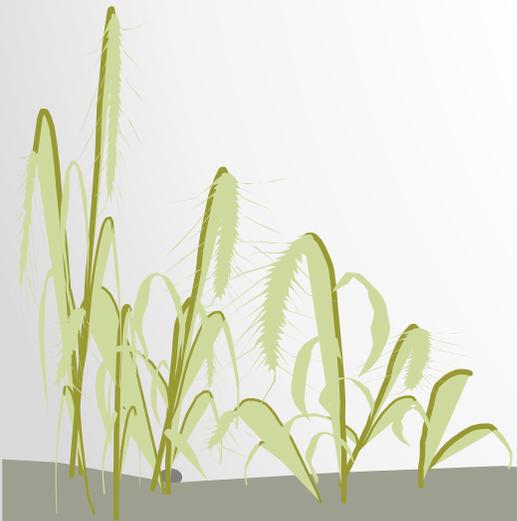
Limitations of social tagging, 1

- Ambiguity (e.g., Ant has been used for Actor Network Theory, and Apache Ant, a Java programming tool)
- Polysemy (Port: Wine; Computer port; left side of a ship; where ships unload, etc.)
- Synonymy (cataloguing/cataloging; flower/flowers)
- Variations in levels of specificity (e.g., Vegetarian versus ovo-lacto vegetarian, ovo vegetarian, lacto vegetarian, fruitarian, pescetarian, etc.)



Limitations of social tagging, 2

- Folksonomies provide no guidelines for the use of compound headings, punctuation, word order, and so forth; for example, should one use the tag vegan cooking or cooking, vegan, vegancooking, or vegan_cooking? Finally, and not insignificantly, the terms could be applied incorrectly.



Examples of inconsistent tagging

No standard
citation order



recipes chickpea tahini

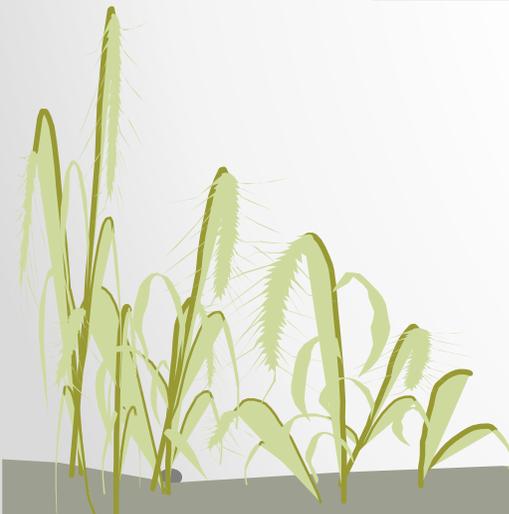
No standard
structure for
compound nouns



coconut-milk

rawvegan

blog_food



Yes, but

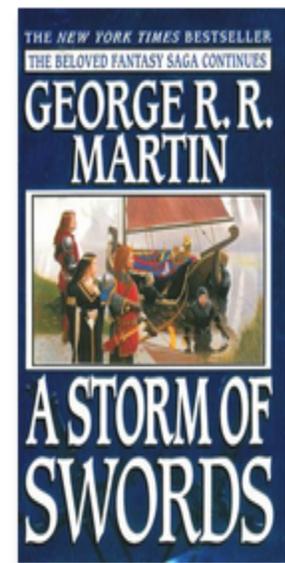
- Users are willing to tolerate the shortcomings of social tagging because ultimately they lower barriers to cooperation.
- Users do not have to agree upon a hierarchy of tags; they strive to achieve a degree of consensus over the general meaning of tags.
- In recommender systems, as a URL receives more and more bookmarks, the set of tags used in those bookmarks becomes stable across different users. From my experience, for example, I am more likely to choose a recommended tag than create my own.



Ubiquity of social tagging

- Given the ease of creating and using tags, nearly any member of the Internet community can make use of this tool. Although interaction through social networking is one of the primary uses of tagging, the process offers benefits for the solitary user as well, namely, the opportunity to access bookmarks online from any computer (e.g., Delicious), to impose structure on written works (e.g., blog posts), academic research and file sharing (e.g., CiteULike), multimedia sites (e.g., Flickr, Photobucket, and YouTube), reading collections (e.g., GoodReads), etc.





A Storm of Swords (A Song of Ice and Fire #3)

by George R.R. Martin

★★★★★ 4.51 ·  rating details · 186,914 ratings · 10,084 reviews

Here is the third volume in George R.R. Martin's magnificent cycle of novels that includes *A Game of Thrones* and *A Clash of Kings*. Together, this series comprises a genuine masterpiece of modern fantasy, destined to stand as one of the great achievements of imaginative fiction.

Of the five contenders for power, one is dead, another in disfavor, and still the wars rage as al...more

Genres

Fantasy	10,271 users
Science Fiction Fantasy	480 users
Fantasy > Epic Fantasy	468 users
Epic	246 users
Fantasy > High Fantasy	199 users
Adventure	192 users
Fantasy > Dragons	191 users
Adult	184 users
Fantasy > Magic	114 users
Science Fiction	108 users



This photo also appears in

[miniBIGS' photostream](#)

[LEGO \(group\)](#)

[Song of Ice and Fire \(group\)](#)

[Minifigs \(group\)](#)

[Custom LEGO Minifigs \(group: 22,928\)](#)

[+ Toys \(玩具,おもちゃ\) \(group\)](#)

[...and 4 more groups](#)

Tags

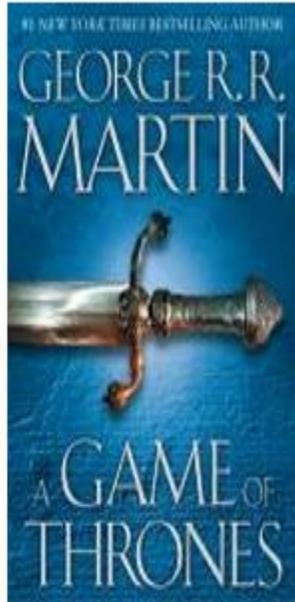
[lego](#) [custom](#) [minibigs](#) [game of thrones](#)
[tyrion](#) [dany](#) [jon snow](#) [arya](#) [ned](#) [stark](#)

Find it at EPL

Items: 35
Volumes: 2
Available: 0
Holds: 76

[Place a Hold](#)

[View Availability Details »](#)



My Rating:



[Add a Comment](#)

[Add Tags](#)

[Add More](#)

[Add to My Shelves](#)

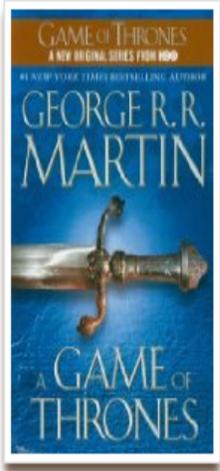
[Save to a List](#)

[Print](#)



Tags

- [fantasy](#)
- [dark](#)
- [medieval](#)
- [war](#)
- [epic](#)
- [political](#)
- [adventure](#)
- [battles](#)
- [betrayal](#)
- [dragons](#)
- [embarrassment of riches blog](#)
- [epic adventure](#)
- [epic fantasy](#)
- [fantasy war](#)
- [fun](#)



A Game of Thrones

by [George R. R. Martin](#)

Other authors: See the [other authors](#) section.

Series: [A Song of Ice and Fire \(1\)](#)

Members	Reviews	Popularity	Average rating	Conversations / Mentions
18,664	538	80	★★★★½ (4.41)	4 / 895

[+ Add to your library](#)

[+ Add to wishlist](#)

▼ Tags

[numbers](#)

[show all](#)

[2011 \(99\)](#)
[2012 \(51\)](#)
[A Song of Ice and Fire \(770\)](#)
[adventure \(94\)](#)
[dragons \(225\)](#)
[ebook \(111\)](#)
[epic \(281\)](#)
[epic](#)
[fantasy \(264\)](#)
[fantasy \(3,717\)](#)
[fiction \(1,438\)](#)
[George R. R. Martin \(107\)](#)
[high fantasy \(118\)](#)
[intrigue \(82\)](#)
[Kindle \(129\)](#)
[knights \(83\)](#)
[magic \(95\)](#)
[medieval \(130\)](#)
[novel \(157\)](#)
[own \(87\)](#)
[paperback \(75\)](#)
[politics \(109\)](#)
[read \(273\)](#)
[read in 2011 \(73\)](#)
[science fiction \(110\)](#)
[series \(328\)](#)
[sff \(138\)](#)
[signed \(50\)](#)
[to-read \(147\)](#)
[unread \(83\)](#)
[war \(143\)](#)

How does social tagging affect indexers?

- Tagging is not going away. When you see the success of sites like GoodReads and LibraryThing, it's evident that people like contributing their own data (in the form also of reviews).
- People follow each other on sites like GoodReads and LibraryThing to see what their site friends are reading. These sites therefore act as recommender sites for other items you might wish to read, and so forth. Tagging is not going away, so it's best to embrace it.



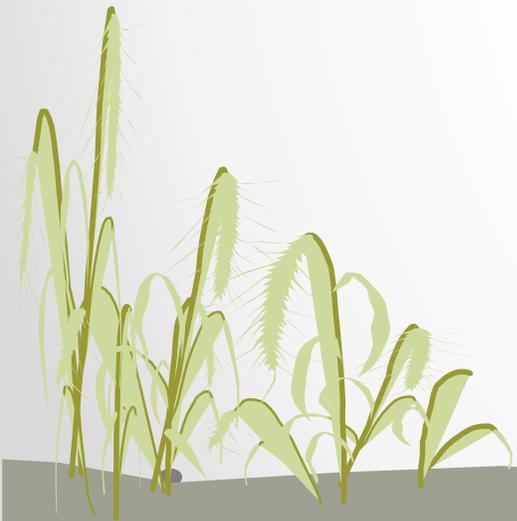
Ideal indexing scenario: Combine controlled vocabulary with tags.

- The ideal scenario is to have a system that includes both controlled vocabularies and tags. Take blogs, for example. The categories are more rigidly controlled; when you create a blog post, you are prompted to assign it at least one category. These categories can be firmly controlled, i.e., outside users cannot add, delete, or modify the categories. You can use the categories for directory-style browsing, which can get a little time consuming if the blog has a lot of posts.
- In a corporate environment, the creation and maintenance of these categories can be assigned to 1-2 administrators. You can add value to the blog by allowing the authors of the posts to add their own tags, in addition to choosing from the assigned categories.



Blogs, continued

- Blog platforms do not generally have tag recommenders, since each post is unique, rather than a common URL. What will happen, however, is that as you are typing a tag, if a similar tag has been assigned, it will be shown as a recommender tag. It has to be an almost exact match for this to happen, however, e.g., If I type in veg, it will prompt me with vegan, since I've used this tag before.



Information retrieval systems, 1

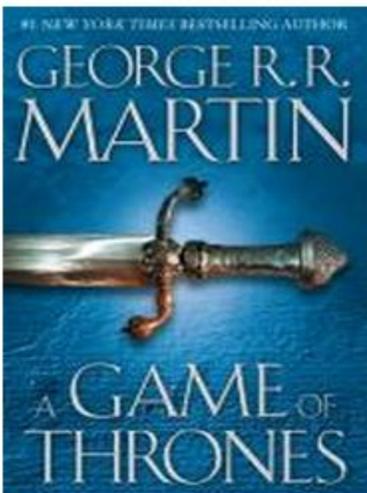
- With systems like library catalogues and bibliographic databases, there is merit to allowing users to add their own tags. The original metadata record (e.g., the MARC record) can't be tampered with and the controlled vocabulary stays intact. In this case, the tags add as a supplement or complement to the controlled vocabulary.
- User tags may reflect more accurately current information, since it takes a while to update thesauri and subject headings; these tags can reflect more idiomatic language, rather than the more formal language that is typical of controlled vocabularies



Information retrieval systems, 2

- In multi-cultural environments, users can add tags in their own language (restricted to roman alphabet), which can add to make the bibliographic items more retrievable and relevant to the client.
- Because tags could be associated with an individual (depending on the system), I can connect to like-minded readers, researchers, and so forth, via their tags, which is not something that can be done via controlled vocabularies.
- User tags help us monitor changes in language and can help us update our thesauri and subject heading lists to reflect the language of our clients.





My Rating:



 Add a Comment

 Add Tags

 Add More

 Add to My Shelves

 Save to a List

 Print



Note: no subject headings have been assigned

Item Details

Description	Full Record	Author Notes	Excerpts	Reviews	Preview
-------------	--------------------	--------------	----------	---------	---------

Author:	Martin, George R. R.
Title:	A game of thrones
Publisher:	Bantam Books
Imprint:	New York : - Bantam Books
Pages:	694
Series:	A song of ice and fire - bk. 1
ISBN:	0553103547, 0553573403, 0553588494, 9780553573404, 9780553386790
Language:	English
Notes:	A fantasy novel. "Spectra." Bantam Spectra pbk. ed. contains maps. A fantasy novel.
Statement of responsibility:	George R.R. Martin
Characteristics:	694 p. ;,25 cm.
Author (Original Script):	Martin, George R. R.

Tags

- › [fantasy](#)
- › [dark](#)
- › [medieval](#)
- › [war](#)
- › [epic](#)
- › [political](#)
- › [adventure](#)
- › [battles](#)
- › [betrayal](#)
- › [dragons](#)
- › [embarrassment of riches blog](#)
- › [epic adventure](#)
- › [epic fantasy](#)
- › [fantasy war](#)
- › [fun](#)

Iron Man

(DVD - 2008)

Average Rating



Find it at EPL

Total Copies: 18

Available: 0

On Hold: 17

[Place a Hold](#)

[View Availability Details »](#)



My Rating:



[Add a Comment](#)

[Add Tags](#)

[Add More](#)

[Add to My Shelves](#)

[Save to a List](#)

[Print](#)

Subject Headings

› [Iron Man \(Fictitious Character\)](#)
— [Drama.](#)

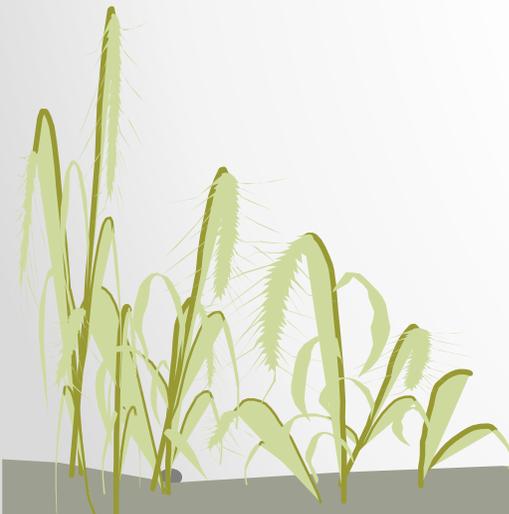
Tags

- › [action](#)
- › [science fiction](#)
- › [action adventure](#)
- › [action film](#)
- › [adventure](#)
- › [adventure film](#)
- › [dark](#)
- › [drama](#)
- › [fastastical](#)
- › [inspirational](#)
- › [international weapons trade](#)
- › [inventors](#)
- › [iron man](#)
- › [science fiction film](#)
- › [super hero](#)

« [Less](#)

Newer forms of social tagging

- Newer variations of social tagging can be found in hashtags used in Twitter, Tumblr, Instagram, and so forth. Hashtags are a quick way to follow a stream of tweets assuming, of course, that people use the same hashtag consistently.
- It's not uncommon for the same thread to be distributed across variations of the same hashtag, e.g., ASIST13; ASIST2013, ASISTCONF, and so forth.



Hashtags, 1

- A hashtag can be used by any person, which means that conference attendees, for example, can create various hashtags for the same conference, depending on who follows whom, and how many different attendees created hashtags for the same event.
- Hashtags suffer from the same problems as tags and any other uncontrolled vocabularies, as discussed earlier.
- In a corporate environment, you can create controlled hashtags to limit the amount of “noise;” it is increasingly common for hashtags to be created officially for public events so that everyone uses the same hashtags.



Hashtags, 2

- Hashtags are not registered or controlled by any one user or group of users
- Hashtags cannot be retired from public usage, which means that hashtags can be used in theoretical perpetuity depending upon the longevity of the word or set of characters in a written language.
- Hashtags do not contain any set definitions, meaning that a single hashtag can be used for any number of purposes determined by their users.



Hashtags, 3

- Hashtags are also used informally to express context around a given message, with no intent to actually categorize the message for later searching, sharing, or other reasons, e.g., “the Leafs blew it again #disappointed, #shouldbeusedtoit, #maybe next time.
- As you can see, there is much potential for the overuse of hashtags, and they can quickly lose their usefulness or appeal.
- Facebook is supposed to be incorporating hashtags soon.



 **BDC**  @BDC_News ×
 Followed by JoAnne Akerboom
 Promoted · Follow

 **JobLIST Library Jobs** @ALA_Jo... ×
 Followed by Jason Harris and oth...
 Follow

 **Sarah V** @SVisin ×
 Followed by DJIM and others
 Follow

[Find friends](#)

Trends · [Change](#)

#mygirlfriendnotallowedto

Tsonga

#heartbreaker

#unpopularopinion

#teenwolf

James Holmes

#Vancouver

Telus

Spartan Assault

Canada

Results for #highered 

Top people · [View all](#)

 **Ed Shahzade**  @Ed ⌵ [Follow](#)

Connecting People on Twitter Since 2006.
 Discreet Communications & Twitter Advisor to
 Fine Clients. Lighthouse to 3 sons,
 Constitutional Patriot, Team Player.

Tweets [Top](#) / [All](#) / [People you follow](#)

20 new Tweets

 **TimesHigherEducation** @timeshighered 1h

"The difference between us is that I can just change my job. You academics are what you do." ow.ly/IGJKX #highered

 Retweeted 10 times

[Expand](#)

 **Higher Education** @GdnHigherEd 29m

Elizabeth Warren's loan bill offers lifeline to America's indebted students | Sam Blum gu.com/p/3gbxv/tw via @commentisfree #highered

Followed by Laura Pasquini and 2 others

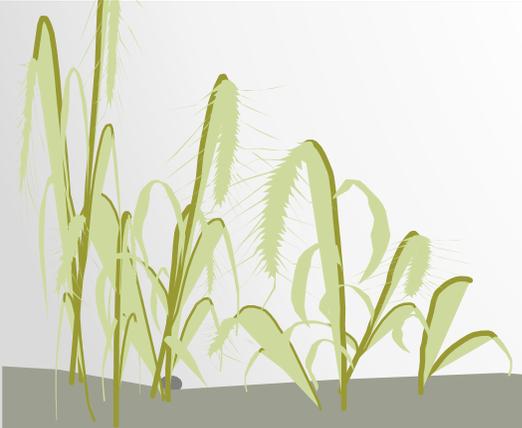
[View summary](#)

 **Bramoso Pizza** @BramosoPizza 1h

Looking for something to do tonight? Still time to buy your ticket for tonight's @LocalTasteTrs! #yummy #halifax #quinpoolroad

 Retweeted by ilovelocalhfx

[Expand](#)



Geotags, 1

- Geotags are another innovative use of social tagging. GeoTagging is the process of adding geographic metadata to images, e.g., in Flickr, QR codes, RSS feeds, and so forth.
- Geotags may consist of latitude and longitude coordinates, altitude, distance, place names, etc.
- Because of the numerical nature of many geotags, you are more likely to find consistency in the tags.



Geotags, 2

- Geotagging-enabled information services can be used to find location-based news, websites, or other resources.
- Geotagging can tell users the location of the content of a given picture or other media.
- With most smartphones, geotags are assigned automatically by the phone; this means that when you post your pictures publicly, this information can be available to anyone. This does raise some privacy concerns, as geotagging can serve as a form of tracking. You have the option to disable this feature, but the default is that it will run in the background.



Example from Panoramio

On The Road - by Bowser

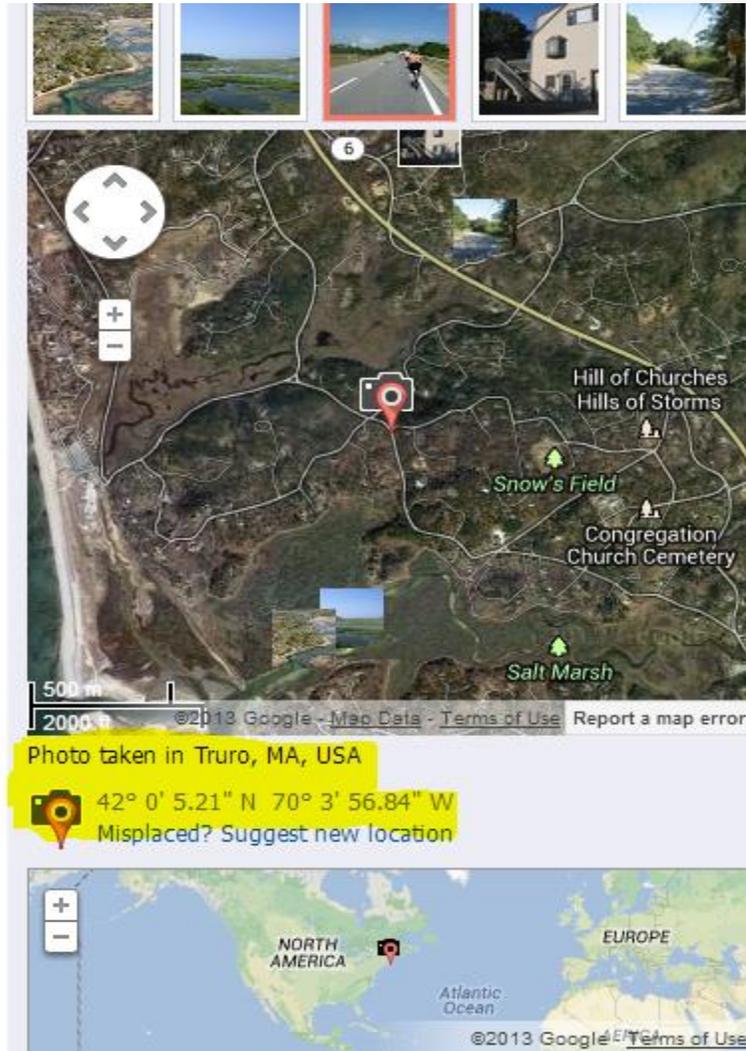
See in Google Earth Share on: 



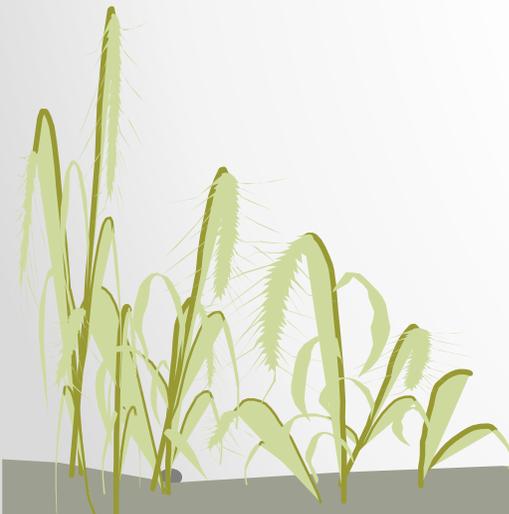
by [shootinphotos](#)

Selected for Google Earth [?] - ID: 77431270

More photos by [shootinphotos](#)



- How does social tagging impact what you do?
- How do you plan to work with social tagging?



Questions?

Louise Spiteri

Louise.Spiteri@dal.ca

