A Tool for Collaborative Research

James Miller
UC Berkeley School of Information
May 2009

Contents

1. Overview

Social bookmarking is a web-based linking system through which internet users apply metadata tags to web resources. Users associate one or more keywords with the URIs of resources. The result is a bookmark, which may be shared with other users of the social bookmarking application. Websites dedicated to social bookmarking facilitate this process and provide a network of users as de facto collaborators. The aggregate product of social bookmarking is a "folksonomy," which is a loosely organized network of users, resources and tags, an informal taxonomy.

The Standardsland Tagging Service (Slants) is a collaborative research tool based on social bookmarking. Its approach is to automatically collect bookmarks relevant to a particular domain from users of the Delicious social bookmarking website. By focusing on a particular domain, Slants targets users who have a common interest, and who therefore select keywords from a relatively narrow vocabulary range and with a reasonably predictable purpose. Creating a community around a collaborative process may also result in individuals' tagging becoming more oriented toward a consensus vocabulary, though the jury is still out on this.[1]

The bookmarks harvested via the RSS feeds are stored in a database. Slants depends on a thesaurus made specifically for the domain to normalize the tagging vocabulary of the individual users. The thesaurus is equally important in retrieving bookmarks from the database, providing the mechanism for collocating search results by subject.

The domain chosen for this project focuses on information and communication technology (ICT) standards as they affect public policy, and, more specifically, issues of identity. However, the basic approach, aggregating and then organizing the social bookmarking work of multiple actors, could be applied to any domain.

2. The Need Case Addressed by Slants

Technical choices have nontechnical consequences. In the debate over the public policy consequences of ITC design decisions, some observors have emphasized technical solutions. Cranor and Reagle, for example, speak of separating mechanism and policy, so that policy is not imbedded in the internet's technical architecture. [2] Clark *et al* see an ongoing "tussle" for control between the competing interests of private enterprise, governments, users, and others, in which technical designers cannot be value-neutral, but should assume responsibility for finding technical solutions to policy-based problems.[3]

Others disagree with an approach that espouses technical fixes, arguing for citizen engagement on the side of important public policy interests. Morris and Davidson argue that a conceptual gulf exists between the domains of public policy and standards development that limits the ability of

[1] Tonkin, Emma *et al.*
[2] Cranor, Lorrie F. and Joseph Reagle.
[3] D.D. Clark, J. Wroclawski, K.R. Sollins, and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet,"

public policy advocates "to identify policy concerns in the large number of standards efforts…."[4] In addition to this knowledge gap, Morris and Davidson identify other challenges, including institutional and cultural factors that alienate technologists from advocates; the difficulty for advocates in maintaining institutional attention throughout the long duration of standards development efforts; the size, number and diversity of standards; and the lack of scalability in the process of identifying standards with public policy issues.

To mitigate these barriers to effective advocacy, Morris and Davidson propose that standards bodies adopt a "public policy impact statement" as an "early warning system" that would alert advocates to those relatively few standards that have public policy implications.[5] However, with more than 400 ICT standards consortia worldwide, even the task of coordinating impact statements would be formidable.

Finally, Beth Simone Noveck calls for a new citizen democracy in which small, decentralized groups engage in "the practice of democracy with a 'small d'…a model of consociational democracy premised on the collective action of small groups. [6]" Noveck sees this as a form of social action made necessary by the defects of institutionalized democratic processes and one that is enabled by the collaborative technologies generally referred to as computer supported cooperative work (CSCW).

The Slants approach falls in with Novacek's belief that technology can be a force multiplier for advocacy by creating a collaborative process that leverages the activities of many to achieve a shared purpose. But there is no reason that multiple approaches to advocacy can't coexist. The alternative of inaction, as voiced by Lawrence Lessig, may have dire consequences:

> We will treat code-based environmental disasters—like the loss of privacy, like the censorship of censorware filters, like the disappearance of an intellectual commons—as if they were produced by gods, not by Man. We will watch as important aspects of privacy and free speech are erased by the emerging architecture of the panopticon, and we will speak, like modern Jeffersons, about nature making it so, forgetting that here, we are nature. We will in many domains of our social life come to see the net as the product of something alien—something we cannot direct because we cannot direct anything. Something instead that we must simply accept, as it invades and transforms our lives.[7]

2.1 The Semantic Gap

The divide between the public policy and standards domains, which Morris and Davison cite as one of the challenges to successfully monitoring standards development, is also a challenge for any lexical search system seeking to detect links between public policy issues and technical

[4] Morris, John/Alan Davidson, 2003: Policy Impact Assessments: Considering the Public Interest in Internet Standards Development
[5] Morris, John, Alan Davidson, 2003: Policy Impact Assessments: Considering the Public Interest in Internet Standards Development, p10.
[6] Noveck, B. S. (2005), 'A democracy of groups'; http://www.firstmonday.org/issues/issue10_11/noveck/index.html
[7] Lessig, L. (2006). Code 2.0. New York: Basic Books, p 338.

standards. In considering what kinds of tools could signify these connections, one is quickly convinced that human understanding of both the issue and the standard is necessary.

Consider the following passages:

- "The Interface Identifier [AARCH] for an Ethernet interface is based on the EUI-64 identifier [EUI64] derived from the interface's built-in 48-bit IEEE 802 address."[8]

- "Under IPv4, the predecessor to IPv6, Internet addressing allowed a reasonable amount of privacy and anonymity, because a numeric address was typically not tied to any particular machine or user. With IPv6, however, the standard provided that in many cases a user's address would be derived from the unique MAC (Medium Access Control) address embedded in the user's Ethernet network card."[9]

The first passage is from an early version of the Internet Protocol version 6 (IPv6) specification. One of the main tasks addressed by this spec was the need to expand the number of IP addresses to accommodate future growth of the Internet. The method proposed was to tie the IP address to the MAC address, as explained in the second passage, from [5].

While the two passages address the same subject—the use of MAC addresses in IP addresses—the explicit content is different. The first passage explains how the IP address is derived, while the second draws a policy inference: that exposing the MAC address raises a privacy concern. Thus, the two passages inhabit different concept spaces, making the discovery of a privacy issue in the first passage unlikely by any natural-language search process. In the case of latent semantic analysis (LSA), for example, Landauer *et al* admit that, "LSA, as currently practiced, induces its representations of the meaning of words and passages from analysis of text alone. None of its knowledge comes directly from perceptual information about the physical world…."[10] A robot would have to be told beforehand what technical issue to look for in order to know how to find it. It is through the human ability to draw inferences that go beyond literal content—to bridge the semantic gap—that the policy significance of the specification emerges.

The decision to base Slants on collaborative bookmarking stems from the need to bridge the semantic gap between the technical and public policy concept spaces. In the approach taken by Slants this conceptual bridge consists of the tag metadata attached to web resources by Delicious users.

2.2 Identity Vocabulary

In this, its initial stage, Slants focuses on a subdomain of public policy: the impact of standards on identity. Identity is, as one commentator states, "a subtle concept, but its all about your private information." In the online context, identity involves a complex of issues, including information privacy and sharing, security of information storage and transmission, authentication, access, online tracking and surveillance, datamining, profiling, third party use of identity information, identity theft, anonymous speech, and so on. This complexity makes the

[8] IETF Draft Spec, IPv6, http://tools.ietf.org/html/draft-ietf-ipngwg-trans-ethernet-03
[9] Davidson, A, J. Morris, R. Courtney.
[10] Landauer, T. K., Foltz, P.W., Laham, D, p4.

identity domain ripe for a service that keeps track of online identity resources, but it also makes a comprehensive mapping of the domain difficult.

Our intent with Slants is to provide a tool that aggregates a large amount of information in one place and makes it easy to find. While Slants is not complete in its coverage of the identity domain, it is in principle extensible and scalable. Although the tools that make it possible to easily expand the Slants thesaurus exist (particularly EVM software) they are not part of this release. Our goal for the initial version of Slants has been to build a large enough "identity thesaurus" to accommodate a broad range of tagging vocabulary.

Slants uses an identity vocabulary as the basis both for requesting tags and filtering the tags returned by RSS feeds. This vocabulary is based on research and on observation of the tags used by Delicious users to signify identity-related resources. This vocabulary contains about 1000 terms that range from broad terms such as identity, privacy, trust, and access to specific terms such as the names or numbers of standards and protocols.

3. Conceptual Design

3.1 Folksonomies

One of the defining characteristics of a folksonomy is that it imposes few rules on the tagger. As observed by Tonkin et al,[11] "The low cognitive cost of tagging items is partially resultant from the fact that no formal structure or vocabulary is involved, and in many cases those searching for objects also benefit from the idiosyncratic nature of tags, since specialised or dialect uses of terminology are represented as well. Tags have great value in serendipitous browsing -- that is, unstructured or undirected searches through which interesting objects are found, although the user may not have set out to find that particular object."

This lack of structure comes at a cost, as Tonkin *et al* continue:

> Despite their low cognitive cost, their capability of matching the user's language, and their great value in unstructured browsing, folksonomies are haunted by a number of important issues related to intrinsic language variability and imprecision: polysemy, homonymy, plurals, synonymy and basic level variation are linguistic issues which do not appear easy to solve, and which can dramatically reduce the effectiveness and benefits brought on by the very use of a tagging system .
>
> Existing social tagging applications have not been designed with information discovery and retrieval in mind. The resulting folksonomies (collections of tags) are completely uncontrolled, lacking even basic control of word forms such as spelling variants, synonyms and disambiguation of homonyms."[12]

[11] Tonkin, Emma *et al.*
[12] Golub, Koraljka *et al.*

Even when an individual user's tagging follows a personal system, these systems are as diverse as the user population. In addition, these systems are tacet and completely opaque to Slants. For example, while one user may use tags in a hierarchical fashion, as in

identity management system federated sso openid

another may use a natural language approach:

openid sso federated identity management system

Because Delicious is indifferent to the ordering of tags, the users' intent , in the first example, to express a hierarchy from general to specific and, in the second, to create a grammatical phrase, is lost. Since Slants is based on the assumption that clusters of tags on a single bookmark are semantically related with respect to the bookmarked resource and with each other, the real possibility that a Delicious user has a different agenda—or is simply being random—may impact the precision and correctness of how Slants associates tags with one another.

By automatically harvesting the results of their Delicious tagging, Slants imposes almost no additional cognitive cost on Delicious users. At the same time, Slants seeks to reduce the ambiguity of tagging by normalizing users' tags against a controlled vocabulary. In Slants, tagging is regarded as the collaborative assembly of a candidate vocabulary for describing resources within a domain.

3.2 Social Bookmarking Sites

The ranks of social bookmarking websites number in the dozens. Many of them offer RSS feeds and some provide APIs. In theory, any or all of these could have served as a platform for Slants. However, there is good reason to use Delicious.

3.2.1 Advantages of Delicious

Delicious has the advantages of scale, with over 5 million users[13] and 100 million unique bookmarks,[14] and operational stability as an owned unit of Yahoo. In addition, for simplicity's sake, we chose to work with just one of the available sources.

Nevertheless, it would be a simple matter to expand Slants to accept feeds from other bookmarking services as well as Delicious. While Delicious uses RSS 2.0, Cite-U-Like, Connotea use RSS version 1.0. However, these formats are close enough to allow parsing of the RSS file with only a few adjustments.

There are as many ways to use Delicious as there are users. While some users accumulate bookmarks for years without deleting them, others clean house periodically, in which case their bookmarks are lost to other Delicious users. Because Slants can harvest bookmarks continuously, it acts as an archiving service for bookmarks that have been subsequently deleted by Delicious users.

[13] http://en.wikipedia.org/wiki/Del.icio.us
[14] http://www.techcrunch.com/2007/09/06/exclusive-screen-shots-and-feature-overview-of-delicious-20-preview/

3.2.2 Disadvantages of Delicious

Delicious privacy settings do not allow the option of making bookmarks selectively available to a subset of delicious users. This is a problem in cases where some level of confidentiality is desired. However, users can hide their networks, and the true identity of a Delicious user need not be evident from his or her Delicious user name.

In addition, under our initial use case, it is desirable to make the Slants user network as transparent as possible, in order to allow more users to join. While it can be argued that some of these users will contribute irrelevant bookmarks, the filtering process provided by the thesaurus should in most cases prevent the addition of irrelevant bookmarks and tags.

4. How Slants Works

Slants uses PHP/CURL to programmatically request RSS feeds from Delicious.com. Delicious provides a syntax that supports a variety of targeted requests. Delicious also provides an API that enables password-protected remote interaction with individual Delicious sites. Slants uses this API to interact with its own Delicious site at http://delicious.com/slants.

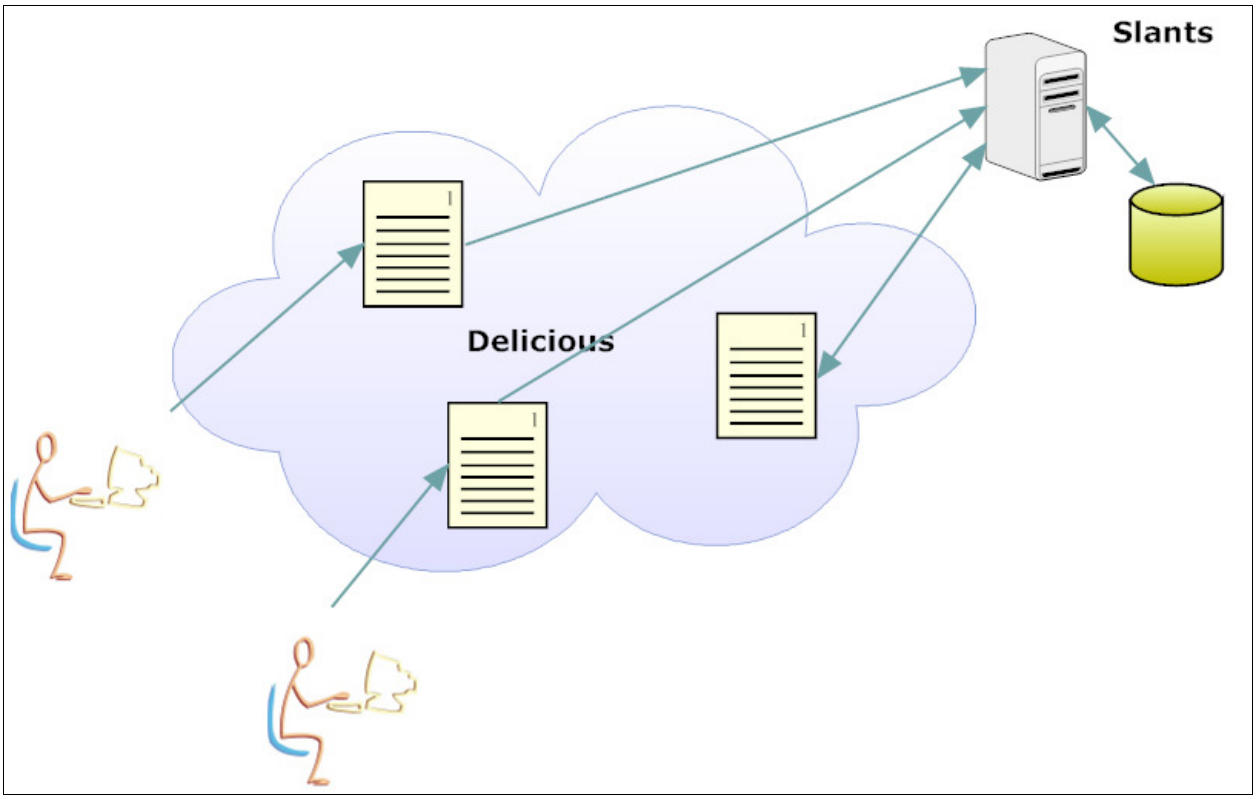

Figure 1. Slants receives RSS feeds that contain bookmarks created by Delicious users and saves them in a database. Slants also maintains its own Delicious account.

4.1 Harvesting Bookmarks

Slants receives RSS feeds from Delicious after sending a request in the form of a URI. While Delicious allows a variety of requests, Slants is mainly designed to request sets of bookmarks selected by user and tag. For example, the following request would return a feed containing all bookmarks from user JimTester that have both the *openid* and *privacy* tags:

http://feeds.delicious.com/v2/rss/JimTester/openid+privacy

Delicious then returns an RSS feed that contains information on as many bookmarks as match the request (see Figure 1). The information includes the following:

- **title .** The text in the title bar of the resource.
- **pubDate .** The date the bookmark was created.
- **guid .** A globally unique identifier for this bookmark.
- **link .** The URI of the resource.
- **dc:creator .** The Delicious user id of the bookmark creator
- **comments.** A globally unique identifier for all Delicious bookmarks of this resource.
- **wfw:commentRss.** A URI that requests an RSS feed from Delicious that contains all bookmarks for a resource.
- **source.** Identifies and links to the main Delicious page of the user who created the bookmark.
- **description.** Text describing the resource, entered by the Delicious user who created the bookmark.
- **category.** Tags assigned by the user to the bookmark.

```
<item>
  <title>LDAP.com - Commentary by Mark Wahl, CISA - Digital ID World and OpenID URLs (20070925)</title>
  <pubDate>Fri, 01 May 2009 20:36:45 +0000</pubDate>
  <guid isPermaLink="false">http://delicious.com/url/1e0ff15fc5fbd696f28dafbe8f0334dd#JimTester</guid>
  <link>http://www.ldap.com/1/commentary/wahl/20070925_02.shtml</link>
  <dc:creator><![CDATA[JimTester]]></dc:creator>
  <comments>http://delicious.com/url/1e0ff15fc5fbd696f28dafbe8f0334dd</comments>
  <wfw:commentRss>http://feeds.delicious.com/v2/rss/url/1e0ff15fc5fbd696f28dafbe8f0334dd</wfw:commentRss>
  <source url="http://feeds.delicious.com/v2/rss/JimTester">JimTester's bookmarks</source>
  <description>That OpenID require the user to enter a personally-identifying URL at arbitrary Relying Party web sites raised
privacy concerns to France Telecom. France Telecom did not auto-assign OpenIDs to their customers, as it would have exposed
their existing customers&amp;amp;#039; user identifiers (currently private to the customer-FT relationship). A customer might not
wish to be required to have their userid become exposed in an OpenID URL.</description>
  <category domain="http://delicious.com/JimTester/">OpenID</category>
  <category domain="http://delicious.com/JimTester/">privacy</category>
  <category domain="http://delicious.com/JimTester/">authentication</category>
  <category domain="http://delicious.com/JimTester/">exposure</category>
</item>
```

Figure 1. RSS output from Delicious. The most important entries are boldface. (For brevity's sake just one bookmark item is included.)

The RSS feed is parsed and the extracted values are entered into a MySQL database. In addition to the tags that match the query, any additional tags assigned by the Delicious users are also returned. These tags may be valuable in establishing additional conceptual links to a resource.

4.2 Retrieving Bookmarks

When a user accesses the Slants website, a MySQL query runs. This query populates the website with the most recent feeds. The query results are formatted in XML and transformed for display using XSLT and CSS.

An index is generated simultaneously and displays in a separate column of the web page. The index is generated by matching tags in the bookmarks with terms in a thesaurus. The thesaurus consists of a hierarchy of domain-specific topics and a controlled vocabulary. Tags find matches in the thesaurus and are thus organized for enhanced retrieval. Stemming software[15] and synonym replacement help match tags with terms. The generated index lists hierarchically all the bookmarks in the database, with additional cross-links to related topics. A keyword lookup window provides an alternative search capability.

Clicking on a category, subcategory, or leaf node sends a new query to the database, to retrieve the requested subset of bookmarks.

4.3 Enhanced Bookmark Harvesting

Slants can act like a robotic Delicious user. It can make multiple requests for feeds by iterating through a list of Delicious users as well as through multiple sets of tags. Slants can also be set to run multiple RSS requests on a recurring basis, such as once a week, in order to harvest new bookmarks. Slants is not limited to bookmarks of users in the Slants community, but can request them from any of more than five million Delicious users.

As more tags accumulate around individual resources, associations build with other resources where tag sets overlap. This improves collocation during information retrieval and creates navigation paths for browsing. Slants can harvest based on tag sets that are independent of any user or resource, thus discovering both new resources that are related to ones already discovered and other Delicious users that have the same interests as Slants community members. Finally, Slants can do expansive harvesting, substituting synonyms, different word forms, and related terms in the tag set to find bookmarks that are conceptually akin to some in the Slants database, but use a different tagging vocabulary.

5. Future Work

Future work should focus on development of the Slants thesaurus. This work includes both refining the identity management vocabulary and extending the vocabulary to include other public policy subdomains.

5.1 EVM Term Recommender

The Entry Vocabulary Module Toolkit (EVMT) was developed at the I-School in 2001 by Aitao Chen, under DARPA contract N66001-97-C-8541; AO-F477, "Search Support for Unfamiliar

[15] The program, class.stemmer.inc, implements the Porter Stemming Algorithm. It is free software available at http://www.chuggnutt.com/stemmer-source.php (Copyright (c) 2003-2007 Jon Abernathy <jon@chuggnutt.com> All rights reserved.)

Metadata Vocabularies." The principal investigator was Prof. Michael Buckland; co-principal investigators were Dr. Fredric Gey and Prof. Ray Larson. EVMT enables mapping of entry terms to catalog terms based on a probability ratio. In the case of Slants, it could calculate the probability that a term used by a Delicious tagger is relevant to a term in the Slants controlled vocabulary. EVMT is teachable: it becomes more accurate with use. This would help overcome the limitation to scalability imposed by the current need to expand the Slants thesaurus manually.

5.2 User Participation

Over time, user participation could expand Slants' scope. As more users are included in bookmark harvesting, resources within the public policy domain will become more visible.

5.2.1 User-requested searches and postings

A function that permits users to create tag sets for bookmark harvesting could take advantage of expansive searches to return results in a targeted area of interest.

5.2.2 User additions to the thesaurus

It would be desirable to allow users to add terms to the thesaurus. This would improve both the finding and selecting power of the thesaurus, provided users make good choices of terms and of the hierarchical and syndetic relationships between terms. To prevent a highly undesirable pollution of the thesaurus, logic would have to be introduced to keep the thesaurus internally consistent and a consensus process would be needed to allow other members to vote on or perhaps veto new thesaurus entries. Given these caveats, the idea of an organically growing thesaursus is appealing, since it distributes the substantial work involved in creating a thesaurus over time and over many users.

5.3 Possible Applications

Slants uses tags to classify information. More abstractly, though, the tags are simply instructions to the Slants application, and could carry other meanings besides specifying a term under which to classify the associated bookmark. This creates opportunities for additional functionality. For example, a controlled vocabulary of tags could provide other processing instructions, such as formatting, queuing, distribution, approval and expiration of a given bookmark. Imagine, for example, that one collaborator finds a story on the web about her nonprofit. She could tag it with a special "whatsnew" tag that tells the service to flag the bookmark for posting in a special "What's New" area of the organization's website and to simultaneously notify any collaborators who need to approve the posting. The bookmark could then be held in a nonpublic queuing area until it is approved, or deleted after a certain date if it is not acted on.

6. Conclusion

Folksonomies store a wealth of metadata describing resources and conceptual connections between resources. As time goes on this metadata layer will only grow richer and more complex. The challenge is to systematize this information to serve the classic purposes of information retrieval: namely to locate, identify, select, obtain, and navigate.[16] Historically, the resources required to achieve a high bibliographic standard of information retrieval have limited the size of the collection that could be catalogued, and creation of large-scale taxonomies of networks has

[16] Svenonius, E., p20.

proven daunting. The popularity and growth of folksonomies, on the other hand, is surely in part due to the low cognitive cost of tagging. The challenge here is to impose some discipline without interfering with the convenience and freedom of using any tag you want. Slants demonstrates one way of keeping that cost low while imposing some structure on the metadata: by limiting the domain, leveraging the efforts of many taggers, and imposing order on the tags after the fact.

Although Slants focuses on the public policy/ICT domain, the same approach can be applied to any relatively well-delimited domain by using a different thesaurus. This is a particularly attractive approach for organizations with limited resources. For example, small nonprofits, advocacy groups and watchdog groups have important research needs, for which they often rely on volunteers. By simply pursuing their usual research routine—tagging web pages they find interesting—volunteer researchers could collaborate in building an always up-to-date, organized body of resources.

7. References

1. Antón, A.I., He, Q., and Baumer, D. "The complexity underlying JetBlue's privacy policy violations." *IEEE Security & Privacy* (to appear).
2. Capocci, A., G. Caldarelli. "Folksonomies and clustering in the collaborative system CiteULike. http://arxiv.org/abs/0710.2835, Oct 2007; http://arxiv.org/PS_cache/arxiv/pdf/0710/0710.2835v2.pdf
3. Clark, D. D.,J. Wroclawski, K.R. Sollins, and R. Braden, "Tussle in Cyberspace: Defining Tomorrow's Internet," Proc. ACM SIGCOMM 2002, pp. 347-356.
4. Cranor, Lorrie F. and Joseph Reagle (1998). Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences Project. In Jeffrey K. MacKie-Mason and David Waterman, eds., Telephony, the Internet, and the Media. Mahwah: Lawrence Erlbaum Associates.
5. Davidson, Alan, John Morris, Rob Courtney. "Strangers in a Strange Land: Public Interest Advocacy and Internet Standards.: Washington, D.C.: Center for Democracy and Technology, 2002. <www.cdt.org/publications/piais.pdf>
6. Golub, Koraljka, Catherine Jones, Brian Matthews, Jim Moon , Marianne Lykke Nielsen, Douglas Tudhope, "Enhancing social tagging with a knowledge organization system ," UKOLN,2008,http://homes.ukoln.ac.uk/~kg249/publ/ALISSv3.pdf,
7. Heymann, P. and Hector Garcia-Molina: Collaborative Creation of Communal Hierarchical
8. Taxonomies in Social Tagging Systems. Stanford InfoLab Technical Report 2006-10. http://ilpubs.stanford.edu:8090/775/1/2006-10.pdf
9. IEEE Security and Privacy Magazine,
10. Vol. 6, No. 2, March-April 2008.
11. Kent, S., Millett, L. (Eds) (2003), Who Goes There? Authentication Through the Lens of Privacy, Committee on Authentication Technologies and Their Privacy Implications, National Research Council, National Academies Press, Washington, DC.

12. Landauer, T. K., Foltz, P.W., & Laham, D. (1998). "Introduction to Latent Semantic Analysis." *Discourse Processes*, 25, 259–284.

13. Morris, J., A. Davidson. "Public Policy Considerations for Internet Design Decisions," Internet Draft, Center for Democracy and Technology, June 2003, http://www.cdt.org/standards/draft-morris-policy-considerations-00.txt.

14. Morris, John, Alan Davidson, 2003. "Policy Impact Assessments: Considering the Public Interest in Internet Standards Development." Washington, D.C., 2003: Center for Democracy and Technology. www.cdt.org/publications/pia.pdf

15. Noveck, B. S. (2005), 'A democracy of groups'; First Monday, 10(11). http://www.firstmonday.org/issues/issue10_11/noveck/index.html (accessed 11 November 2005)

16. Reagle, Joseph M., Jr, "Social Protocols: An Introduction," Internet Draft, http://www.w3.org/Talks/980922-MIT6805/SocialProtocols.html

17. Svenonius, E. *The Intellectual Foundation of Information Organization* (MIT Press, Cambridge, MA, 2000).

18. Tonkin, Emma; Corrado, Edward M.; Moulasion, Heather Lea; Kipp, Margaret E. I.; Resmini, Andrea; Pfeiffer, Heather D.; Zhang, Quiping. (2008). Collaborative and social tagging networks. Ariadne 54. Retrieved March 19, 2008, from http://www.ariadne.ac.uk/issue54/tonkin et-al/.

19. Van Damme, C., M. Hepp, and K. Siorpaes (2007). Folksontology: An integrated approach for turning folksonomies into ontologies. In Bridging the Gep between Semantic Web and Web 2.0 (SemNet 2007), pp. 57-70, http://celinevandamme.com/vandammeheppsiorpaes-folksontology-semnet2007-crc.pdf.