



Studying Social Tagging and Folksonomy: A Review and Framework

Item type	Journal Article (On-line/Unpaginated)
Authors	Trant, Jennifer
Citation	Studying Social Tagging and Folksonomy: A Review and Framework 2009-01, 10(1) Journal of Digital Information
Journal	Journal of Digital Information
Downloaded	31-Oct-2016 03:34:34
Link to item	http://hdl.handle.net/10150/105375

Studying Social Tagging and Folksonomy: A Review and Framework

J. Trant, University of Toronto / Archives & Museum Informatics

158 Lee Ave, Toronto, ON Canada M4E 2P3 jtrant [at] archivumuse.com

Abstract

This paper reviews research into social tagging and folksonomy (as reflected in about 180 sources published through December 2007). Methods of researching the contribution of social tagging and folksonomy are described, and outstanding research questions are presented. This is a new area of research, where theoretical perspectives and relevant research methods are only now being defined. This paper provides a framework for the study of folksonomy, tagging and social tagging systems. Three broad approaches are identified, focusing first, on the *folksonomy* itself (and the role of tags in indexing and retrieval); secondly, on *tagging* (and the behaviour of users); and thirdly, on the nature of *social tagging systems* (as socio-technical frameworks).

Keywords: Social tagging, folksonomy, tagging, literature review, research review

I. Introduction

User-generated keywords – tags – have been suggested as a lightweight way of enhancing descriptions of on-line information resources, and improving their access through broader indexing. “Social Tagging” refers to the practice of publicly labeling or categorizing resources in a shared, on-line environment. The resulting assemblage of tags form a “folksonomy”: a conflation of the worlds ‘folk’ and ‘taxonomy’ used to refer to an informal, organic assemblage of related terminology (Vander Wal, 2005b). When shared with others, or viewed in the context of what others have tagged, these collections of resource identifiers, tags and people begin to take on additional value through network effects. Searching tags can enable the discovery of relevant resources, and the social relationships that develop among taggers become a means of information discovery in and of themselves (Marlow et al., 2006a). Those who actively use Web-based resources, such as on-line databases of scientific articles, developed tools to enable the creation of personal collections of ‘bookmarks’ or

pointers to networked resources, that are described or ‘tagged’ with words that identify and categorize them (Mathes, 2004; Quintarelli, 2005) – see for example, del.icio.us (Golder & Huberman, 2005), Connotea (Hammond et al., 2005; Lund et al., 2005) and PenTags (Pennsylvania State University Library, 2005-).

This paper reviews research into social tagging and folksonomic description (identifying 180 sources from before December 2007). Methods of researching the contribution of social tagging and folksonomy are described, and outstanding research questions are presented. An examination of the nature of on-line searching, the changing context for information retrieval, and efforts to identify alternative sources for indexing vocabulary, sets the stage for this discussion of the emerging literature of social tagging and folksonomy. This is a new area of research, where theoretical perspectives and relevant research methods are only now being defined. This paper provides a framework for both the study of social tagging and folksonomy and the analysis of their contribution to the on-line information landscape.

2. Social Tagging and Folksonomy

2.1 Context

Within the “Web 2.0” focus on user-driven design and social participation (O'Reilly, 2005), user tagging – enabling user contribution of keywords for personal or shared information organization and retrieval – has attracted attention in the library, museum and archives community (Bearman & Trant, 2005; Chun & Jenkins, 2005; Miller, 2006; Styles, 2006), as well as in education more broadly (Alexander, 2006; P. Anderson, 2007). Tagging has been seen as a possible solution to improved searching of networked resources, as well as a means to support their personalized use.

Interest in tagging and folksonomy derives in part from the information discovery needs of the ‘long tail’ (C. Anderson, 2004). Anderson re-assessed concepts of value in on-line information and posited that the ‘shelf-life’ of digital content is much longer and more attenuated than that of physical content. This long life span, when combined with the potential to reach micro-markets without significant incremental costs, changes the economics of content creation and distribution. Akst (2005) notes that the availability of microcontent changes our approach to value. Access at the data-level to a broad range of content enables the construction of a large number of secondary service applications; there is value both in the service layer and in the “long tail” of the content itself. (Kilkki (2007) has proposed a model for analysing long tails.)

Morville (2005) posits that within an environment with an excess of content, findability becomes a critical value. Personal re-discovery, finding the “stuff I’ve seen” is a major

challenge. Before the emergence of tagging applications, Dumais et al. (2003) describe a system to support personal re-discovery that recognises that much work with information involves re-visiting materials previously consulted. They note “studies have shown that 58-81% of Web pages accessed were re-visits to pages previously seen”. Improving personal information recall becomes critical in this environment of abundance.

Theories of indexing and retrieval have been constructed in the context of scarcity of information resources. Mediated searching, most often by professional library staff, meant systems could assume an in-depth knowledge of complex record structure. Precision was essential – and preferable over recall – when pricing structures for the use of research networks were based on a fee-per-search (as they were in those run by the Research Libraries Group [RLG] in the 1980s) and on the number of records returned. But end-user searching is now far more common than professionally-mediated searching, and browsing large results sets is an accepted form of query refinement: one that is expected in art history (where looking is a professional skill) and is becoming more common in image searching generally (Cunningham & Masoodian, 2006).

Success with keyword access to textual sources (as in Google searches) has led to an exploration of alternative methods of generating access points in the field of textual indexing. These could be within the resource itself; for example, Voorbij (1998) compares subject index terms to keywords found in subjects, in Dutch humanities and social sciences records. Ansari (2005) found a greater than 70% match between title keywords and assigned descriptors in study based on medical theses in Farsi. Frost (1989) looked at title words compared to Library of Congress Subject Headings assigned by cataloguers. Gil-Leiva and Arroyo (2007) looked at author keywords compared to professional index terms. Image captions would seem to be a logical place to find additional image information that could support retrieval, even in cross-language/multi-lingual environments (Smeaton & Quigley, 1996). All these studies point to the possibility to enhance professional indexing with materials from other sources

User assignment of keywords has been proposed as a way to improve image retrieval on the Web. Von Ahn and Dabbish (2004) introduce the ESP Game [<http://www.espgame.org>] an application designed to encourage the labeling (indexing or tagging) of images through a gaming interface in order to improve the searchability of the Web. The system is described: in a time-limited environment, two users are given the same image at the same time and enter words until they match, then their performance is scored and they move on to the next image. The goal is to get as many matches as quickly as possible. Taboo words (not accepted towards the match) were introduced to ensure that the words assigned to the images continued to vary. Evaluation shows that “people will want to play the ESP game and that the labels it produces are useful”. Players were prolific. In the first four months (to

December 10) “almost 1.3 million labels were collected with only 13,630 players. Tags submitted were evaluated: “All (100%) of the images retrieved [during the test] made sense with respect to the test labels”. Tags were also compared to the results of in-lab indexing tests, where 15 participants were asked to provide 5 labels for each of 20 random images. “For all (100%) of the 20 images, at least 5 (83%) of the 6 labels produced by the game were covered by the participants (i.e., each of these labels was entered by at least one participant). Moreover, for all (100%) of the images, the three most common words entered by participants were contained among the labels produced by the game.” A manual assessment of the labels was also made, where 15 participants were shown the first 6 words applied in the game to each of 20 images, and asked “how many of the words above would you use in describing this image to someone who couldn't see it” and “how many of the words have nothing to do with the image”. “5.105 words or (85%) of the words for each image would be useful in describing it”; 105, or (1.7%) of the words were felt not to relate to the image at all. The authors feel that using labels for images will significantly improve image search on the Web, which at the moment is skewed by text-based systems that depend on annotation, file names and context. The ESPgame technology now underpins the Google Image Labeller (Google, 2007a), part of Google Image Search (Google, 2007b). The success of the ESPgame was pivotal in establishing the museum community's interest in tagging (Bearman & Trant, 2005).

2.2 Defining Tagging and Folksonomy

A number of different ways to approach the study of user tagging and folksonomy are emerging in a rapidly developing, but young literature. (This study reviews materials published through the end of 2007.) Authors interested in “tagging” as an activity focus attention on the user's role in linking terms to resources. Others, interested in the vocabulary that results, direct their study to “folksonomy” – the collective assemblage of tags assigned by many users. Still others profile the social and technological context within which tagging takes place and folksonomies are constructed. We can think of *tagging* as a process (with a focus on user choice of terminology); of *folksonomy* as the resulting collective vocabulary (with a focus on knowledge organization); and of *social tagging* as a socio-technical context within which tagging takes place (with a focus on social computing and networks).

Other terms have also been used to describe this phenomenon, including ‘community cataloguing,’ ‘cataloguing by crowd’, (Chun & Jenkins, 2005), social classification (Furner & Tennis, 2006; Landbeck, 2007; G. Smith, 2004; Trant, 2006), and ethnoclassification (boyd, 2005; Merholz, 2004a, 2004b; Walker, 2005). The definition of ‘folksonomy’ itself is not firm, with discussions focusing on required elements of sociability, collaboration and

cooperation (Vander Wal, 2005b; Weinberger, 2006c). The Wikipedia “Talk” page for “Folksonomy” reveals this disagreement ([Collective], 2007).

There is a tendency in the relatively young literature of social tagging and folksonomy, to conflate concepts, and to equate *social bookmarking* (systems that enable the sharing of tagged links to Web resources) with tagging and folksonomy. For example Noruzi (2006) discusses folksonomy and social bookmarking as if they were synonymous.

2.3 Framing the issues

Discussion of social tagging and folksonomy emerged on the Web, in late 2004 and 2005, in papers and blog postings (particularly from the information architecture and social computing communities). Early discussions focused on naming and describing the phenomenon, and positioning it in the changing Web 2.0 landscape. The rapidly developing literature – all from the past 3 years – clusters into three broad approaches to the study of social tagging and folksonomy, focused first, on the *folksonomy* itself (and the role of tags in indexing and retrieval); secondly, on *tagging* (and the behaviour of users); and thirdly, on the nature of *social tagging systems* (as socio-technical frameworks). Three early, and influential papers illustrate each of these approaches, and are discussed here in detail to frame the studies that follow.

Mathes (2004, a widely-cited paper originally written for a LIS course at the University of Illinois Urbana-Champaign) provides the first review and survey of social tagging / folksonomy systems. He describes the functionality of these systems, compares user tagging behaviour in Flickr (<http://www.flickr.com>) and del.icio.us (<http://del.icio.us>) and notes the following limitations of user assigned tags, seen as “inherent in an uncontrolled vocabulary”: ambiguity, inconsistent orthography, and synonymy. Mathes identifies the following strengths: serendipitous discovery through browsing; and tracking desire lines. He asserts that folksonomies “work” because of low barriers to entry; low cognitive costs; a tight feedback loop; and the ability to share tags. Mathes cites the following areas for future research: quantitative tag analysis, qualitative user analysis and applicability to other systems. Mathes’ focus is on *language (tagging vocabulary – i.e. folksonomy – as indexing)* and the possible role of folksonomy in *information retrieval*.

Golder and Huberman (2005; published as Golder & Huberman, 2006) provide a description of the structure of collaborative tagging systems (based on del.icio.us) and analyse the dynamics of their use, finding “regularities in user activity, tag frequencies, kinds of tags used and bursts of popularity in bookmarking, as well as remarkable stability in the relative proportions of tags within a given url.” They define “collaborative tagging” as “the practice of allowing anyone – especially consumers – to freely attach keywords or tags to

content”, though the process they describe is really social or collective, not actively collaborative, and note that “tagging is most useful when there is nobody in the “librarian” role or there is simply too much content for a single authority to classify”. Following a review of semantic and cognitive aspects of classification that explores polysemy, synonymy and “basic level variation” (i.e. the level of description used in a hierarchical system), the authors assert that “tagging is fundamentally about sensemaking”. Golder and Huberman’s observation that “sensemaking is a retrospective process” positions bookmarking and tagging in the process of re-discovery (finding things that you’ve found before), and is used to explain an observed refinement in tags over time. They note that as well as being descriptive, tags perform functions: “Identifying What (or Who) it is about”; “Identifying What it Is”, “Identifying Who Owns It”, “Refining Categories”, “Identifying Qualities or Characteristics”, “Self Reference”, “Task Organizing” (though it is noted that things can fit in more than one category). Tags were assigned in stable patterns. After a relatively small number of bookmarks [fewer than 100], a nascent consensus seems to form, that is not affected by the addition of further tags. Stabilization of terminology is encouraged by the way del.icio.us interface suggests commonly used tags, facilitating their re-use. Convergence does not negate minority views (the majority simply adds to the weight of existing tags; the minority continues to co-exist). Golder and Huberman harvested tags from del.icio.us for their quantitative analysis, focused primarily on *user behaviour when tagging*.

Marlow, Naaman, boyd, & Davis (2006a, 2006b) focus on the social aspect of tagging and present a model in which “tags are represented as typed edges connecting users and resources” (Marlow et al., 2006a, 1). Though they note that “system design and attributes” and “user incentives” also are significant variables, this group sees relationships between users as a critical element of tagging systems. Characteristics of tagging systems include: tagging rights (who can tag what); tagging support (whether or not you see other tags, or if tags are suggested); aggregation (duplicate tags for the same resource); type of object (or what is tagged); source of material (from participants, system or Web); resource connectivity (links using tags or not); and social connectivity (links between users). User incentives are divided into organizational and social motivations, and include future retrieval, contribution and sharing, attracting attention, play and competition, self-presentation, and opinion expression. Marlow et al. also present a case study of Flickr (chosen to contrast with Golder and Huberman’s analysis of del.icio.us (Golder & Huberman, 2006)), and propose measures such as distribution of distinct tag collections across users; correlations between the number of tags, photos and contacts of a particular user; distinct tags over time; and vocabulary overlap across random users, as ways to characterize tagging. This group offers a typology that focuses on *the socio-technical aspects of tagging systems*.

The popular press embraced tagging and folksonomy in 2005. In the *New York Times* Todras-Whitehill (2005) used tagging as a key element in defining three genres of

networking sites. In *Wired*, Andrews provided a “how-to” focused piece (Andrews, 2005b). Pink included folksonomies as part of the *New York Times* “The Year in Ideas” for 2005, and mentioned the Guggenheim and the Metropolitan Museum of Art’s exploration of the concept (Pink, 2005). Smith summarized tagging activity in 2005, as the year “tags went mainstream”, including Yahoo!’s acquisition of both Flickr and del.icio.us (G. Smith, 2005b).

These early studies – and the resulting popular buzz – set the stage for a flurry of activity studying tagging, folksonomy, and social tagging systems. These studies are reviewed in the sections that follow. Some may be referenced more than once, as their approaches are hybrid, and they address multiple aspects of these complex systems.

2.4 Folksonomy and Vocabulary: Studying Tags

Groups interested in on-line information retrieval have studied folksonomy from the perspective of its ability to support information retrieval. Folksonomy is often criticized because tags are not drawn from a controlled vocabulary. The aggregated terminology drawn from tagging is expected to be inherently inconsistent, and therefore flawed, according to theories of indexing. The strand of literature that focuses on folksonomy and information retrieval explores this premise.

2.4.1 General studies

Merholz (2004b) celebrates the user-focus of tagging. His metaphor, paths that are worn in the grass, where people walk rather than where the landscape designer wanted them to go, is likened to Susan Leigh Star’s “ethnoclassification” (Star, 1996). But, he notes tagging’s drawbacks: synonymy and inaccuracy. Merholz reasserts the value of established knowledge in classification and retrieval (2005) emphasizing the contribution of the traditions of classification and vocabulary control, and the value of hierarchy for disambiguation.

Quintarelli (2005) defines folksonomy as “user-generated classification, emerging through bottom-up consensus”. Virtue is found in the “bottom-up distributed and collaborative grassroots approach”, where “relationships naturally emerge”. Meaning comes from the aggregation of tags and their clustering. Quintarelli’s overview – focused on the nature of vocabulary – presents the following drawbacks of folksonomy-driven information retrieval: “a lack of precision”, a lack of hierarchy, a low findability quotient, and problems with scaling. These are contrasted with benefits including the fact that folksonomies “reflect the population’s conceptual model” and are capable of “matching users’ real needs and language”, in part because they are inclusive. Quintarelli’s sense of inevitability, that folksonomies are a “forced move” (citing Shirky (2005a)), is based in part on a sense that controlled vocabularies are “not practically and economically extensive”. A summary of ‘side

benefits' includes "possibilities of aggregation and analysis" and the "creation of communities" in the context of an emerging "metadata ecology" – the co-existence of controlled vocabulary and folksonomy advocated by Rosenfeld (2005).

Kroski (2005) also outlines the pros and cons of folksonomic descriptive strategies, based on a review of social sites like del.icio.us, 43 Things, Flickr and Technorati. Her pros all relate to the "Wisdom of Crowds" and include that folksonomies: "are inclusive", "are current", "offer discovery", "are non-binary", "are democratic and self-moderating", "follow desire lines", "offer insight into user behavior", "engender community", "offer a low cost alternative", and "offer usability". She sums up with "resistance is futile", citing the weight of on-line information. Cons include "no synonym control", "lack of precision", "lack of hierarchy", "the 'basic level' problem", "lack of recall", and that – like any social system – folksonomies "are susceptible to gaming".

Peterson (2006) argues that folksonomy's relativism means that it is inherently flawed, and will always fail to produce the accuracy of formal classification (drawn from Aristotelian categories), to the detriment of the user. Weinberger, however, argues for the value of relativism, and the need to accommodate inaccuracy and inconsistency in knowledge representation systems (Weinberger, 2006a). He views tagging positively, for its ability to incorporate multiple perspectives, reflect emerging concepts, enable the emergence of social groups through shared semantics, and accommodate diversity (Weinberger, 2005b), all ideas explored later in his book-length study (Weinberger, 2007).

2.4.2 Vocabulary control and evolution

Theories of self-normalization assert that folksonomic tags will self-regulate – that the collective vocabulary will become more consistent over time, without external imposition of controls. In a reaction to the constraints of controlled vocabulary, Shirky (2004) asserts that synonym control reduces nuance and sacrifices meaning, that hierarchy is often forced and false, and that poly-hierarchy is essential to understanding the multi-faceted nature of meaning. While admitting that "lack of precision is a problem" with folksonomic terminology, he attributes this to user-behaviour, rather than to the nature of folksonomy itself, and predicts that tags will become self-normalizing.

Changes in popular usage are reflected in the terms used in tagging systems. Speroni di Fenizio (2005) reports an analysis of del.icio.us tags for Shirky's "Power Laws, Weblogs, and Inequality" (Shirky, 2003) and illustrates the shift in tagging vocabulary following the introduction of the phrase 'long tail'. Smith is skeptical about the success of popular tags, citing of the lack of control over synonymy, poor precision in single-word tags, and missing hierarchy (G. Smith, 2004).

Udell illustrates the theory of stabilization of vocabulary around common choices in tags in an interactive exploration, showing how vocabulary evolution operates within an individual's tag set in del.icio.us (Udell, 2005). This stabilization of vocabulary was confirmed by Millen and Feinberg's study of dogear at IBM: "there is a gradual decline in new tags as end-users enter more bookmarks", determined by calculating the percent of new tags as a function of existing tags, and "considerable tag re-use" (Millen & Feinberg, 2006, 6).

Weinberger also explores the pros and cons of the normalization of language promoted by shared information retrieval spaces (such as his example of the choice by CNN to use 'tsunami' instead of 'tidal wave'). He comes to no conclusion, except that things are grey and people and language will evolve (Weinberger, 2005a). Weinberger later notes that folksonomies are an anti-authoritarian symbol, that embrace a multiplicity of world views (Weinberger, 2006b). They harness the power of 'emergence' and 'excess', and represent a rejection of reductivist 'essentialism'. This anti-authoritarianism, and inclusivity, is seen by Weinberger to have inherent benefit.

In an approach similar to Weinberger, Porter contrasts 'discovery' with 'finding,' noting that "users who discover information didn't need to know it was there to begin with, and so couldn't have been trying to find it. In a word: serendipity." Controlled vocabulary marginalizes those who do not use the same words (Porter, 2005a, 2007). This difference in tagging vocabularies is present between teachers and students (Bateman et al., 2007), where tagging terms reflect mastery of a specialized terminology.

Macgregor and McCulloch (2006) contrast folksonomies and controlled vocabularies and conclude that "the need for lexical control, hierarchical structure and associated coding is essential for attaining meaningful subject interoperability across distributed systems (perhaps using different dialects or languages), as well as maintaining the efficacy of subject searching on local systems." They feel that public relations might solve users' difficulties with controlled vocabularies. Guy and Tonkin (2006) also posit that training users might provide more consistency in tagging, removing inconsistencies in terminology, including synonymy (same sense/different words) and polysemy (same word/different senses), all seen barriers to retrieval. This is similar to the approach of Ali Mejias (2005), and sounds a lot like the Training that Markey (2007a) said users didn't want.

Shirky asserts that formal classification systems are a bad fit when applied to electronic resources. Classification works well, he maintains, when the domain to be organized has a "small corpus, formal categories, stable entities, restricted entities and clear edges", with participants who are "expert cataloguers, authoritative source of judgment, coordinated users, expert users". So in the large undefined corpus of the Web, populated with naive users, there is a problem: "The only group that can categorize everything is everybody",

working together in an organic, evolving way (Shirky, 2005b). He feels there are false equivalencies drawn in discussions of tagging; tags can be multi-dimensional, and are used that way – in combination with each other and with user data (Shirky, 2005c).

It is unclear how much system design influences tagging behaviour and tag vocabulary choice. What happens when systems prompt for terms from source data, from terms used by a particular user, or from terms used by others about this resource, e.g. the del.icio.us recommender? Does this limit expression, as asserted by Weinberger (2005b)? Sood et al. (2007) tested this with a system that provides tag suggestions for new blog posts, building on the work of Mishne (Mishne, 2006). Xu et al. (2006) propose a method for collaborative tag suggestion that takes into account the reputation of users and a number of criteria for good tags, including high coverage, high popularity, least effort, and uniformity (in format) for use in Yahoo!'s myWeb2.0. Sen et al., present a mathematical model of tagging based on personal tendency and community influence on personal tagging behaviour, within the context of the MovieLens tagging system, and show how prior exposure to a tag influences tag selection, towards consistency (Sen et al., 2006).

There are a number of questions still to be explored in the relationship between folksonomies and controlled vocabularies. It has not yet been determined how popular and expert vocabularies differ and where each can be employed to most effect. Vocabularies evolve over time, and study is needed to determine if tagging and folksonomy could provide clues to vocabulary evolution. For example, tagging might offer a straightforward way to collect alternate terms to include in an existing thesaurus. Tags might reflect emerging concepts that are of interest in the enterprise, and serve to capture employees' knowledge. Tagging could enable the development of a corporate folksonomy that could be integrated into an enterprise-wide vocabulary (Hayman, 2007, 19).

2.4.3 Vocabulary analysis

A folksonomy can be studied in relationship to other indexing vocabularies to identify the contribution of social tagging. Lin et al. (2006) describe "emerging characteristics of social classification" through three cases studies, of Connotea, Flickr, and del.icio.us, looking at tag distribution, categories of tags, and the relationship between tags and index terms.

Kipp compares the vocabulary of users, authors and cataloguers. She first analysed tags, author keywords and professionally supplied descriptors for 176 entries from citeulike.org. CiteUlike tags, author keywords and indexer-supplied terms were compared on a 7-point scale. New vocabulary emerged that included 'time management tags', and geographic descriptors (supplied only by indexers) (Kipp, 2006a, 2006b). In a further study, Kipp (2007b) again examined author, intermediary and tagging vocabularies (building on Kipp,

2006b). This time she focusing on biology articles tagged in CiteUlike and catalogued in PubMed. Results were similar to the earlier study: taggers' terminology differs from that of authors and indexer, though it may be related (as alternate forms of terms). The scholarly community using the CiteUlike tagging tool revealed an interest in methodology that was not reflected in the traditional cataloguing; these methods-focused tags made a distinct contribution to the description of biology articles. A similar result was found in a CiteUlike study focusing on health information (Kipp, 2007a).

Berendt and Hanser (2007) report on a study that examines the tags assigned to blogs in relation to domains, parts of speech (using the Stanford POS) tool, and sense (using WordNet) to see what can be learned about the semantics of tags and the behaviour of individual taggers. Smith's small study (2007) explores the relationship between folksonomy and subject analysis in a study of tags assigned in LibraryThing and the subject headings assigned by the Library of Congress Catalog. It appeared, in her small sample, that LibraryThing tags were better than subject headings at identifying identified latent subjects (ones that did not meet the coverage rules for library subject cataloguing).

2.4.4 Finding structure in folksonomy

Kipp and Campbell (2006) used co-word analysis to elicit patterns from tags used in the social bookmarking service del.icio.us. They harvested del.icio.us tags for 64 popular bookmarks (58,728 posts, with a range of 53-5172 posts per URL, and 49-13,809 tags per URL, for a total of 165,831 tags). A plot of the distribution of tags across URLs had a very long tail: 30% of tags occurred only once in the sample. However, the slope of the power-law curve was gentler than expected in some cases. Tags per user per URL were modest: "65% of users assigned between 1 and 3 tags to a URL". Co-word frequencies were plotted using multi-dimensional scaling to graphically illustrate the relationships between tags. Clusters of terms did not include all of the terms for a concept as might be expected. This may reveal different user groups, with distinct vocabularies. A group of popular tags, such as 'toread', were functionally related, making them distinct from traditional classification terminology.

Cattuto et al. "adopt a tag-centric view of the system, ... [investigating] the evolving relationship between a given tag and the set of tags that co-occur with it" using the approach of 'Semiotic Dynamics'. They explore ways to describe the exponent of the power law curve [a means of describing the steepness of the curve], expecting "a shallower behaviour for tags co-occurring with generic tags (e.g. *blog*) and a steeper behaviour for semantically narrow tags (e.g. *ajax*)" (Cattuto et al., 2006).

Smith (2005a) provides a record of a tagging application that includes the ability to tag tags in relation to each other, to make it possible to identify tags as synonyms and identify one term as 'preferred'. This allows a more formal structure to be defined within a folksonomy. Disambiguation remains a challenge. De Smet and Moens (2007) demonstrate the automatic creation of a “text-based topic hierarchy” from a document corpus focused on a specific geographic area, with a known dialect, which they describe as a “city-based folksonomy”. Tag disambiguation might be enabled by prompting users with particular facets, in which to situate their tags (Bar-Ilan et al., 2006). But it has not been established that this improves the resulting folksonomy.

Heyman and Garcia-Molina (2006) present an algorithm that automatically builds a hierarchy of tags from data in a tagging system. They propose that *agreement* between users on appropriate tags; *density*, “the frequency with which users annotate objects” [expressed as annotated objects over all objects]; and *overlap*, the “frequency with which users are annotating the same objects as one another” [expressed as shared annotated objects over annotated objects] are salient features of tagging vocabularies. Graph centrality [a concept from social network theory] is used to expose “importance” when latent hierarchies derived from tags are translated into a similarity graph.

Begelman, Keller and Smadja (2006) explore the use of tag clustering techniques, proposing an algorithm that assigns similarity based on co-occurrence, and groups tags based on ‘spectral clustering’. They note that clustering is time-sensitive (co-occurrences are temporal) and community sensitive (meaning varies in different contexts).

Tonkin (2006b) suggests a method for pre-processing tags to form a keyword index that includes possible ways of parsing compound tags, and addressing divergences of orthography, this is just a first step towards extracting value from tags. Fokker et al. suggest ways to use tags to enhance navigation of Wikipedia, using automatically extracted tags to bootstrap user tagging in a P2P environment (Fokker et al., 2006). Iturrioz et al. (2007) explore the problem of fragmented tag spaces (individual’s tags spread across various applications) by proposing a desktop-based tag aggregator. Peters and Stock (Stock & Peters, 2007) propose natural language processing methods to identify entities referenced by tags that could be useful for retrieval. Quintarelli et al. (2007) posit that faceted tag structures will improve tag-based retrieval.

2.4.5 Tag spaces as ontologies

Semantic Web developers are interested in folksonomies as emergent ontologies. Mika presents an “Actor-Concept-Instance” model of ontologies, and explores it within the context of social networks. In this model, tags are 'hyperedges' that represent “the

commitment of a user” that a resource and a concept are related. Social Context is critical because it makes it possible to study meaning in its “emergence from user actions.” ... “ontologies are us: inseparable from the context of the community in which they are created and used” (Mika, 2007).

Ohmukai et al. (2005) report a design for a system that enables users to position their tags [and bookmarks] alongside those of others in their personal network, enabling the construction of a community-based ontology, and include a proposed RDF expression of the model. Beckett (2006) offers a method for using wiki-space to disambiguate and structure tags into meaningful semantics. Passant (2007) proposes a way users can link tags to an expressed ontology, in use in a defined corporate sphere. In the SemKey system, Marchetti et al. (2007) propose a method for doing much of this in an automated fashion, with the assistance of relationships between terms made explicit in WordNet and Wikipedia.

In a more theoretical study of emergent ontologies, Dix et al. (2006) note that “collaborative tagging systems are an alternative mechanism to the semantic Web approach where experts build ontologies” (p.1). A “*semantic halo*” of meaning is constructed around a tag by automated querying. Relationships among tags, including broader, narrower and related terms, are computed, taking into account temporal shifts in definition. They propose an algorithm that takes into account ‘aggregation’ (all tags linked together by co-occurrence), ‘abstraction’ (generalized and specialized related tags), ‘ambience’ (context) and ‘age’ (ordered ambience over time). A small-scale test made with del.icio.us data is reported.

Halpin et al. (2007) propose to extract meaningful ontologies by focusing on the ‘short head’ rather than the long tail, looking at commonalities in most-used tags, combined with measures of stability and information value (the measure of a tag based on the number of pages it retrieves). Their model also takes into account the essential feedback cycle in tagging (Halpin et al., 2007, 220). Schmitz (2006) proposes a statistically-based method for inducing semantic structure from tags in Flickr that is also based on similarities, but that allows that clusters share some proportion of a set of characteristics (polythetic), but not necessarily all the same features (monothetic).

It remains to be determined if tagging and folksonomy enable the discovery of latent ontology. Methods for manually exposing ontology – or inducing it technically – need to be studied further.

2.4.6 Economics and scale

The low costs of user tagging grounds the economically deterministic view that folksonomy will prevail over structured metadata (Shirky, 2005a), particularly in an era of “mass amateurization” (Shirky, 2002). But Davis notes that savings in collective classification may be illusory. Costs may not be lowered, but only shifted, as efficiencies in one area of the system are made at the expense of costs in another: “Tagging bulldozes the cost of classification and piles it onto the price of discovery” (Davis, 2005). Costs can only be determined by examining the entire complex system.

2.4.7 Other approaches to folksonomy

There are other theoretical approaches to studying tagging vocabulary. Tennis looks at folksonomy within the context of theories of classification. Specifically, he distinguishes library classification (focused on organization by subject and physical accessibility), boundary infrastructures (frameworks that connect disparate areas of an organization (Bowker & Star, 1999)) and social tagging (focused on personal information management), and explores the functions of each (Tennis, 2006). In contrast, Voss asserts that tagging simply represents ‘manual indexing’ (Voss, 2007).

Tonkin puts tagging in an historical context, looking at other ways to manage personal collections of documents, and contrasting approaches based on local file systems, and those that are internet-based (Tonkin, 2006a). She summarises research into fields as diverse as indexing, language games, discourse communities, code switching, hermeneutics, indexing and situationalism, annotation, and file [and filing] systems, noting a continuum of interests rather than an opposition (Tonkin, 2007).

Van Setten et al. view “annotations” – very broadly defined as any assertions or metadata about a digital resource (including tags) – as opinion, requiring a knowledge of their authorship for assessment. Based on case studies (that look at problems in existing systems), they argue for annotation on economic and coverage grounds, but note that “knowing who (or what) provided the annotations can help people in determining the relevance of the annotations for their own goals” (p. 9) (Van Setten et al., 2006). Berendt and Hanser (2007) assert that in the blogosphere the tags an author attaches to a post should be considered as ‘just more content’ because their choice is an assertion of the relevance of a post to a particular area of interest. Feinberg (2006) looks at authorship and authority – and particularly the choice of the form of tags – in the context of different goals of tagging (indexing personal collections, sharing personal collections, and the collective creation of merged collections).

2.5 User Behaviour: Studying Tagging

The second approach to studying social tagging and folksonomy focuses on user behaviour in tagging systems. These researchers focus on what users do when they tag. Again, these studies are early, and primarily report basic descriptive statistics. Since tagging systems are relatively new, understanding what users are doing with them help develop an appreciation of their contribution to the digital information landscape. As an example of this type of study, Gospodnetić (2005a, 2005b, 2005c, 2005d, 2005e) reports on tagging in Simpy (a social bookmarking tool). He examines tags, noting that “98% of all tags are single-word tags”, and that the large majority of bookmarks imported from browsers are not tagged (Gospodnetić, 2005b). Not all tagging is public. He notes “7.3 % of tags in Simpy are private links” [this was at a time when del.icio.us didn't have an option not to share tags]. Private links have a higher average number of tags: public 4.94, private 6.63 (Gospodnetić, 2005c). He finds that analysing tagging data by time is problematic, as not all users have locations in their profiles, and imputing time of tagging from home location is a problematic. Impressionistic findings point to evenly distributed additions to the system, and searches clustered later in the day (Gospodnetić, 2005d, 2005e). In dogear (IBM's internal tagging system), usage studies (of 16,577 unique tags generated for over 50,000 bookmarks, by 2,579 users between July 2005-March 2006) show an average of 2 tags per bookmark, with “71% of bookmarks having three or fewer tags” (Millen & Feinberg, 2006, 6). These results compare to the framing studies of del.icio.us and Flickr (Golder & Huberman, 2006; Marlow et al., 2006a).

Despite the difficulties in precisely defining the time of tagging, models need to take into account the temporal aspect of tagging systems' use. Both user interests and tag frequency shift over time. Kelkar et al. propose a model that considers tagging within the context of activity, and posit ways of identifying trends in interests and communities by looking both “activity by users across different tags and activity in tags across different users”, considering “intensity” (activity level) and “spread” (the range of tagging activity) (Kelkar et al., 2007, 1).

Vander Wal proposes to study social tagging systems by defining Broad and Narrow Folksonomies. “Broad” folksonomies (such as del.icio.us) result when many people tag the same item: analysing tags reveals a power law distribution, tapering to a 'long tail' of items in which only a few people exhibit interest. “Narrow” folksonomies (like Flickr) result when only one person (or a few people) tags an object, usually one they themselves created. In this narrow, individually-defined context it may be more difficult to determine meaning in relationships between tags, because language may be personal (Vander Wal, 2005a, 2005c).

2.5.1 User Motivation

The motivations for tagging or categorizing can be approached in terms of cognitive science. Sensemaking organizes change, by identifying variances and labeling them; “categories have plasticity because they are socially defined, because they have to be adapted to local circumstances” (Weick et al., 2005). Tagging can be considered as an act of sensemaking, with shared tags becoming a form of collective meaning. Users tag to make sense of – or organize – a part of their world.

The immediate payback of personal information management is one of the prime motivations for investing in social tagging systems (Vander Wal, 2005d). Re-discovery and self-organization offer direct personal benefits (Terdiman, 2005). This has come to be referred to as the del.icio.us lesson: “Personal value precedes network value: ... selfish use comes before shared use” (Porter, 2005b). Personal collections form the basis of many tagging applications. Users’ own photos are tagged in Flickr; MyLibrary enables the cataloguing of personal book collections (Rutkoff, 2006; Spalding, 2007). Organizing my stuff in shared space may have network effects, but the benefit is first and foremost that my stuff gets organized.

The act of tagging can be placed within the context of cognitive processes. In a summary of the cognitive stages/process involved in tagging a networked resource sinha asserts that tagging is “lighter” than categorization because it is comprised only one stage (associating an object with multiple concepts) rather than two (associating an object with multiple concepts, and then choosing the most appropriate) (sinha, 2005a).

Small scale user studies (interviews) and tag analysis show that tagging in del.icio.us is seen primarily as a personal memory aid (Trevino, 2005, 2006). These results are echoed in a study by Maarek et al. (2006) that looked at management of information gathered in a conference environment. Other motivations include exposure, opinion, attention getting, performance and activism (Zollers, 2007) as well as time management (Kipp, 2006b), self-organization (Kipp, 2007c), and to enable other functionality (Morrison, 2007). Tags might be characterized as an allocation of attention, something that could be measured in an “attention economy” (Ali Mejías, 2005).

But it has been noted that *not tagging* is the majority default position (Ames & Naaman, 2007); the matrix of self-organization, self-communication, social organization, and social communication might provide a framework to organize user studies exploring tagger motivation. We still need to know more about why people tag, whether taggers’ motivations differ in different circumstances, and how people might be encouraged to tag or to tag ‘well’.

The relationships between tagging and sociality also need further examination. The social nature of Web 2.0 and the rise of community-created content have been broadly observed (Prescott, 2007). But how do these factors relate to tagging? For example, Johnson notes, in blog posting following the DC 2006 meeting, that it is 'social' aspects – community and collaboration – distinguish tagging from individually created metadata (Johnston, 2006). But it remains unclear what factors exert social influence.

Lee (2006) posits that tagging behaviour changes when users are aware of the social nature of a tagging application. “Perceived social presence” may affect the way that users tag; in her study, users with more social data exposed on del.icio.us were more apt to annotate their bookmarks. How important is the social network (or social position) to tagging and folksonomy?

Lawley notes that tagging systems are open to “inevitable systematic disruptive behaviour” (*aka* gaming the system), that tags taken out of context can lose their meaning, and that tagging behaviour is influenced by context. She explores this last point by reflecting on her own behaviour using the ESP Game (<http://www.espgame.org>), and notes that the unintended consequences of tagging in the context of a game include skewing vocabulary towards ‘winning’ terms. She posits that similar factors are at play in other folksonomic environments and the social nature of these systems influences the way people tag (Lawley, 2005).

2.6 Social Tagging: Studying Socio-technical Systems

The third group of studies of social tagging and folksonomy examines user actions positioned within a broader framework. These studies describe systems and the inter-relationships of their parts. This includes the use of tags as navigational / retrieval tools, and methods of processing tags to facilitate their use within particular contexts.

The creators of Connotea (an academic social bookmarking application) offer an overview of existing social bookmarking tools that identifies their inherent characteristics as socio-technical systems: personal (or group) user accounts; a mechanism for entering data; tools to facilitate entry; free tagging; search by tag or user (sometimes Boolean); querying by popularity, users, or tags; RSS feeds; and extensions such as browser plug-ins (Hammond et al., 2005, from the conclusion). These characteristics are explored further in the introduction to Connotea, an open source, “free online reference management and social bookmarking service for scientists” that was inspired by del.icio.us (Lund et al., 2005). Novel features include the ability to identify a work as “mine” and an optional “comments” or annotations, from any user. References tagged with particular terms are available as an RSS feed. When studying the use of the system, the authors observed “Second-Order

Effects” characteristic of the system as a whole, including “tag convergence”, “implicit recommendation”, and the creation of “semi-automatic directory pages” (as tag pages are indexed by Google). CiteUlike.org has similar characteristics (Emamy & Cameron, 2007).

2.6.1 Information discovery

Other papers focus on particular aspects of the socio-technical tagging system. The value of social bookmarking as a navigation tool is explored in a study of the dogear social bookmarking service (internal to IBM). Millen and Feinberg found that showing tags to others enables browsing, an immediate benefit of tagging. “Considerable” browsing of others’ bookmarks took place: “74% of the total pages visited [using URLs in dogear] were bookmarked by someone else”. Tags that were used more often were also browsed more often (Millen & Feinberg, 2006, 4). Damianos et al. (2006) report similar results at Mitre Corporation. Sterling also highlights the role of folksonomy can play in new forms of information discovery, because it offers a window into the interests of the broad pool of users (Sterling, 2005). Following the tag-stream of trusted users becomes a new form of recommendation or current awareness service. Tags support a new kind of information discovery, across different types of Web pages, for time-sensitive, changing information (for example, searching blogs, and Flickr to pull together the on-line traces of an event such as a conference (Jarvis, 2006)). Other studies of tag-based identification of relevant content include Hayes and Avesani (2007), who cluster blog postings based on tags, and find that patterns in tagging behaviour (as well as tags themselves) reveal meaningful groupings. Hyde et al. (2007) propose a means of using tags for shared bookmarks alongside keywords extracted from the sites tagged to generate recommendations of related materials.

The role of tag-based discovery also needs to be studied, particularly vis à vis other means of providing enhanced access to collections, including thesaurally-enhanced searching.

2.6.2 Searching using tags

Winget (2006) explores ways that tags support information discovery within Flickr, noting that while searching with tags does find relevant and even ‘interesting’ images, navigation is more like exploration than known-item searching. sinha (2005b) offers observations on clustering in Flickr (flickr.com, 2006); there are no names for the clusters and no attempt at complex visualizations. She contrasts this with grokker (<http://www.grokker.com>), where results are displayed in named groups or as a ‘zoomable map’, or Clusty (<http://www.clusty.com>) where clusters are named, partially because it is difficult to read meaning into lists of URLs.

Lerman and Jones (2007) note that *social browsing* – viewing pictures posted by pre-identified personal contacts – is one of the major ways that users navigate in Flickr. Dubinko

et al. (2006) present an approach to creating a visualization of photos from Flickr, based on tags, processed by factors such as 'interestingness' as well as categories of tags such as "events", "personalities" and what the authors' call 'social media tagging' — clusters of photos based on emergent themes like 'What's in your fridge?'.

Aurnhammer et al., (Aurnhammer, Hanappe et al., 2006; Aurnhammer, Steels et al., 2006) integrate tagging with search based on visual features (CBIR), and show in a small experiment that there may be promise in this method, both for using visual elements to disambiguate tag vocabulary and for using visual search to improve tag-based navigation. Jorgensen (2007) argues for this kind of hybrid approach, integrating indexing, content based image retrieval and tagging, to take into account the many different facets of meaning in visual expression.

Tag streams can also be combined with other data sources. For example, Biddulph demonstrates in a prototype application, how tags from del.icio.us could be linked into recordings made available at the BBC Radio on-line site. Real-time integration of BBC materials that shared tags might function as a recommender, providing links to related materials (Biddulph, 2005).

2.6.3 Visualization

The availability of tag streams as RSS data has enabled the development of a number of tag/data visualization tools. The most prevalent is the tag cloud, a visualization of the tags represented in a system, with tags weighted (most often by size and typographic weight) based on frequency of occurrence. del.icio.us offers a simple tag cloud interface for its popular or recent tags (del.icio.us, 2007). grokker allows the hierarchical or graphical display of the results of a keyword search (against Yahoo! Wikipedia, and/or Amazon books); found sites are presented as thematic clusters in an hierarchy or as 'bubbles' within a circular conceptual space (grokker, 2005a, 2005b). Chudnov has shown more sophisticated examples of visualizations of unallog tag data, that include a temporal element (Chudnov, 2006). Hassan-Monteroa & Herrero-Solanaa (2006) offer a method for improving tag clouds by basing them on clustering rather than alphabetical order. Kaser and Lemire (2007) offer improved algorithms for tag cloud typography. Zeldman criticises the use of tag clouds as a primary navigational metaphor, noting that they sacrifice hierarchy in favour of a visualization of the popular; he asserts that hierarchy is a proven way to parse a large body of materials, and to communicate the linkages between different sub-classes, two essential keys to navigating large information resources (Zeldman, 2005a, 2005b).

2.6.4 Ranking

The availability of tag data linked to a particular user enables ranking of users, based on the characteristics of their tagging. John and Seligman present a scenario for the identification of experts in the enterprise, through a ranking system – ExpertRank – based on expertise and clusters of related tags (John & Seligmann, 2006). A similar approach is taken by Michlmayr and Cayzer (2007), who propose a method to visualize a user's tag space, and use the resulting profile as a way of identifying relevant material.

Collaborative Rank (no longer operational) was an experimental tool that harvested data from del.icio.us and provided user rankings based on influence – defined as how often a user's tags are copied (Andrews, 2005a; Michail, 2005). High ranking may also be a motivation for tagging (Chanchal Gupta, a top-ranked tagger according to CollaborativeRank quoted in Andrews, 2005a). Similarly the creators of Bibsonomy designed an algorithm – FolkRank (adapted from PageRank) – that computes relevance from users, tags and resources, supplemented by tag assignment (a particular user's use of a tag), that was tested with large body of data scraped from del.icio.us: 75,242 users, 533,191 tags, 3,158,297 resources. The algorithm computed weighting based on the premise that “a resource which is tagged with important tags by important users becomes important itself. The same holds, symmetrically, for tags and users”. FolkRank offers: recommendation, suggestion of related tags, integration of other users' behaviour into suggestions, and suggestion of other users with shared interests (Hotho et al., 2006).

Tagging people offers another way of ranking them, as it makes expertise explicit. It has been shown to enable the creation of community (through the declaration of a field of knowledge, geographic area, affiliation with a project, or association with an event) in a corporate context (Farrell & Lau, 2006; Farrell et al., 2007). Explicit tagging of expertise is favourably compared by Bogers et al. (2006) to the extraction of implicit evidence of expertise from other sources. But what if tagging becomes a corporate function? In introducing the concept of ‘induced tagging’ – tagging done by digital library professionals for a corporate group – Sánchez et al. (2007) challenge the personal motivation (self-benefit) seen by some to be the central to tagging. Shared tagging of content can also be helpful for group formation, and for the sharing of knowledge (Yew et al., 2006).

2.6.5 Social networks and tags: finding latent structure

Socio-technical tagging systems can also be described according to network theory, studying the relationships between users, tags and resources. Trevino (2005) provides three possible visualizations of the relationship between users, tags and Web pages, as ways to approach social bookmarking systems. The first shows users between multiple tags and a single Web

page (a Web page is tagged by many users with many different tags). The second shows users between a tag, and many Web pages (a tag is assigned by many users to different Web pages). The third shows tags between a single user and many Web pages (a user assigns many tags to many Web pages). Each version is valid. Which visualization is more appropriate depends on where you are in the process: doing the tagging, looking at tags, or finding Web pages using tags, i.e. whether you are interested in tagging, folksonomy or social tagging systems.

Shen and Wu (2005) define the following properties of a folksonomy network: “degree distribution” (“the number of tags which share at least one piece of content with the tag”); “clustering coefficient” (a measure of “the interrelatedness of i's [a node's] neighbors”), and “average path length” (the average “minimum length of path” between nodes) (p. 2-3). Using data drawn from del.icio.us, they constructed a network between urls and tags; user data was ignored. The two days of test data had the characteristics of a “small world network” (a small number of hops between nodes) and a “scale free network” (some nodes are highly connected, most are not and scale doesn't change this). Schmitz et al. (2007) study the networks structure of Bibsonomy and del.icio.us and also find similar characteristics of a small world network. They also looked at relative path lengths across the tripartite *user | tag | resource* network, and studied tag co-occurrence, noting that “evidence is compatible with the existence of complex, possibly hierarchical structures in the network of tag co-occurrence.” (C. Schmitz et al., 2007, section 6.1). These hierarchical structures are the subject of the ontological deduction research described earlier.

Lambiotte and Ausloos (2005) propose mathematical methods to characterise relationships between users, items and tags that enable profiling. In a bipartite network of users and items, “vectors are signatures of the users/items, that account for their interests/audience” (p. 7). If only a single axis is analysed (the relationship between tags, for example) it is possible to use mathematical methods to reduce a branched network structure to a hierarchical representation of its strongest relationships (Fig 2). It is proposed that these 'blocks' “could be identified as families/genres/communities” (p. 8).

Mieszkowski (2005) distinguishes tagging from other social networking activities, because the networks formed by shared tags are emergent, not pre-declared as ‘friends’ are.

2.7 Alternative Perspectives on Tagging and Folksonomy

There are some studies that look at social tagging and folksonomy from analytical traditions other than information retrieval (folksonomy); user experience (tagging); or network analysis (social tagging systems). Within the context of the historical development of hypertext, and the tension regarding authorship that emerged in critical theory (Barthes declaring

authorship dead vs. Foucault asserting that authorship was the last bastion of rationality), the tagging environments of social tools such as Flickr or del.icio.us are described by critical theorist Walker as 'feral' (or having reverted to the wild) in contrast to the more 'domesticated' access provided by formal classification systems. These tagging sites are provided as models for a more feral form of hypertext fiction, of which blogging is used as an emergent example (Walker, 2005).

Campbell (2006) uses phenomenology to integrate models of tagging and the semantic Web (which he sees as more "highly structured"). With Foucault as context, he asserts Husserl's dialogic relationship between the individual and the world, between a person and concepts, that uses concepts held internally as a referent. Shared inter-subjectivities, a community of shared concepts, underpin collective understanding. He sees a tension between inter-subjectivity defined by control (vocabularies and ontologies) and one defined by emergent semantics (tagging), that pits consistent activities by trained experts which create contrived and elaborate structures against loose, evolving structures created by end-users of information. These two visions seem to be in conflict, but he points, instead, to a possibility of creative tension.

Campbell and Fast explore other theoretical frameworks, including *pace layering* (Brand, 1994) which posits that "different rates of change are beneficial to a complex system", noting that Morville uses this theory in *Ambient Findability* (2005) [discussed earlier] to explain how Information Architecture can accommodate new things like 'tagging'. They also examine *complexity theory* noting that hostility and volatility may characterise complex systems. *Resilience theory* of information ecology extends pace theory, focusing on the role of change in complex systems. Both theories point out the perils of denying change. Campbell and Fast assert that the "complex interactions of productive revolt that characterize adaptive systems" could help define a research agenda to move Information Architecture ahead (Campbell & Fast, 2006).

3. Conclusions

These early studies of tags, folksonomy, tagging and social tagging systems, are descriptive and theoretical. They often present basic quantitative analysis of data (descriptive statistics), gathering data sets by harvesting publicly available sources such as Flickr, del.icio.us, CiteUlike, or Connotea. These datasets are incomplete (as they do not represent all users or the entire tag vocabulary or tagging activity of any particular user). Tools such as vocabulary analysis and classification, user interaction theory, and social network theory are used to describe and analyse the nature of tagging and folksonomy, both characterised as new phenomena.

Proponents of tagging systems cite a number of theoretical advantages derived from tagging, from the resulting folksonomy, and from network effects. Tagging begins as a personal information management and re-discovery tool. New concepts often emerge in personal tags that are then shared in social systems, where social information discovery leads new users to content. This is where social tagging seems to offer a number of affordances, primarily related to the use of the resulting folksonomy by others for information retrieval, browsing, searching or current awareness. User involvement in tagging may also have inherent advantages, promoting a sense of ownership of content, and developing social cohesion in a group. Broad distribution of effort offers a way to conquer the economics of indexing.

But tagging and folksonomy are not without their critics. Particularly in the information retrieval community, there is concern over the relativistic nature of socially created vocabularies, the lack of term consistency, problems with synonymy and polysemy, and the inherent inconsistency of a user-generated vocabulary. All of these characteristics may limit the value of a folksonomy as an indexing language and retrieval tool.

The value and contribution of social tagging and folksonomy has not yet been fully established. Studies of tagging and folksonomy need to come together with studies of searching in holistic analyses of information discovery and use (and re-discovery and re-use (Bearman & Trant, 1998)) that are sensitive to the complex personal information-seeking context alluded to by Markey (2007b).

Most studies of tagging and folksonomy reflect one of three broad perspectives, focused on folksonomic vocabularies and information retrieval, on user behaviour in tagging systems, or on the socio-technical nature of those systems themselves and their description using network theory. Tagging can be examined as a personal information management tool, or a retrieval tool. Tags can be described with distribution statistics. Folksonomic vocabularies can be analysed by subject or type of term. The development and use of systems can be placed within various theoretical frameworks, and their social and technical contexts of use described and analysed. The social context of tagging can be studied through the lense of authorship, or described and analysed using constructs from social network theory. Each of these approaches adds to our appreciation of the role of user-generated content in the information landscape.

4. Acknowledgements

This work was supported in part by a grant from the US Institute of Museum and Library Services (IMLS) (The Metropolitan Museum of Art et al., 2006).

5. References

- [Collective]. (2007). Talk: Folksonomy. *Wikipedia* Retrieved December 12, 2007 from <http://en.wikipedia.org/wiki/Talk:Folksonomy>
- Ahn, L. v., & Dabbish, L. (2004). Labeling Images with a Computer Game. CHI 2004, Vienna, Austria.
- Akst, D. (2005, December 11). Unbundles of Joy (Column: "On the Contrary"). *New York Times* Retrieved December 12, 2005 from <http://www.nytimes.com/2005/12/11/business/yourmoney/11cont.html>.
- Alexander, B. (2006). Web 2.0: A New Wave of Innovation for Teaching and Learning? *EDUCAUSE Review*, 41(2), 32-44. from <http://www.educause.edu/ir/library/pdf/erm0621.pdf>
- Ali Mejías, U. (2005, April 26, 2005). Tag Literacy. *Ideant [blog]*
- Ames, M., & Naaman, M. (2007). *Why We Tag: Motivations for Annotation in Mobile and Online Media*. CHI 2007, San Jose, California.
- Anderson, C. (2004). The Long Tail. *Wired*, 12(10). from <http://www.wired.com/wired/archive/12.10/tail.html>
- Anderson, P. (2007). *What is Web 2.0? Ideas, technologies and implications for education*: Joint Information Systems Committee (JISC) Retrieved January 31, 2008 from <http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf>
- Andrews, R. (2005a). Tag, You're It: Best Bookmarker. *Wired News*. Retrieved January 31, 2008 from <http://www.wired.com/news/technology/0,69083-0.html>
- Andrews, R. (2005b). Tips from Top Taggers. *Wired News*. Retrieved January 31, 2008 from <http://www.wired.com/news/technology/0,1282,69084,00.html>
- Ansari, M. (2005). Matching Between Assigned Descriptors and Title Keywords in Medical Theses. *Library Review*, 54(7), 410-414.
- Aurnhammer, M., Hanappe, P., & Steels, L. (2006). Augmenting Navigation for Collaborative Tagging with Emergent Semantics. International Semantic Web Conference (ISWC2006); Lecture Notes in Computer Science, Athens, Georgia, USA Retrieved January 31, 2008 from <http://iswc2006.semanticweb.org/items/Aurnhammer2006ve.pdf>.

- Aurnhammer, M., Steels, L., & Hanappe, P. (2006). Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/17.pdf.
- Bar-Ilan, J., Shoham, S., Idan, A., Miller, Y., & Shachak, A. (2006). Structured vs. unstructured tagging – A case study. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland Retrieved December 4, 2007 from http://www.ibiblio.org/www_tagging/2006/12.pdf.
- Bateman, S., Brooks, C., McCalla, G., & Brusilovsky, P. (2007). Applying Collaborative Tagging to E-Learning. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_56.pdf.
- Bearman, D., & Trant, J. (1998). Unifying our cultural memory: Could electronic environments bridge the historical accidents that fragment cultural collections? In L. Dempsey, S. Criddle & R. Heseltine (Eds.), *Information Landscapes for a Learning Society, Networking and the Future of Libraries* (Vol. 3).
- Bearman, D., & Trant, J. (2005). Social Terminology Enhancement through Vernacular Engagement: Exploring Collaborative Annotation to Encourage Interaction with Museum Collections. *D-Lib Magazine*, 11(9). Retrieved September 11, 2006 from <http://www.dlib.org/dlib/september05/bearman/09bearman.html>
- Beckett, D. (2006). Semantics Through the Tag. XTech 2006: Building Web 2.0, Amsterdam, The Netherlands Retrieved January 31, 2008 from <http://xtech06.usefulinc.com/schedule/paper/135>.
- Begelman, G., Keller, P., & Smadja, F. (2006). *Automated Tag Clustering: Improving search and exploration in the tag space*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/20.pdf.
- Berendt, B., & Hanser, C. (2007). *Tags are not metadata, but "just more content" - to some people*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/2--Berendt-Hanser.pdf>.
- Biddulph, M. (2005, January 2005). Radio 3 versus del.icio.us. *hackdiary.com [blog]* Retrieved December 20 from <http://www.hackdiary.com/misc/firefox-delicious-demo.html>

- Bogers, T., Thoonen, W., & van den Bosch, A. (2006). Expertise classification: Collaborative classification vs. automatic extraction. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/1792/01/bogers.pdf>.
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out* (Paperback published 2000 ed.). Cambridge Massachusetts / London England: The MIT Press.
- boyd, d. (2005, January 28, 2005). issues of culture in ethnoclassification/folksonomy. *apophenia :: making connections where none previously existed [blog]* (also published on many2many Retrieved December 20 from http://www.zephorio.org/thoughts/archives/2005/01/28/issues_of_culture_in_ethnoclassificationfolksonomy.html
- Byde, A., Wan, H., & Cayzer, S. (2007). Personalized Tag Recommendations via Tagging and Content-based Similarity Metrics. *Weblogs and Social Media*, Boulder, Colorado, USA Retrieved December 5, 2007 from <http://www.icwsm.org/papers/4--Byde-Wan-Cayzer.pdf>.
- Campbell, D. G. (2006). A phenomenological framework for the relationship between the Semantic Web and user-centered tagging systems. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved December 7, 2007 from <http://dlist.sir.arizona.edu/1838/01/campbell.pdf>.
- Campbell, D. G., & Fast, K. V. (2006). From Pace Layering to Resilience Theory: the Complex Implications of Tagging for Information Architecture. *IA Summit 2006* Retrieved January 31, 2008 from http://www.iasummit.org/2006/files/164_Presentation_Desc.pdf.
- Cattuto, C., Loreto, V., & Pietronero, L. (2006). Collaborative Tagging and Semiotic Dynamics. from <http://arxiv.org/pdf/cs/0605015>.
- Chudnov, D. (2006, January 6). visualizing folksonomy data from unalog using Starlight. *One Big Library [blog]* Retrieved January 10 from <http://onebiglibrary.net/node/13>
- Chun, S., & Jenkins, M. (2005). Cataloguing by Crowd; a proposal for the development of a community cataloguing tool to capture subject information for images (A Professional Forum). *Museums and the Web 2005*, Vancouver Retrieved January 31, 2008 from http://www.archimuse.com/mw2005/abstracts/prg_280000899.html.

Cunningham, S. J., & Masoodian, M. (2006). Looking for a picture: an analysis of everyday image information searching. Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries Chapel Hill, NC, USA.

Damianos, L., Griffith, J., & Cuomo, D. (2006). Onomi: Social Bookmarking on a Corporate Intranet. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/28.pdf.

Davis, I. (2005). Why Tagging Is Expensive. *Panlibus [was the Silkworm blog]*. from http://blogs.talis.com/panlibus/archives/2005/09/why_tagging_is.php

De Smet, W., & Moens, M.-F. (2007). *Extraction of Folksonomies from Noisy Texts*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/4--Smet-Moens.pdf>.

del.icio.us. (2007, December 7, 2007). Popular Tags on del.icio.us: Tag cloud visualization. Retrieved December 7, 2007 from <http://del.icio.us/tag/>

Dix, A., Levialdi, S., & Malizia, A. (2006). *Semantic Halo for Collaboration Tagging Systems*. Workshop on the Social Navigation and Community-Based Adaptation Technologies In Conjunction with Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), Dublin, Ireland. Retrieved January 31, 2008 from http://www.sis.pitt.edu/~paws/SNC_BAT06/crc/malizia.pdf.

Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., & Tomkins, A. (2006). Visualizing Tags over time. 15th International Conference on World Wide Web, Edinburgh, Scotland Retrieved January 31, 2008 from <http://doi.acm.org/10.1145/1135777.1135810>.

Dumais, S., Cuttrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D. C., et al. (2003). Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *Proceedings of the ACM Conference on Information Retrieval*, 72-79.

Emamy, K., & Cameron, R. (2007). Citeulike: A Researcher's Social Bookmarking Service. *Ariadne*, 51. from <http://www.ariadne.ac.uk/issue51/emamy-cameron/>

Farrell, S., & Lau, T. (2006). *Fringe Contacts: People-Tagging for the Enterprise*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/25.pdf.

Farrell, S., Lau, T., & Nusser, S. (2007). Building Communities with People-Tags. Human-Computer Interaction – INTERACT 2007; Lecture Notes in Computer Science Retrieved January 31, 2008 from <http://www.springerlink.com.myaccess.library.utoronto.ca/content/e541852635m2561w/fulltext.pdf>.

Feinberg, M. (2006). An examination of authority in social classification systems. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved December 6, 2007 from <http://dlist.sir.arizona.edu/1783/01/feinberg.pdf>.

flickr.com. (2006). flickr Tags / museum / clusters. In flickr-museum060123.jpg (Ed.). Retrieved January 23, 2006 from <http://www.flickr.com/photos/tags/museum/clusters/>.

Fokker, J., Pouwelse, J., & Buntine, W. (2006). *Tag-Based Navigation for Peer-to-Peer Wikipedia*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/9.pdf.

Frost, C. (1989). Title word as entry vocabulary to LCSH: correlation between assigned LCSH term and derived term from title. *Cataloging and Classification Quarterly*, 10, 165-179.

Furner, J., & Tennis, J. T. (2006). *Advances in Classification Research, Volume 17: Proceedings of the 17th ASIS&T Classification Research Workshop*, Austin, Texas, USA.

Gil-Leiva, I., & Alonso-Arroyo, A. (2007). Keywords given by authors of scientific articles in database descriptors. *Journal of the American Society for Information Science and Technology*, 58(8), 1175-1187.

Golder, S. A., & Huberman, B. A. (2005). The Structure of Collaborative Tagging Systems. *Journal of Information Science [preprint]* ([preprint Sept 8, 2005]). from <http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf>

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208.

Google. (2007a). Google Image Labeller. Retrieved November 29, 2007 from <http://images.google.com/imagelabeler>

Google. (2007b). Google Image Search. Retrieved November 29, 2007

Gospodnetić, O. (2005a, 28 September). Web 2.0 Quick Study: Tag Use Patterns. *Simpy Chichimichi [blog]* from <http://blog.simpy.com/blojsom/blog/2005/09/28/Quick-Study-Tag-Use-Patterns.html>

Gospodnetić, O. (2005b, 29 September). Web 2.0 Tagging Study, Part 1: How Much We Tag. *Simpy Chichimichi [blog]* from <http://blog.simpy.com/blojsom/blog/2005/09/29/Tagging-Study-Part-1-How-Much-We-Tag.html>

Gospodnetić, O. (2005c, 5 October). Web 2.0 Tagging Study, Part 2: Private vs. Public Tagging. *Simpy Chichimichi [blog]* from <http://blog.simpy.com/blojsom/blog/2005/10/05/Tagging-Study-Part-2-Private-vs-Public-Tagging.html>

Gospodnetić, O. (2005d, 19 October). Web 2.0 Tagging Study, Part 3: Timing of Read vs. Write *Simpy Chichimichi [blog]* from <http://blog.simpy.com/blojsom/blog/2005/10/19/Web-2-0-Tagging-Study-Part-3-Timing-of-Read-vs-Write.html>

Gospodnetić, O. (2005e, 19 October). Web 2.0 Tagging Study, Part 4: Nocturnal Activity Pattern *Simpy Chichimichi [blog]* from <http://blog.simpy.com/blojsom/blog/2005/10/19/Web-2-0-Tagging-Study-Nocturnal-Activity-Pattern.html>

grokker. (2005a, [ongoing]). grokker.com: folksonomy - diagrammatic view. Retrieved December 22 from <http://www.grokker.com/applet.html?query=folksonomy>

grokker. (2005b, [ongoing]). grokker.com: folksonomy - hierarchical view. Retrieved December 22 from <http://www.grokker.com/applet.html?query=folksonomy>

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up Tags? *D-Lib Magazine*, 12(1). Retrieved April 21, 2006 from <http://www.dlib.org/dlib/january06/guy/01guy.html>

Halpin, H., Robu, V., & Shepherd, H. (2007). *The Complex Dynamics of Collaborative Tagging*. WWW2007, Banff, Alberta, Canada. Retrieved January 31, 2008 from <http://www2007.org/papers/paper635.pdf>.

- Hammond, T., Hannay, T., Lund, B., & Scott, J. (2005). Social Bookmarking Tools (I): A General Review. *D-Lib Magazine*, 11(4). Retrieved December 12, 2005 from <http://www.dlib.org/dlib/april05/hammond/04hammond.html>
- Hassan-Monteroa, Y., & Herrero-Solana, V. c. (2006). *Improving Tag-Clouds as Visual Information Retrieval Interfaces* International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006, Medira, Spain. Retrieved January 31, 2008 from http://www.nosolousabilidad.com/hassan/improving_tagclouds.pdf.
- Hayes, C., & Avesani, P. (2007). *Using Tags and Clustering to Identify Topic-Relevant Blogs*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/2--Hayes-Avesani.pdf>.
- Hayman, S. (2007). Folksonomies and Tagging: New developments in social bookmarking. Ark Group Conference: Developing and Improving Classification Schemes, Rydges World Square, Sydney Retrieved January 31, 2008 from <http://www.educationau.edu.au/jahia/webdav/site/myjahiasite/shared/papers/arkhayman.pdf>.
- Heyman, P., & Garcia-Molina, H. (2006). *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*: Stanford University Retrieved January 31, 2008 from <http://dbpubs.stanford.edu/pub/showDoc.Fulltext?lang=en&doc=2006-10&format=pdf&compression=&name=2006-10.pdf>
- Hotho, A., Jäschke, R., Schmitzz, C., & Stumme, G. (2006). Information Retrieval in Folksonomies: Search and Ranking. In *The Semantic Web: Research and Applications: Lecture Notes in Computer Science (3rd European Semantic Web Conference, ESWC 2006 Budva, Montenegro, June 11-14, 2006 Proceedings)* (pp. 411-426). Heidelberg: Springer Berlin.
- Iturrioz, J., Diaz, O., & Arellan, C. (2007). Towards Federated Web2.0 Sites: The TAGMAS Approach Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_34.pdf.
- Jarvis, J. (2006). Tagging - the latest way to search the web. *MediaGuardian*. from <http://media.guardian.co.uk/broadcast/story/0,,1676388,00.html>
- John, A., & Seligmann, D. (2006). *Collaborative Tagging and Expertise in the Enterprise*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop,

- Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/26.pdf.
- Johnston, P. (2006, November 16, 2006). The "social" in social tagging. *eFoundations: metadata, middleware, e-learning [blog]* Retrieved November 21
- Jørgensen, C. (2007). Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift. 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research,, Milwaukee, Wisconsin, USA. J. Lussky (Ed.), Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/2064/01/Jorgensen%5FUpdate%5FSIG-CR%5F2007.doc>.
- Kaser, O., & Lemire, D. (2007). Tag-Cloud Drawing: Algorithms for Cloud Visualization. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_12.pdf.
- Kelkar, S., John, A., & Seligmann, D. (2007). An Activity-based Perspective of Collaborative Tagging. International Conference on Weblogs an Social Media, Boulder, Colorado, USA Retrieved January 31, 2008 from <http://www.icwsm.org/papers/2--Kelkar-John-Seligmann.pdf>.
- Kilkki, K. (2007). A practical model for analyzing long tails. *First Monday*, 12(5). from http://www.firstmonday.org/issues/issue12_5/kilkki/index.html
- Kipp, M. E. (2006a). Complementary or Discrete Contexts in Online Indexing: A Comparison of User, Creator and Intermediary Keywords. Canadian Association for Information Science, Toronto, Ontario, Canada Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/1533/01/mkipp-caispaper.pdf>.
- Kipp, M. E. (2006b). Exploring the context of user, creator and intermediate tagging. IA Summit 2006, Vancouver, BC Retrieved January 31, 2008 from http://www.iasummit.org/2006/files/109_Presentation_Desc.pdf.
- Kipp, M. E. (2007a). Tagging for health information organisation and retrieval. North American Symposium on Knowledge Organization, Toronto, Ontario. J. T. Tennis (Ed.), Retrieved December 12, 2007 from <http://dlist.sir.arizona.edu/1909>.

- Kipp, M. E. (2007b). Tagging Practices on Research Oriented Social Bookmarking Sites. Canadian Association for Information Science, Montreal, Quebec, Canada Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/2027/01/kipp%5F2007.pdf>.
- Kipp, M. E. (2007c). @toread and Cool: Tagging for Time, Task and Emotion. 8th Information Architecture Summit, Las Vegas, Nevada, USA Retrieved December 12, 2007 from <http://dlist.sir.arizona.edu/1947/01/mkipp-iasummit2007.pdf>.
- Kipp, M. E., & Campbell, D. G. (2006). Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. American Society for Information Science and Technology, Austin, TX Retrieved December 12, 2007 from <http://dlist.sir.arizona.edu/1704/01/KippCampbellASIST.pdf>.
- Kroski, E. (2005). The Hive Mind: Folksonomies and User-Based Tagging. *InfoTangle [blog]* Retrieved December 12 from <http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>
- Lambiotte, R., & Ausloos, M. (2005). Collaborative tagging as a tripartite network [Electronic Version]. *arXiv: Computer Science*. Retrieved January 1, 2006 from <http://arxiv.org/abs/cs/0512090v2>.
- Landbeck, C. (2007). Trouble in Paradise: Conflict Management and Resolution in Social Classification Environments by Richard Urban. *ASIS&T Bulletin: Special Section: Folksonomies*. from <http://www.asis.org/Bulletin/Oct-07/landbeck.html>
- Lawley, L. (2005, January 20, 2005). social consequences of social tagging. *Many2Many* Retrieved December 14
- Lee, K. J. (2006). *What goes around comes around: an analysis of del.icio.us as social space*. Computer Supported Cooperative Work, Banff, Alberta, Canada. Retrieved January 31, 2008 from <http://doi.acm.org.myaccess.library.utoronto.ca/10.1145/1180875.1180905>.
- Lerman, K., & Jones, L. (2007). *Social Browsing on Flickr*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/3--Lerman-Jones.pdf>.
- Lin, X., Beaudoin, J. E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. *Advances in Classification Research, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop*, Austin, Texas, USA. J. Furner & J.

- T. Tennis (Eds.). Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/1790/01/lin.pdf>.
- Lund, B., Hammond, T., Flack, M., & Hannay, T. (2005). Social Bookmarking Tools (II): A Case Study - Connotea. *D-Lib Magazine*, 11(4). Retrieved April 22, 2008 from <http://www.dlib.org/dlib/april05/lund/04lund.html>
- Maarek, Y. S., Marmasse, N., Navon, Y., & Soroka, V. (2006). *Tagging the physical world*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 4, 2007 from http://www.ibiblio.org/www_tagging/2006/21.pdf.
- Macgregor, G., & McCulloch, E. (2006). Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55(5), 291 - 300. from http://eprints.rclis.org/archive/00005703/01/CollaborativeTaggingToolPaperGmEm_preprint.pdf
- Marchetti, A., Tesconi, M., Ronzano, F., Rosella, M., & Minutoli, S. (2007). SemKey: A Semantic Collaborative Tagging System. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_45.pdf.
- Markey, K. (2007a). Twenty-five years of end-user searching, Part 1: Research findings. *Journal of the American Society for Information Science and Technology*, 58(8), 1071-1081. from <http://www3.interscience.wiley.com.myaccess.library.utoronto.ca/cgi-bin/fulltext/114210851/PDFSTART>
- Markey, K. (2007b). Twenty-five years of end-user searching, Part 2: Future research directions. *Journal of the American Society for Information Science and Technology*, 58(8), 1123-1130. from <http://www3.interscience.wiley.com.myaccess.library.utoronto.ca/cgi-bin/fulltext/114210853/PDFSTART>
- Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006a). *HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead*. Proceedings of the seventeenth conference on Hypertext and Hypermedia Odense, Denmark. Retrieved December 7, 2008 from <http://doi.acm.org/10.1145/1149941.1149949>.
- Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006b). Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead World Wide Web 2006 (WWW2006):

- Collaborative Web Tagging Workshop, Edinburgh, Scotland Retrieved December 9, 2007 from http://www.ibiblio.org/www_tagging/2006/29.pdf.
- Mathes, A. (2004, December). Folksonomies - Cooperative Classification and Communication through Shared Metadata. Retrieved August 6 from <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Merholz, P. (2004a, August 30, 2004). Ethnoclassification and vernacular vocabularies. *Peterme.com: Links, thoughts, and essays from Peter Merholz [blog]* Retrieved November 28, 2007 from <http://www.peterme.com/archives/000387.html>
- Merholz, P. (2004b, October 19, 2004). Metadata for the Masses. Retrieved December 15 from <http://www.adaptivepath.com/publications/essays/archives/000361.php>
- Merholz, P. (2005, 2005 08 07). Clay Shirky's Viewpoints are Overrated. *peterme.com [blog]* Retrieved November 29, 2007 from <http://www.peterme.com/archives/000558.html>
- Michail, A. (2005, [continuously updated]). Collaborative Rank. Retrieved January 2 from <http://www.collabrank.org/>
- Michlmayr, E., Cayzer, S., & Shabajee, P. (2007). Learning User Profiles from Tagging Data and Leveraging them for Personal(ized) Information Access. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_29.pdf.
- Mieszkowski, K. (2005). Steal this bookmark! *salon.com*. from <http://www.salon.com/tech/feature/2005/02/08/tagging/>
- Mika, P. (2007). Ontologies are us: A unified model of social networks and semantics. *Journal of Web Semantics*, 5(1), 5-15.
- Millen, D. R., & Feinberg, J. (2006). Using Social Tagging to Improve Social Navigation. Workshop on the Social Navigation and Community-Based Adaptation Technologies In Conjunction with Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), Dublin, Ireland Retrieved November 20, 2007 from http://www.sis.pitt.edu/~paws/SNC_BAT06/crc/millen.pdf.
- Miller, P. (2006). Web 2.0: Building the New Library. *Ariadne*, 45(October). from <http://www.ariadne.ac.uk/issue45/miller/>

Mishne, G. AutoTag: A Collaborative Approach to Automated Tag Assignment for Weblog Posts. 15th International World Wide Web Conference (WWW2006), Edinburgh, Scotland Retrieved December 13, 2007 from <http://staff.science.uva.nl/~gilad/pubs/www2006-blogtags.pdf>.

Mishne, G. (2006). AutoTag: A Collaborative Approach to Automated Tag Assignment for Weblog Posts. 15th International World Wide Web Conference (WWW2006), Edinburgh, Scotland Retrieved December 13, 2007 from <http://staff.science.uva.nl/~gilad/pubs/www2006-blogtags.pdf>.

Morrison, P. J. (2007). Why Are They Tagging, and Why Do We Want Them To? *ASIS&T Bulletin: Special Section: Folksonomies, October/November 2007*. from <http://www.asis.org/Bulletin/Oct-07/morrison.html>

Morville, P. (2005). *Ambient Findability: What we find changes who we become*. Cambridge: O'Reilly.

Noruzi, A. (2006). Folksonomies: (Un)Controlled Vocabulary? *Knowledge Organization*, 33(4), 199-203. Retrieved January 31, 2008 from http://eprints.rclis.org/archive/00011286/01/Folksonomy%2C_UnControlled_Vocabulary.pdf

O'Reilly, T. (2005, September 30, 2005). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *O'Reilly.com [blog]* from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>

Ohmukai, I., Hamasaki, M., & Takeda, H. (2005). A Proposal of Community-based Folksonomy with RDF Metadata. 4th International Semantic Web Conference, Galway, Ireland Retrieved January 31, 2008 from http://www.ifi.uzh.ch/ddis/fileadmin/events/iswc2005ws/CameraReady/Ohmukai_Folksonomy_021.pdf.

Passant, A. (2007). *Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs: Theoretical background and corporate use-case*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/2--Passant.pdf>.

Pennsylvania State University Library. (2005-). PennTags. Retrieved April 15, 2006 from <http://tags.library.upenn.edu/>

- Peterson, E. (2006). Beneath the Metadata: Some Philosophical Problems with Folksonomy. *D-Lib Magazine*, 12(11). from <http://www.dlib.org/dlib/november06/peterson/11peterson.html>
- Pink, D. (2005). Year in Ideas: Folksonomy. *The New York Times Magazine*. Retrieved January 31, 2008 from <http://select.nytimes.com/search/restricted/article?res=F10914F938550C728DDDB0994DD404482>
- Porter, J. (2005a). Controlled Vocabularies Cut Off the Long Tail. *Bokardo.com: Social Web design [blog]*. from http://bokardo.com/archives/controlled_vocabularies_long_tail/
- Porter, J. (2005b, December 21 2005). Learning More about Structured Blogging. *bokardo, a blog about web 2.0 [blog]* Retrieved December 22 from <http://bokardo.com/archives/learning-more-about-structured-blogging/>
- Porter, J. (2007). Taxonomies and Tags. *Bokardo [blog]* Retrieved October 15 from <http://bokardo.com/archives/taxonomies-and-tags/>
- Prescott, L. (2007). *Hitwise US Customer Generated Media Report*: Hitwise
- Quintarelli, E. (2005). Folksonomies: power to the people. ISKO Italy - UniMIB Meeting, Milan Retrieved December 13, 2005 from <http://www.iskoi.org/doc/folksonomies.htm>.
- Quintarelli, E., Resmini, A., & Rosati, L. (2007). FaceTag: Integrating Bottom-up and Top-down Classification in a Social Tagging System. *ASIS&T Bulletin, June/July 2007*. from http://www.asis.org/Bulletin/Jun-07/quintarelli_et_al.html
- Rosenfeld, L. (2005, January 6, 2005). Folksonomies? How about Metadata Ecologies? Retrieved December 22 from http://www.louisrosenfeld.com/home/bloug_archive/000330.html
- Rutkoff, A. (2006). Social Networking for Bookworms. *The Wall Street Journal*.
- Sánchez, J. A., Arzamendi-Pétriz, A., & Valdiviezo, O. (2007). *Induced tagging: promoting resource discovery and recommendation in digital libraries*. International Conference on Digital Libraries, Vancouver, BC, Canada. Retrieved January 31, 2008 from <http://doi.acm.org.myaccess.library.utoronto.ca/10.1145/1255175.1255252>.

- Schmitz, C., Grahl, M., Hotho, A., Stumme, G., Cattuto, C., Baldassarri, A., et al. (2007). Network Properties of Folksonomies. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_13.pdf.
- Schmitz, P. (2006). *Inducing Ontology from Flickr Tags*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/22.pdf.
- Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., et al. (2006). *Tagging, Communities, Vocabulary, Evolution*. 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada. Retrieved January 31, 2008 from <http://doi.acm.org/10.1145/1180875.1180904>.
- Shen, K., & Wu, L. (2005). Folksonomy as a Complex Network. *arXiv: Computer Science*. from <http://arxiv.org/pdf/cs.IR/0509072>
- Shirky, C. (2002). Weblogs and the Mass Amateurization of Publishing. *Clay Shirky's Writings About the Internet [blog]* Retrieved November 28, 2007 from http://shirky.com/writings/weblogs_publishing.html
- Shirky, C. (2003). Power Laws, Weblogs, and Inequality. *Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community [blog]* Retrieved December 5, 2007 from http://www.shirky.com/writings/powerlaw_weblog.html
- Shirky, C. (2004, August 25, 2004). Folksonomy. *Many2Many [blog]* Retrieved december 11 from <http://www.corante.com/many/archives/2004/08/25/folksonomy.php>
- Shirky, C. (2005a). Folksonomies are a forced move: a response to Liz. *Many 2 Many [blog]* from http://many.corante.com/archives/2005/01/22/folksonomies_are_a_forced_move_a_response_to_liz.php
- Shirky, C. (2005b). Ontology is Overrated: Categories, Links, and Tags. *Clay Shirky's Writings About the Internet: Economics & Culture, Media & Community [blog]* Retrieved April 26 from http://www.shirky.com/writings/ontology_overrated.html
- Shirky, C. (2005c, January 24, 2005). Tags != folksonomies && Tags != flat name spaces. *Many2Many [blog]* Retrieved January 2 from

- http://www.corante.com/many/archives/2005/01/24/tags_folksonomies_tags_flat_name_spaces.php
- sinha, r. (2005a, September 27, 2005). A cognitive analysis of tagging, (or how the lower cognitive cost of tagging makes it popular). *rashmi sinha, thoughts on technology, design & cognition [blog]* Retrieved December 22 from http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html#000143
- sinha, r. (2005b, August 4, 2005). Clustering comes to Flickr. *rashmi sinha: thoughts on technology, design and cognition [blog]* Retrieved December 22 from http://www.rashmisinha.com/archives/05_08/flickr-clusters.html
- Smeaton, A. F., & Quigley, I. (1996). Experiments on Using Semantic Distances Between Words in Image Caption Retrieval. *Research and Development in Information Retrieval*. from <http://ucdata.berkeley.edu:7101/sigir-2002/sigir2002CLIR-14-sanderson.pdf>
- Smith, G. (2004). Folksonomy: social classification. *Atomiq / information Architecture [blog]* Retrieved December 15 from http://atomiq.org/archives/2004/08/folksonomy_social_classification.html
- Smith, G. (2005a). Tagging tags to make synonyms. *Atomiq [blog]*. Retrieved December 23, 2005 from http://atomiq.org/archives/2005/10/tagging_tags_to_make_synonyms.html
- Smith, G. (2005b). The Year in Tags. *You're It! [a blog on tagging]*. Retrieved November 26, 2007 from <http://tagsonomy.com/index.php/the-year-in-tags>
- Smith, T. (2007). Cataloging and You: Measuring the Efficacy of a Folksonomy for Subject Analysis. 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research,, Milwaukee, Wisconsin, USA. J. Lussky (Ed.), Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/2061/01/Smith%5FUUpdated.doc>.
- Sood, S., Owsley, S., Hammond, K., & Birnbaum, L. (2007). *TagAssist: Automatic Tag Suggestion for Blog Posts*. Weblogs and Social Media, Boulder, Colorado, USA. Retrieved December 5, 2007 from <http://www.icwsm.org/papers/2--Sood-Owsley-Hammond-Birnbaum.pdf>.
- Spalding, T. (2007). When tags work and when they don't: Amazon and LibraryThing. *Thingology Blog*. Retrieved November 19, 2007 from

- <http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php>
- Speroni di Fenizio, P. (2005). On Tag Clouds, Metric, Tag Sets and Power Laws. *P.S. [blog]* Retrieved November 24, 2007 from <http://blog.pietrosperoni.it/2005/05/25/tag-clouds-metric/>
- Star, S. L. (1996, February 16-17, 1996). Slouching towards infrastructure. *Social Aspects of Digital Libraries Workshop* Retrieved December 27 from <http://is.gseis.ucla.edu/research/dl/star.html>
- Sterling, B. (2005). Order Out of Chaos. *Wired*, 13(04). Retrieved January 31, 2008 from <http://www.wired.com/wired/archive/13.04/view.html?pg=4>
- Stock, W. G., & Peters, I. (2007). *Folksonomy and Information Retrieval*. Joining Research and Practice: Social Computing and Information Science. American Society for Information Science and Technology, 2007, Milwaukee, Wisconsin. Retrieved January 31, 2008 from http://www.walt.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1194344432asist_am07.pdf.
- Styles, C. (2006). How Web 2.0 will change history - Possible futures for websites of the National Archives of Australia. Australian Historical Association, Canberra Retrieved November 21, 2007 from <http://maeg.textdriven.com/wp-content/uploads/2006/09/Web2+history.pdf>.
- Tennis, J. T. (2006). Social tagging and the next steps for indexing. 17th ASIS&T Classification Research Workshop; Advances in Classification Research, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/2091/01/sigcr-06tennis.pdf>.
- Terdiman, D. (2005). Folksonomies Tap People Power. *Wired News*. from <http://www.wired.com/news/technology/1,66456-0.html>
- The Metropolitan Museum of Art, Indianapolis Museum of Art, Chun, S., Stein, R., & Trant, J. (2006). Researching Social Tagging and Folksonomy in Art Museums: U.S. Institute of Museum and Library Services (IMLS). Retrieved January 31, 2008 from <http://www.imls.gov/applicants/samples/NLG%20RD%20Sample%20The%20Met.pdf>.
- Todras-Whitehill, E. (2005, 5 October). 'Folksonomy' Carries Classifieds Beyond SWF and For Sale. *The New York Times* Retrieved January 31, 2008 from

- <http://www.nytimes.com/2005/10/05/technology/techspecial/05ethan.html?ei=5090&en=0fbe9805541dc103&ex=1286164800&adxnnl=1&partner=rssuserland&emc=rss&adxnnlx=1134504875-DAJrygaL0tIo9mfbD84AJQ>.
- Tonkin, E. (2006a). Folksonomies: The Fall and Rise of Plain-text tagging. *Ariadne*, 47(April). from <http://www.ariadne.ac.uk/issue47/tonkin/>
- Tonkin, E. (2006b). Searching the long tail: Hidden structure in social tagging. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved December 12, 2007 from <http://dlist.sir.arizona.edu/1791/01/tonkin.pdf>.
- Tonkin, E. (2007). Signal and Noise: Social Construction and Representation. 18th Workshop of the American Society for Information Science and Technology Special Interest Group in Classification Research,, Milwaukee, Wisconsin, USA. J. Lussky (Ed.), Retrieved January 31, 2008 from <http://dlist.sir.arizona.edu/2058/01/Tonkin%5F%5Fasist2007.doc>.
- Trant, J. (2006). Social Classification and Folksonomy in Art Museums: early data from the steve.museum tagger prototype. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, TX. J. Furner & J. T. Tennis (Eds.). Retrieved January 31, 2008 from <http://www.archimuse.com/papers/asist-CR-steve-0611.pdf>.
- Trevino, E. M. (2005, November 15, 2006). Three Ways to View Social Bookmarking Systems. *technology and the social [blog]* Retrieved December 24 from <http://blog.erickamenchen.net/2005/11/15/three-ways-to-view-social-bookmarking-systems/>
- Trevino, E. M. (2006). *Social Bookmarks: Personal Organization and Collective Discovery on the Web*. Unpublished Masters, University of Illinois at Chicago, Chicago, Illinois.
- Udell, J. (2005). Jon Udell: language evolution in del.icio.us. Retrieved December 15, 2005 from <http://weblog.infoworld.com/udell/gems/delicious.html>
- Van Setten, M., Brussee, R., Van Vliet, H., Gazendam, L., van Houten, Y., & Veenstra, M. (2006). On the Importance of "Who Tagged What". Workshop on the Social Navigation and Community-Based Adaptation Technologies, In Conjunction with Adaptive Hypermedia and Adaptive Web-Based Systems (AH'06), Dublin, Ireland Retrieved November 22, 2007 from http://www.sis.pitt.edu/~paws/SNC_BAT06/crc/vansetten.pdf.

- Vander Wal, T. (2005a, February 21, 2005). Explaining and Showing Broad and Narrow Folksonomies. *Off the Top [blog]* Retrieved January 3 from http://www.personalinfocloud.com/2005/02/explaining_and_.html
- Vander Wal, T. (2005b). Folksonomy Definition and Wikipedia. *Off the Top [blog]* from <http://www.vanderwal.net/random/entrysel.php?blog=1750>
- Vander Wal, T. (2005c, January 18). Folksonomy Explanations. *Off the Top [blog]* Retrieved November 23, 2007 from <http://www.vanderwal.net/random/entrysel.php?blog=1622>
- Vander Wal, T. (2005d, July 1). Tagging for Fun and Finding. *OK / Cancel [blog]* Retrieved November 22
- Voorbij, H. J. (1998). Title Keywords and Subject Descriptors: A Comparison of Subject Search Entries of Books in the Humanities and Social Sciences. *Journal of Documentation*, 54(4), 466-476. from <http://www.emeraldinsight.com.myaccess.library.utoronto.ca/10.1108/EUM0000000007178>
- Voss, J. (2007). Tagging, Folksonomy & Co – Renaissance of Manual Indexing. 10th international Symposium for Information Science Retrieved January 31, 2008 from <http://eprints.rclis.org/archive/00010165/01/Tagging-Voss-ISI2007.pdf>.
- Walker, J. (2005). Feral Hypertext: When Hypertext Literature Escapes Control. Hypertext 2005 - HT05, Salzburg, Austria.
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, 16(4), 409–421. Retrieved January 31, 2008 from <http://web.gsm.uci.edu/~dobstfel/Articles/Organizing%20and%20the%20Process%20of%20Sensemaking.pdf>
- Weinberger, D. (2005a, March 26, 2005). Controlled and suggested vocabularies: Are tags making us dumb? *Joho the Blog [blog]* Retrieved December 15 from <http://www.hyperorg.com/blogger/mtarchive/003836.html>
- Weinberger, D. (2005b, May 2005). Why Tagging Matters. *Berkman Center for Internet & Society* Retrieved November 20 from <http://cyber.law.harvard.edu/home/uploads/507/07-WhyTaggingMatters.pdf>

- Weinberger, D. (2006a). Beneath the Metadata - a reply. *Joho the Blog [blog]* Retrieved November 20 from http://www.hyperorg.com/blogger/mtarchive/beneath_the_metadata_a_reply.html
- Weinberger, D. (2006b). Folksonomy as Symbol. *Joho the Blog [blog]*. from http://www.hyperorg.com/blogger/mtarchive/folksonomy_as_symbol.html
- Weinberger, D. (2006c). What is a folksonomy anyway? [and comments]. *Joho the Blog [blog]* Retrieved December 7, 2007 from http://www.hyperorg.com/blogger/mtarchive/what_is_a_folksonomy_anyway.html
- Weinberger, D. (2007). *Everything Is Miscellaneous: The Power of the New Digital Disorder*. Times Books.
- Winget, M. (2006). User-defined classification on the online photo sharing site Flickr ... Or, How I learned to stop worrying and love the million typing monkeys. *Advances in Classification Research*, Volume 17; Proceedings of the 17th ASIS&T Classification Research Workshop, Austin, Texas, USA. J. Furner & J. T. Tennis (Eds.). Retrieved December 12, 2007 from <http://dlist.sir.arizona.edu/1854/01/winget.pdf>.
- Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). *Towards the Semantic Web: Collaborative Tag Suggestions*. World Wide Web 2006 (WWW2006): Collaborative Web Tagging Workshop, Edinburgh, Scotland. Retrieved December 3, 2007 from http://www.ibiblio.org/www_tagging/2006/13.pdf.
- Yew, J., Gibson, F. P., & Teasley, S. D. (2006). Learning by Tagging: The Role of Social Tagging in Group Knowledge Formation *MERLOT Journal of Online Learning and Teaching*, 2(4), 275-285. from <http://jolt.merlot.org/vol2no4/yew.pdf>
- Zeldman, J. (2005a, 4 May 2005). Remove Forebrain and Serve: Tag Clouds II. *Jeffrey Zeldman Presents The Daily Report [blog]* Retrieved December 23 from <http://www.zeldman.com/daily/0505a.shtml>
- Zeldman, J. (2005b, 19 April 2005). Tag Clouds are the New Mullets. *Jeffrey Zeldman Presents The Daily Report [blog]* Retrieved November 28, 2007 from <http://www.zeldman.com/daily/0505a.shtml>
- Zollers, A. (2007). Emerging Motivations for Tagging: Expression, Performance and Activism. Tagging and Metadata for Social Information Organization; a workshop at WWW2007, Banff, Alberta, Canada Retrieved December 5, 2007 from http://www.ibiblio.org/www_tagging/2007/paper_55.pdf.