Jorgensen, C. (2007). Image access, the semantic gap, and social tagging as a paradigm shift. 18th Annual ASIS SIG/CR Classification Research Workshop doi: 10.7152/acro.v18i1.12868

Image Access, the Semantic Gap, and Social Tagging as a Paradigm Shift

Corinne Jörgensen, Florida State University

Classification Research Workshop, 2007

Abstract

The recent phenomenon of "social tagging" or "distributed indexing" raises a number of questions regarding long-held beliefs and practices of the classification and indexing community. This workshop paper covers several of these issues, such as locus of authority, control, and meaning, and suggests we may be observing the emergence of a new paradigm of knowledge organization.

The Semantic Gap

The "semantic gap" is mentioned frequently in the literature of image access. The term originated in computer science (Smeulders et al., 2000) and is still used in the CS literature today to refer to the difference between two descriptions of an object using different languages, specifically the difference between a human-readable description and a computational representation. In a computational representation, a simple image of an object moves from the level of individual pixels to assemblages of image primitives such as color, shape/region, and texture, to the assemblage and recognition of an object, at least at the level of a simple "basic object." Object recognition necessitates a level of "understanding" of what is being represented; this is achieved by inferring what different combinations of primitives may represent, e.g. black spots or black stripes on an orange-tan background and an assemblage of potential "leg," "body," "tail," and "head" shapes, perhaps combined with "nature colors," could be interpreted as a leopard or tiger. The process is fraught with stumbling blocks such as occlusion, angle of view, scale, shadow, and lack of uniqueness, to mention a few.

However, it is at the level of object recognition that human image description often begins. With the development of automated methods of content-based image retrieval the term "semantic gap" has come to refer to the larger issue of the gap between these image primitives, or low-level features, and the context-sensitive meanings human beings associate with these. This brings us beyond object recognition and understanding into more abstract levels of semantic meaning, and the meanings or emotions associated with even one image can be many, and can vary across time and place. For a human, recognition of familiar objects is instantaneous, and an image of a tiger, once recognized, can represent multiple concepts such power, ferocity, freedom (or a lack thereof, as in a caged tiger), or even endangered species. These concepts form a gestalt of the object, gestalt being a German word roughly translated as a complete pattern or configuration. There are three parts to a definition of gestalt: a thing, its context or environment, and the relationship between them (Wymore 2002). Studies in cognitive science suggest that this gestalt may have equal importance with sensory stimuli in the process of actual recognition of the object.

Objects by far make up the largest portion of image content for pictorial images. On a physical level, objects can often be broken down into smaller component parts, a useful paradigm for the bottom-up process of automated object recognition. For example, an image of an object as simple as an apple has basic primitives such as color *color* (red),

ISSN: 2324-9773

shape (round), and *texture* (smooth, shiny, hard) and is composed of *parts* such as skin, flesh, seeds, and stem. Objects are most frequently named at the "basic level," (Rosch, 1978) that is, the level at which objects within a category will share the maximum number of attributes and at which the objects in the category will be maximally different from objects in other categories. Some progress has been made at bridging the semantic gap here, but assembling object primitives into understandable objects is much more successful within limited domains and processes, such as manufacturing and quality control. Objects, such as apples, also have associated contexts: *uses*: ingredient (pie, applesauce); *weapon* or *target*; *activities* (eating, cooking, bobbing for); *histories* or *stories* (discovery of gravity); or *meanings* (temptation, poison).

As of yet we do not have systems powerful enough or organizations with sufficient resources to bridge the semantic gap: in other words, to permit image indexing to be as detailed or as comprehensive as research is beginning to reveal human image description and searching can be. Translating a human query into one that can be handled by automated methods is beyond most users' abilities, with the problem only exacerbated by some of the search interfaces that have been created. Additionally, the gap between a human's image descriptions and those within an IR system has become a reality that is compounded by today's unmediated, cross-collection, and cross-culture searching of the web or of large digital repositories. Combined with the multivalent nature of images, bridging this gap sometimes seems an intractable problem.

Issues of Authority and Control

From the system side, the ability to provide access to information bestows the provider with a certain amount of power beyond that of merely controlling physical access: the power to name, the power to filter, the power to control the information context, and thus the power to shape the perception of reality. Classification systems and algorithms both have the capacity to mirror and instantiate the assumptions, beliefs, and desires of a group or society (Olsen, 1998; Fleischmann and Wallace, 2005), and once created, they are slow to change, as the paradigms of their creation underlie everything that is built upon them. Change of such a system requires not only new definition and understanding but also a reordering of concepts and shifts in the balance of power.

Indexing and classification has traditionally been the work of a community of educated professionals. Standard text-based methods of indexing or classifying images (controlled vocabulary and thesauri) have emphasized the importance of authority and consistency in description, while automated systems are also constrained by explicit and implicit rule-based methods. There are many benefits to these approaches as they overcome a number of semantic gaps that are created by issues with synonyms, homonyms, and heteronyms (Macgregor and McCulloch, 2006). However, there are other semantic gaps that exist, between groups with different needs, goals, and interests, between providers and users. This has been addressed most extensively in the museum literature, and especially in the art museum literature, as the audiences (both amateur and professional) for arts and cultural heritage objects and images have vastly different experiences, training, and levels of interaction with these objects and images.

These traditional systems are in marked contrast to more recent innovations (e.g. Flickr, www.flickr.com) that permit spontaneous social tagging (effectively, distributed and

"democratic" indexing) of images by a larger community. In these systems, the lines blur between providers and users, and between individual and collective uses, and social tagging and the folksonomic approach have a number of advantages that are seen as a means to overcoming a variety of semantic gaps (Kroski, 2005).

Thus, in addition to manual indexing and automated methods we now have a third approach to image access, that of a wide range of users engaged in social tagging processes. These projects vary in the amount of control exerted over the process, and range from minimal control to attempts to insure a standard of tagging through control of who is engaged in the process. Several projects are currently underway which are facilitating user tagging of images and analyzing the products of this tagging, and as a result non-traditional and creative uses of scholarly and consumer image collections, such as storytelling and reminiscence, are beginning to emerge (Trant, 2007). Additionally, when there is no pre-imposed authority structure controlling what is indexed or how, subgroups form which create their own criteria for inclusion and authority (Stvilia and Jörgensen, 2007). This expands the range and scope both of materials indexed and the range of attributes that are addressed in the indexing process.

What do We Know about Language?

At this point, we might ask, what does the vocabulary of tagging look like? What and how can tags contribute to the vocabulary of description? Are the results comparable to the short term or postulated long-term effect of a million monkeys with typewriters (the infinite monkey theorem)[1]?

Several recent studies have used available data and analyzed the characteristics of language used in social tagging. Among the results, image tags generated in this way appear to follow the characteristics of a Power Law[2] in the form of the Zipf distribution of a natural language corpus (Mathes, 2004; Guy and Tonkin, 2006). The "Long Tail" of the Zipf distribution is being viewed economically as an opportunity for niche markets and philosophically as a catalyst for creativity and diversity. In terms of vocabulary, the Long Tail contains infrequently used words, as compared to the peak of the most commonly used words, thus the Long Tail is seen as having the potential to expand indexing, and therefore retrieval.

However, from an IR standpoint, Luhn's (1958) model proposes that in fact the mid-range terms are the best index terms and relevance discriminators, not the very infrequent words of the long tail. While natural language composes a searcher's query, indexing languages typically employ highly precise and specific terms relevant to the community that uses the indexing language. This suggests that a closer look at the vocabulary generated in the tagging process might be useful to understanding and bridging the

[1] For a thorough discussion of this theorems origins and occurrences, as well as the underlying probability theory, see the Wikipedia entry available at http://en.wikipedia.org/wiki/The_Total_Library

[2] The Power Law states that the frequency of an item is inversely proportional to its rank. This has been found to hold true for such divergent statistics as populations of cities, words in a corpus, and size of web sites. The cause and meaning of this phenomenon is not known but is theorized that it may be a possible statistical artifact.

"semantic gap" between current indexing vocabularies and user's natural language queries.

One recent study uses the NISO guidelines (NISO, 2005) pertaining to the choice and formation of concept terms for thesaurus construction within three tagging systems, De.licio.us, Furl, and Technorati, as a benchmark with which to evaluate their tag structure (Spiteri, 2007). Although the majority of terms within the sample conformed to the guidelines regarding the types of concepts, the use of single tags, the predominance of nouns, the use of recognized spelling, and the use of primarily alphabetic characters, there remained problems with ambiguity. This research concludes that in order to integrate tags within standard systems such as library catalogs, clear recommendations for tag choice and formation be established.

Thus, there remains a number of issues related to the role of established tools such as controlled vocabulary in relation to social tagging, and recommendations for "improving" tagging usually require constraints on how tags are formed, moving tags closer to a controlled vocabulary. Social tagging, as a distributed activity, is also being viewed as a relatively inexpensive way to increase access points to collections (and also likely to be as unstoppable as Google).[3] However, specifying rules for tag formation will increase the effort associated with tagging, and thus could increase cost and lessen tag production.

Where is the Research?

As tagging is such a recent phenomenon, research is only beginning to reach the publication stage and much of the discussion before 2006 is found on the web rather than in the print journal literature. There are a number of assumptions that appear in discussions of tagging questioning its usefulness. For instance, it is widely assumed that tagging will improve recall, but can only hurt precision. Beyond the fact that this inverse relation has not been empirically proven, we should remember that precision and recall themselves are only two (somewhat disputed) measures of retrieval effectiveness and apply well only to specific types of queries.

Another set of assumptions is the effect of a lack of authority control on the consistency and "goodness" of tags.[4] Interestingly, at the same time there has been, in the literature, an increase in both defense of controlled vocabularies against the competition of tags, and calls to "train the user" (to create the "right" kind of tags), a response well known in times of perceived threats to entrenched systems. There is also tacit acknowledgement of the need to placate the current authority, the controlled vocabulary community: "Optimisation of user tag input, to improve their quality for the purposes of later reuse as searchable keywords, would increase the perceived value of the folksonomic tag approach" (Guy and Tonkin, 2006).

---

[3] The author is aware, and the reader should note, that both the "social tagging" activity and many of the terms associated with this activity (for instance, whether it is really "social" tagging) are still in flux and the subject of much debate, most of which is taking place on the Internet. The phenomenon is so new that only recently have articles on it begun to appear in the journal literature, and most research is very preliminary at this point. Nevertheless, the activity of social tagging has grown tremendously in a short time.

[4] Here, we see an example of the power of naming: "tags," as opposed to the more formal terms used for controlled vocabulary items, such as descriptors or preferred terms).

These assumptions are associated with a particular understanding of how to effectively mine and retrieve user-generated tags and are based on the controlled vocabulary paradigm: the only "good" tag is a controlled tag, i.e. one which lends itself well to a specified community and method of retrieval. Here we see a sharp contrast between academic discussions and commercial practice; while a number of Google-like search engines are thriving, searchers are less and less inclined to train themselves to use traditional authoritative indexing and retrieval systems, and use of these resources is declining. Why should users trade a process that allows spontaneity and even fun for one which requires effort and seriousness, not to mention learning?

Research in new phenomena needs to investigate not only the phenomena but also the assumptions upon which the research is based. With the invention of the printing press, the church recognized that the populace needed to be literate to benefit from the widespread production of printed materials; i.e. in order to maintain its authority, the populace needed to be able to read the church's religious teachings. Ironically, we know that the spread of literacy historically created multiple conflicting authorities. A parallel can be drawn to the emergence of the Internet in phenomena ranging from blogs to social tagging. The underlying thread in much discussion of the Internet is the need to train users: to cite, to search, to discriminate, to tag, to fit in with a particular community's current model of authoritative sources, which often derives from centralized (and profitable) publication of print materials. The current paradigm is controlling the production and practice of indexing, rather than eliciting new types of indexing behaviors and new participants in the process. Distributed *description* and *annotation of documents* and distributed *collection building* have the potential to stimulate distributed *knowledge creation* (Jörgensen, 2004). We need to ask what paradigms this could possibly threaten, and place investigation of social tagging within the larger contexts of those paradigms.

Is there a Middle Ground?

As we now see, in addition to content-based (using computer algorithms to describe low-level features) and concept-based (using human intelligence to assign higher-level descriptors) indexing (Rasmussen, 1997), there is now a third alternative available for providing access to images, that of social tagging or cooperative indexing.[5] While social tagging still may be done by an individual in isolation, the tags associated with an image may be contributed by several sources, and recent observations indicate that over time a consensus forms around the tags assigned to an item (Mathes, 2004). We could also refer to this as consensus-based indexing.

While one can't meaningfully create a distribution for a controlled vocabulary (as it would be a flat line since each unique term occurs once), preliminary research demonstrates that the terms from social tagging appear to conform to a Zipf distribution, and the long tail of the distribution is assumed to be where terms capable of the most discrimination occur. A quick sample of 164,663 terms collected from Flickr (Table 1)[6]

[5] In addition to social or voluntary tagging systems, at least one system, Mechanical Turk (www.mturk.com), is experimenting with compensating distributed work such as indexing, writing, and locating information (this has become known as "crowdsourcing").

[6] In September 2006 a random sample of 3,000 photos and their associated descriptions was drawn from the Flickr Recent Photos page using scripts developed in the Java language and a Flickr Java API.

shows what would appear to be useful[7] indexing terms appearing in the long tail all the way down to single word occurrences (which are by and large much less useful as they are often numbers or concatenations of terms). Interestingly, 75% of the 21 most frequently occurring terms were labels that have also seen some success in content-based processing (e.g. sunset, tree, sky, water, flower, clouds, and several colors), supporting the idea that tags could be used as training data for a machine learning environment.

Table 1. Common terms in the "long tail" and their frequency.

Count	Terms
10	pelicans, nativity, thread
9	charcoal, pudding, nurse
8	cheerleader, tears, pigtails
7	clowns, mosquito, snowfall
6	wizard, chasing, cobblestones
5	despair, dimples, estuary
4	clam, coffeehouse, dynamite
3	cages, key, tailor
2	beetles, hothouse, sponges

There are several projects that have focused on integrating tags and image features in a machine learning environment as a means to improve indexing. Hauptman and others (2007) report using this method in multilingual broadcast news indexing and retrieval as part of the TRECVID 2006 Workshop. The research used a "light" ontology of 39 concepts derived from the Large-Scale Concept Ontology for Multimedia (LSCOM)(Naphade et al., 2006), a previous project which developed use cases in the broadcast news area and a larger ontology of 834 concepts using a variety of thesauri and controlled vocabularies as source material.

Aurnhammer, Hanappe, and Steels (2006) propose combining collaborative tagging and visual feature analysis as a way to overcome the problems of each method used alone. Their preliminary work suggests that the addition of visual features can overcome the inherent problems of tagging (ambiguity resulting from synonyms, homonyms, and so forth), and tags are demonstrated to be useful in restricting the search space, supporting classification on visual features. They note that their method does not assume that a tag corresponds to a category, as tags are contributed by multiple users.

[7] "Useful" being defined primarily as common occurring nouns, many at the "basic level" of object description.

The idea of integrating collaborative tagging and visual features derives from the concept of "emergent semantics," where the meaning of an image emerges in the interaction between it and the user, and between it and the context it is placed in, such as the particular image collection or set of returned hits. As Santini (2002) explains:

> In this sense, the goal of the interaction between the user and database is not so much to retrieve images based on a preexisting semantics but to create image semantics. The interaction itself is not configured as a query but as a navigation in which the user dictates similarities and associations between images and, through this activity, reorganizes the database to embody the desired semantic (p. 81).

Meaning in images has not been addressed by most visual indexing vocabularies,[8] as meaning has been considered too subjective and changeable to be a reliable access point. Emergent semantics turns this restriction into an enabler, by allowing the creation of meaning in interaction and capturing this meaning for others to access.

Is There an Ontology of Images?

Many years ago a colleague posed this query to the author. We now know, in fact, that there are multiple ontologies of images. Each vocabulary created for image indexing carries the assumptions and desires of a particular community as well as its own particular knowledge, and each vocabulary creates its own authority and world of meaning. The problem has been with the inherent constraints resulting from having to choose from among (and within) vocabularies and working within limits on the record structures that carry these vocabulary terms, as well as within organizational constraints such as the amount of time to devote to provision of access.

New paradigms are taking shape. The act of tagging embodies the concepts of interactivity, connectivity, and access, which are characteristic of the digital age, and which presume the user to be capable and seeking connection, rather than needing guidance and protection (Dresang, 1999). The concept of emergent semantics places image context on equal footing with image content in determining what a viewer sees in an image as well as the meaning of an image, and supplements the concept of consistency with the concept of particularity, as well as allowing a communicative structure within which new worlds of shared meaning can emerge. The challenge for the research community is to recognize old assumptions and understand new paradigms when framing research in this dynamically evolving area of image access.

[8] One exception is IconClass, which addresses meaning related to a stable iconography.

Bibliography

Aurnhammer, M., P. Hanappe, and L. Steels (2006). Integrating collaborative tagging and emergent semantics for image retrieval. Proceedings WWW2006, Collaborative Web Tagging Workshop, May 2006. Available at: http://www.isrl.uiuc.edu/~amag/langev/paper/aurnhammer06semanticsWWW.html (accessed September 1, 2007)

Dresang, E. (1999). *Radical Change: Books for Youth in a Digital Age*. Bronx, NY: H.W. Wilson.

Fleischmann, K. A. and W. A. Wallace (2005). A covenant with transparency: opening the black box of models. *Communications of the ACM* 48(5): pp. 93-97.

Guy, M. and E. Tonkin (2006). Folksonomies: Tidying up tags? *D-Lib Magazine*, 12(1), January. Available at: http://www.dlib.org/dlib/january06/guy/01guy.html (accessed September 1, 2007).

Hauptmann, A.G., M.-Y. Chen, M. Christel, D. Das, W.-H. Lin, R. Yan, J. Yang, G. Backfried and X. Wu (2007). Multi-lingual broadcast news retrieval. Available at http://www-nlpir.nist.gov/projects/tvpubs/tv6.papers/cmu.pdf (accessed September 1, 2007)

Jörgensen, C. (2004) Unlocking the museum: A manifesto. *Journal of the American Society for Information Science and Technology* 55 (5): 462–464.

Kroski, E. (2006). The hive mind: Folksonomies and user-based tagging. Available at: http://infotangle.blogsome.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/ (accessed September 1, 2007).

Luhn, H. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, *2*(2), 159–16.

Mathes, A. (2004). Folksonomies - cooperative classification and communication through shared metadata. Available at: http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed September 1, 2007).

Macgregor, G. and E. McCulloch (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(5): pp. 291-300.

Naphade, M., J. R. Smith, J. Tesic, S.-F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, J. Curtis (2006). Large-scale concept ontology for multimedia, *IEEE MultiMedia* 13(3): pp. 86-91.

National Information Standards Organization (2005). *Guidelines for the construction, format, and management of monolingual controlled vocabularies*. ANSI/NISO Z39.19-2005. Bethesda, MD: National Information Standards Organization

Olson, H. A. (1998). Mapping beyond Dewey's boundaries: Constructing classificatory space for marginalized knowledge domains. In Geoffrey C. Bowker and Susan Leigh Star, eds., *How Classifications Work: Problems and Challenges in an Electronic Age*, a special issue of *Library Trends* 47(2): 233-254.

Rasmussen, E. M. (1997). Indexing images. *Annual Review of Information Science and Technology* 32: pp. 169-196.

Rosch, E. (1978) "Principles of categorization," in Rosch, E. & Lloyd, B.B. (eds), *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 27–48.

Santini, S. (2002). Image retrieval. *IEEE Intelligent Systems* (January/February): pp. 79-81.

Smeulders, A. W. M., M. Worring, S. Santini, A. Gupta, and R. Jain (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(12): pp. 1349–1380.

Spiteri, L. F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology* 4(2): Article 41. Available at: http://www.webology.ir/2007/v4n2/a41.html (accessed September 1, 2007).

Staab, S., editor (2002). Emergent semantics. *IEEE Intelligent Systems* (January/February): pp. 78-86.

Stvilia, B. and C. Jörgensen (2007). End-user collection building behavior in Flickr. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*, Oct. 19-24 (forthcoming).

Trant, J., with the participants in the Steve.Museum Project (2007). Exploring the potential for social tagging and folksonomy in art museums: Proof of concept. *New Review of Hypermedia and Multimedia*, 12(1): pp. 83–105.

Wymore, J. (2002). Gestalten. *Gestalt!* 6(2). Available at: http://www.g-gej.org/6-2/gestalten.html (accessed September 1, 2007).

ISSN: 2324-9773