

Tagging As A Communication Device: The Impact of Communities on Transforming Tag Information

Co-Sponsored by SIG-TAG, SIG-KM and SIG-CR

Scope

The panel will discuss tagging of documents where a particular vocabulary, language or classification is used for communication. At the individual or conceptual level, tags avoids some of the issues of fossilized terms or meanings, but at the context, language or category level, the meaning must be of a more community or social network nature. There is even a 'tag' to 'tag' relationship where the "to" object may either be a user or information. Therefore, tags as a language can be tailored to improve communication accuracy for the object. The panel will attempt to show how tagging can indicate an appropriate semantics given the user's understanding of the information's context, and why one tag was chosen over another.

The topic areas proposed for this conference panel are:

Ontology of Tags: Do users communities use ontologies and tagging both as metadata for information?

Metadata use: Do small communities have a stable use of terminology?

Tag Clouds: What did I mean by that tag the 1st time I used it? Is this how I mean to use this tag?

Tags as Communication Device: Is tagging an activity to integrate an individual's past, present and (anticipated) future experience?

Questions that will be discussed among panelists and with the audience will focus on looking at tags as a way of communicating information between users. Within a community, is it possible to decide whether to interpret a tag within the context of the document or just treat it as a primary key or keyword, and what are the feasible/required mechanisms that enable that decision? More questions below.

Moderator/Panelist:

Heather D. Pfeiffer

New Mexico State University, hdp@cs.nmsu.edu

Website: <http://www.cs.nmsu.edu/~hdp/>

Heather Day Pfeiffer is a Post-Doc at New Mexico State University. She is interested in Knowledge Engineering, Database Management, and collaborative testbeds. Heather has presented papers and posters at ICCS, KCAP, VL, ASIST, and various university conferences on topics relating to knowledge bases, databases, Artificial Intelligence and other forms of knowledge engineering. She has also published papers in journals such as JETAI and JMMS. Dr. Pfeiffer has a PhD in Computer Science from New Mexico State University, a MS in Computer Science from New Mexico State University, and a BS in Biology/Microbiology from University of Washington.

Emma Tonkin

UKOLN, University of Bath, e.tonkin@ukoln.ac.uk

Website: <http://www.ukoln.ac.uk/>

Emma Tonkin is an Interoperability Focus Officer working for UKOLN at the University of Bath. Her job often leads her into working directly with end-users of cataloguing systems. As her background is in human-computer interaction, her major interest is in developing an understanding of user behaviour and the factors that influence the uptake and use of these technologies. Her current project is the IEMSR, a metadata schema registry designed to support registration, use and reuse of DC and LOM metadata schemas and profiles. She serves as co-moderator of the Dublin Core Metadata Initiative's Schema Registry community, and as a member of the DCMI advisory board. She is currently pursuing a PhD on the topic of wearable computing at the University of Bristol.

Mark R. Lindner

University of Illinois at Urbana-Champaign, mlindner@uiuc.edu

Website: <http://marklindner.info/>

Mark Lindner is finishing a Certificate of Advance Study (CAS) at GSLIS, University of Illinois at Urbana-Champaign. He worked his way through graduate school as a distance education technology technician, thesaurus maintainer and, most recently, as both a serials cataloging and a rapid monographic cataloging graduate assistant. His focus during his CAS has been on the organization of information, with a specific focus on classificatory structures. The topic of his CAS capstone paper is on language use and theory within LIS and, specifically, what adopting the Integrationist view might mean for LIS. Mark serves on the ASIST Standards Committee and has given an invited talk to the UIUC ASIST Student Chapter on the topic of which LIS blogs a student should be reading. He has a MS in LIS from the University of Illinois at Urbana-Champaign and a BS in Philosophy from Illinois State University.

Margaret E. I. Kipp

College of Information and Computer Science, Long Island University, margaret.kipp@gmail.com

Website: <http://www.eskimo.com/~meik/>

Margaret Kipp is an Instructor in the College of Information and Computer Science, Long Island University and a PhD candidate in the Faculty of Information and Media Studies, University of Western Ontario. She has a background in computer science and worked as a programmer/analyst before moving to information science. Her research interests include classification systems, collaborative web technologies and the creation and visualizations of structures in information organization systems. She has presented papers and posters at ASIST, JCDL, IA Summit and CAIS and published papers in Scientometrics. She is currently working on her thesis which examines information organization on the Internet, specifically concentrating on social tagging and social bookmarking.

David R. Millen

IBM T.J. Watson Research, david_r_millen@us.ibm.com

Website: http://domino.research.ibm.com/cambridge/research.nsf/pages/david_r_millen.html

David R Millen is a research manager in the Collaborative User Experience group at IBM Research in Cambridge, MA. His group develops new social software applications, and explores the social, business, and technological implications of these new tools through field studies with small teams and communities. Recent projects include activity-centric collaboration, social bookmarking (dogear), and an enterprise social networking service (beehive). Recent research includes the studies of collaborative tagging in social search and organizational patterns in social tagging. Dr. Millen has a BA from Columbia University and a PhD in Cognitive Psychology from Rutgers University.

Each Panelist's Contribution:

Tagging as Metadata: Ontological Architecture of Tags (Heather D. Pfeiffer)

Knowledge Representation, KR, is how knowledge is represented within a computer for processing, but some definitions of *knowledge* are the "perception of true" or "the fact or state of knowing." Within a community, knowledge is communicated to the individuals through language, and semantics of this language can be conceptually seen as an *ontology*. The concepts within the ontology can be represented in a hierarchical structure, and both this representation and concept are socially defined by the community building the ontology. Guarino saw this ontological level outside the KR for the machine, and as metadata for the representation. Tags placed on a document by a user community giving a context for the document. These tags are the "jargon" or keywords within the community's social network, and are metadata relating this document to other documents with the same context. In this way, the metadata can be seen as conceptual tags given to the KR designed by the community defining context. The architecture of the ontology or its hierarchical links is then built from these tags. However, the ontology can be modified by other users within the domain community.

Pfeiffer, H.D. & Pfeiffer, Jr., J.J. (2007) Representation Levels Within Knowledge Representation. In U. Priss, S. Polovina, and R. Hill (eds), *Conceptual Structures: Knowledge Architectures for Smart Applications*, vol. 4604 of LNAI, pages 484-487, ICCS.

Guarino, N. (1994). *The Ontological Level*. *Philosophy and the Cognitive Science*, pages 443–456, Holder-Pivhler-Tempsky, Vienna.

Investigating lexical choice in online communities? (Emma Tonkin)

Social network analysis allows the ways in which groups of people are linked to be studied. This information may permit the study of some of the features of language use in social domains/contexts, as well as the changes that occur through time. Sites that link social tagging and social network analysis in particular allow features of the evolution of negotiated terms to be studied. The role of terminology within a community can be examined - that is, the motivators behind language acquisition as a socialization process. Message formation is related to intended audience, but also to the process of negotiating identity. Connecting sociolinguistics and classification explicitly in active study of identifiable communities online is a step towards quantitative examination of complex phenomena of variation in language use. However, there are many challenges: online community structures are not uniformly well-understood, and there are few mature tools available with which to work; of the extensive body of theory dealing with linguistic development and shift, only a small subset deals with quantitative measures in text corpora.

Tonkin E (2007). *Signal and Noise: Social Construction and Representation*. 18th Annual ASIST SIG/CR Classification Research Workshop.

Integrating tagging (Mark R. Lindner)

Defending particular theories and epistemologies is important, and if they have no practical implications then they are of no consequence (Hjørland 2007). One must argue against a discipline's theoretical and methodological principles if one has issues with them (Hjørland 1997). Roy Harris has been doing just that with orthodox linguistics for over 30 years. Harris is the founder of Integrationism, a radical view of communication and language. What do the orthodox view of linguistics and Integrationism each offer in explaining tagging behavior, or even simply the meaning of tagging as communication? How would an Integrationist view tagging for the individual and for the community? How would they view it diachronically and synchronically?

International Association for the Integrational Study of Language and Communication. "What is Integrationism?" <http://www.integrationists.com/integrationism.html>

Hjørland, B. (2007). *Semantics and Knowledge Organization*. In *Annual Review of Information Science and Technology* (pp. 367-405). Medford, N.J.: Information Today.

Harris, R. (1998). *Introduction to Integrational Linguistics*. Oxford : Pergamon.

Hjørland, B. (1997). *Information Seeking and Subject Representation: An Activity-Theoretical Approach to Information Science*. Westport, Conn: Greenwood Press.

@toread and Cool: Subjective, Affective and Associative Factors in Tagging (Margaret E. I. Kipp)

The idea behind social bookmarking tools is to harness the power of the network effect on the web to create a more useful search system by combining the efforts of users in the assignment of labels (tags) to a bookmarked item. However, many users of del.icio.us, citeulike and connotea appear to want to store more than just the subject of the documents they are bookmarking. Tags such as @toread, tobuy, todo, fun and cool suggest that users see their relationship to these documents in different ways. While the latter tags express an emotional connection to the document, the former show evidence of a desire to attach personal information management information to documents. This desire to combine personal information management and document classification echoes findings in document use research which users categorized items in order to

better understand their relationship to other items and to tasks the users wished to perform. (Sellen and Harper 2002; Kwasnik 1991; Malone 1984) The use of such non subject tags, tags which are deliberately excluded from traditional classification systems due to their potentially temporary or task specific nature, shows that users may see classification as a holistic process closely tied to themselves, their work and their groups.

Kwasnik, B. H. (1991). The Importance of Factors That Are Not Document Attributes in the Organization of Personal Documents. *Journal of Documentation* 47(4), 389-398.

Malone, T. W. (1983). How Do People Organize Their Desks? Implications for the Design of Office Information Systems. *ACM Transactions on Office Information Systems* 1(1), 99-112.

Sellen, A. J. & Harper, R. H. R. (2002). *The Myth of the Paperless Office*. Cambridge, MA: MIT Press.

Patterns of collaborative tagging in a large organization (David R. Millen)

Few empirical studies of social bookmarking within a corporate environment have been reported and relatively little is known about the patterns of actual use. In this presentation, highlights of a field study of social bookmarking and collaborative tagging will be shared from three perspectives. First, personal (or private) tagging will be compared to public (enterprise-wide) tag collections. Second, differences in social tagging among organizational entities within the enterprise will be explored. And finally, tag use, reuse and influence among articulated (friend) communities will be examined.

Thom-Santelli, J., Muller, M.J., & Millen, D. R. (to appear) Social Tagging Roles: Publishers, Evangelists, Leaders. *SIGCHI conference on Human Factors in computing system*. Florence, Italy, April 5-10.

Pan, Y. X., Millen, D. R. (2008) Information Sharing and Patterns of Social Interaction in an Enterprise Social Bookmarking Service. HICSS – January 7-10.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007) Getting our head in the clouds: Toward evaluation studies of tagclouds. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. San Jose, CA, ACM Press.

Millen, D. R., Yang, M., Whittaker, S. and Feinberg, J. (2007) Social bookmarking and exploratory search. *ECSCW 2007*, Limerick, Ireland, Sept 26-28, 2007.

Millen, D. R., J. Feinberg, & Kerr, B. (2006) Dogear: Social bookmarking in the enterprise. *Proceedings of the SIGCHI conference on Human Factors in computing systems*. Montreal, Quebec, Canada, ACM Press.

Moderator's Questions for Panelists:

Could tagging be equal to semi formal or informal metadata?

To what extent is terminology, whether from a natural language, classification or ontology, emergent from a community context? That is, to what extent is it acquired in a top-down process and to what extent is it contributed or developed in a bottom-up process?

Do we see a difference in the integration of tags within a community that are used a single point in time (synchronic) or across time (diachronic)?

Is there a difference in the tagging when the communities are friendly or just associative?