

SOCIAL TAGGING AND THE NEXT STEPS FOR INDEXING

Joseph T. Tennis <jtennis@interchange.ubc.ca>

School of Library, Archival and Information Studies

University of British Columbia

301 - 6190 Agronomy Road

Vancouver, BC V6T 1Z3, Canada

1. Introduction

Social tagging, as a particular type of indexing, has thrown into question the nature of indexing. Is it a democratic process? Can we all benefit from user-created tags? What about the value added by professionals? Employing an evolving *framework analysis*, this paper addresses the question: what is next for indexing? Comparing social tagging and subject cataloguing; this paper identifies the points of similarity and difference that obtain between these two kinds of information organization frameworks. The subsequent comparative analysis of the parts of these frameworks points to the nature of indexing as an authored, personal, situational, and referential act, where differences in discursive placement divide these two species. Furthermore, this act is contingent on implicit and explicit understanding of purpose and tools available. This analysis allows us to outline desiderata for the next steps in indexing.

2. Background

The conceptualization and the act of indexing change in response to its socio-technological environment. Indexing, as the interpretation and representation of significant characteristics of documents for information systems is an act with many different manifestations. Social tagging is a manifestation of indexing based in the open – yet very personal – web. Indexing in this environment presents itself different from professional indexing services done for catalogues and databases. This apparent difference, and the desire to know what the future of indexing holds in light of these developments, begs the question as to how social tagging and subject cataloguing – a LIS professional manifestation of indexing – are similar and how they are different. In order to answer this question, this paper employs a *framework analysis* to compare two different types of indexing: social tagging and subject cataloguing.

Framework analysis is a rubric and a set of questions used to compare different systems, methods, and work processes of information organization frameworks. For the purposes of this paper, an information organization framework consists of information organization structures (classification schemes, taxonomies, ontologies, bibliographic descriptions, etc.) and the work processes involved in maintaining these systems (the act of classifying, the act of tagging, the act of creating a bibliographic description). Each framework sits within a discourse or set of discourses and understanding these is part of the analysis.

Framework analysis posits that all information organization systems and work processes have four elements: purpose, predication, function, and context. Where purpose is defined as an explicit and/or implicit intention of the framework – the reason why the system is built and maintained; a predication is the coordinated operationalization of achieving the purpose of the framework – the link between purpose and function; functions are actions enabled in the framework, and context is the technological and social environment the framework inhabits. Context is a developing concept in Information Science, one that in framework analysis admits of many levels and units of analysis – ranging from tasks to socio-political discourse, and from queries to evolving information needs.

Social tagging and subject cataloguing are examples of information organization frameworks. It can be argued that these frameworks are a type of indexing, because both of them are systems, methods, and work processes that analyze documents and create representations of significant characteristics for inclusion in information systems. However there are differences between these two frameworks. The scope of documents considered for indexing differs. They admit to different purposes, and they fix themselves in different discursive regions. Further, they create different predications and functions in order to make manifest the individual framework. Finally these frameworks operate in different contexts. Social tagging is not built in the same context, with the same tools, by the same methods, or even for the same purposes as subject cataloguing. These differences lead us to examine their epistemic and discursive claims.

3. Rationale

Information Science sees similarities in various information organization frameworks. Vickery (1997), Soergel (1999), have both commented on the similarity between classification and ontology engineering. In this discussion, the comparison often points to a superiority of one type

of knowledge – and by extension one type of framework. This superiority is often rooted in precedence.

However, using a framework analysis that makes explicit the diversity in purpose, predication, function, and context shows us that an appreciation of diversity in these frameworks would allow us to make clearer statements about the effective performance of systems.

Other studies have seen benefit in examining diversity information organization frameworks. Comparative studies of classification have revealed interesting differences between naïve and professional classificationists (Beghtol, 2003). Beghtol, here, is a comparativist that see utility in identifying difference over and against the reinvention discourse. Our work follows the ethos of Beghtol's.

Finally, as is evident from the call for papers for this workshop, Information Science assumes similarities between extant frameworks (indexing systems like subject cataloguing) and social tagging. Beyond the similarity on the act of interpreting and representing documents in an information system, Information Science assumes that people create indexing tools and the work processes around them, for identical purposes, with complementary functionality and in contexts that would not affect differences in the former three.

We posit that this is a problematic position for scholarship on systems, methods, and work practices – problematic for work on information organization frameworks. Where other Information Scientists see co-opted knowledge, and corrupted identity, we see, along with thinkers like Beghtol, a necessary and interesting diversity. It is only by fully understanding this diversity that we can create effective evaluation rubrics for these frameworks – so that research may improve systems design, development, and implementation.

There are two camps: those that see reinvention and those that see difference.

4. Species of Indexing: Subject Cataloguing and Social Tagging

Indexing is, minimally, the analysis of documents for their significant characteristics in order to represent those characteristics in an information system (Langridge, 1989). This definition can be expanded, in order to highlight similarities and differences between various acts of analysis and

representation. Such a definition might look like this. Indexing is an act where an indexer in a particular context, goes through a process of analyzing a document for its significant characteristics, using some tools to represent those characteristics in an information system for a user.

These two species of indexing can be compared along two lines of manifestation: a prescriptive (textbook) manifestation and descriptive (observed) manifestation. Prescriptively subject cataloguing manifests as a practice that identifies users' needs for finding and collocating stock in a library by subject. In order to fulfill those needs, subject cataloguing uses a list of subject headings that are precoordinated for specific entry. The descriptive manifestation of subject cataloguing is a bit different. In Sauperl's work she identified three meanings that were interpreted: user's, author's, and cataloguer's, and found that the last of these three often found its way into the catalogue. This was due to the nature of collocation and extant collections as represented in the catalogue.

Social tagging does not have a *textbook* manifestation. The best we can do to attribute prescriptive manifestation to social tagging is to look at purposes of systems. Tagging systems are built to enable: sharing and managing citations, photos, and web pages. However, much of the sharing is done through observing someone else's personal tagging practices or through natural language (tag) use. Some of the tags used in tagging systems are idiosyncratic and only meaningful to the individual's interaction with the material indexed. As a result, tagging systems have tags like "todo", "tobuy", "want", "don't have", and "7.20.06 AIDS Vaccine - Design.Immunogenicity.Efficacy." These tags reflect significance in relation to tasks (buying, etc.) and sorting or differentiating between other tags (dates appended to AIDS Vaccine...).

From these two manifestations, prescriptive and descriptive, we can see points of departure, but not many areas of overlap. Tagging seems intensely personal, whereas subject cataloguing is an act of delegation mediated by institutions (the library and the Library of Congress Subject Headings). A more thorough analysis would offer us insight into the similarities as well as the differences that obtain between these two acts of indexing.

5. Discourse, Work Process, and Structure

In order to understand the similarities and differences between social tagging and subject cataloguing we examined three areas of indexing work. This process required a lens. We found that lens in Ron Day's *Fordist critique* (Day, 2001), Tennis's work on document interpretation processes (Tennis, 2005), and Jacob's rubric of vocabulary structures (Jacob, 2000). Taken together these three offer us the fodder for a comparative framework.

6. Framework Analysis

In order to thoroughly compare different species of indexing we need a common rubric or framework. This framework should be able to highlight the similarities and differences that obtain among species of indexing. What follows is an emerging framework analysis that compares the (1) processes, (2) structures, of indexing and (3) the contexts in which these species of indexing occur. These together comprise the rubric. I will introduce this rubric, and then present the results of using it to compare social tagging and subject cataloguing.

The process of indexing, as discussed in the theoretical literature, has a number of factors: steps, constraints, and decisions. It is influenced by approach (user or document centered), and it is influenced by the indexer, users, and tools used to represent documents (Mai, 2005). Factors at work in the indexing processes have been compiled here in the form of a rubric. This work is based on Tennis's work (Tennis, 2005).

Table 1 presents eleven factors at work in the indexing process. Some of them are artifacts and others are interpretive constructs. All figure in theories and practice of indexing. The second rubric relevant to our work is characteristics of structures of representation.

The indexing process uses structures of representation. Social tagging and subject cataloguing use different structures, and it is this point that allows us to see many differences.

In Table 2 we see twelve characteristics or elements of indexing structures. These characteristics and elements offer us a way making comparative statements about tags and subject headings in these two types of indexing. The final rubric we want to present frames the components of the discourse of indexing.

17th SIG/CR Classification Research Workshop, November 4, 2006

	Work Process of Indexing
1	Analysis / Interpretation Process(es)
2	Significant Characteristic(s)
3	Document(s)
4	Context(s)
5	Indexer(s)
6	Tool(s)
7	Representation(s)
8	Information System(s)
9	User(s)
10	Purpose(s)
11	Reflection(s) on the Process

Table 1. Work processes of indexing as seen through framework analysis

	Structures in Indexing
1	Type of Control (policy?)
2	Degree of Control (institution/personal)
3	Freedom from Control (work within or outside of vocab?)
4	Type of Combination (pre or post?)
5	Composition of Vocabulary (warrant?)
6	Consistency of Vocabulary
7	Specificity of Descriptors
8	Levels of Hierarchy
9	Lead-in Vocabulary
10	Syndetic Structure
11	Definitions/Scope Notes
12	Purpose(s)

Table 2. Characteristics and elements of structures used to represent indexing as seen through framework analysis

In Table 3 we see six components of indexing discourse. These components of discourse, when cast as a rubric here, allow us to categorize discourse surrounding the practice of social tagging and subject cataloguing. The follow section presents these rubrics alongside assertions about how social tagging and subject cataloguing compare given these categories.

	Discourse of Indexing for Indexers
1	Authority
2	Authorship
3	Technique
4	Links between texts (Intertextuality)
5	Scope
6	Language deployment

Table 3. Discourse in indexing as a framework

	Elements of Frameworks
1	Purpose
2	Predication
3	Function
4	Context

Table 4. Elements of frameworks

7. Findings

The findings are first presented here in tabulation. A brief summary of each table is presented, and some salient differences are noted. We then expand on them in the discussion below. It is important to note that these comparisons all have counter examples. The root of their utility is in identifying dominant discourse through texts and community acceptance. For example, there are social taggers that are very routinized in their behaviour, however, the findings from Golder and Huberman (2006) point to the consistent inconsistency in utilization of whole systems. In a similar vein, future work would match more empirical data to the discursive contours of the rubric presented below.

Table 5 presents a comparison of the work processes of both types of indexing. Clear differences in work process show up in categories 1, 3, 9, 10, and perhaps 11. Category number 1 presents social tagging as a multiple-purpose analysis process. This results in different significant characteristics (seen in category 2) and even different documents (category 3). Social tagging systems have grown up around communities that want to share goals. For example Zaadz (www.zaadz.net) and 43things (www.43things.com) do not tag *documents* in the form of text, images, movies, or the like. Instead these systems allow users to create tags for the tasks they want to complete, and then other users add that tag to their profile. For example, x on 43things

17th SIG/CR Classification Research Workshop, November 4, 2006

	Work Process of Indexing	Social Tagging	Subject Cataloguing
1	Analysis / Interpretation Process(es)	Task management, identification of topics or subject matter, considering future use by the indexer	Identification of subject matter, considering future use by a user (user-oriented, content-oriented, etc.)
2	Significant Characteristic(s)	Whole work or part of work – names (who owns the resource), topics, genre, place in a grouping (number), evaluation (funny, scary), relation to self (mystuff), related to task	Whole work – topics, forms of knowledge, geographic areas, genre, etc.
3	Document(s)	Web documents, ideas, not just works	Books, web, etc., works
4	Context(s)	On the Web	In a library
5	Indexer(s)	Personal relationship with material	Professional relationship with the material
6	Tool(s)	Tags, collections of tags	LCSH, catalogues, logs
7	Representation(s)	Tags (uncontrolled, postcoordinate)	Precoordinate, controlled, subject headings
8	Information System(s)	Social Tagging System (at its purposes – not unitary across systems)	Catalogue (and its purposes – supposed to be unitary across systems)
9	User(s)	(1) themselves (2) others [group?]	Catalogue users (never really themselves because professional mores)
10	Purpose(s)	Management, Sharing, Interaction	Finding, Collocation
11	Reflection(s) on the Process	Blogs and talks show evidence of this	Sauperl, UC report, and blogs see evidence of this

Table 5. Factors of work processes in indexing used to compare social tagging and subject cataloguing

wants to climb Grouse Mountain near Vancouver. He has this in common with y, z, and # of other users of the system.

Categories 9, 10, and 11 stand out as points of difference between social tagging and subject cataloguing because of the nature of the work: personal versus delegated. Social tagging is done for personal reasons. As such the purpose and reflection on that process are personal in nature. Likewise, since the act of tagging is for oneself, not someone else.

Table 6, adapted from Jacob's rubric (2000), outlines a schematic of the structures in subject cataloguing and social tagging. The differences between these structures are mostly related to purpose and local or professionally accepted policy. Policy shapes the interpretive control of subject cataloguing, and helps it fulfill its purpose. This purpose, collocation and precision, stands out as very different from folksonomies used in social tagging. It is not clear from any purpose statements of social tagging systems that they want to provide precision in collocation. They talk about management and sharing of documents or tags.

Table 7 outlines the discursive contours of these frameworks. In this table, we are concerned with the scope, authority, and technique of these examples of indexing. Taking a nod from Ron Day's analysis of the discourse of knowledge management, we here apply a modified rubric to social tagging and subject cataloguing. In this rubric, below, we can see that much of the discourse of indexing that situates social tagging stands in apogee to the discourse of subject cataloguing. They are two poles. Cataloguing admitting to a Fordist approach to indexing and social tagging to a post-Fordist approach. The latter is an approach that throws off routinization, institutionalization, totalizing discourse, and what Day calls rational productivity – the mode of production that maximizes profit and discourse that shapes thought on maximizing profit. In this case, subject cataloguing is an expensive activity, and work at improving cataloguing practice often means reducing costs.

17th SIG/CR Classification Research Workshop, November 4, 2006

	Structures in Indexing	Folksonomies (Social Tagging)	LCSH (Subject Cataloguing)
1	Type of Control (policy?)	No policy	Local policy and LCSH policy
2	Degree of Control (institution/personal)	No control or personal commitment to control vocab construction	Institutional control on vocabulary construction
3	Freedom from Control (work within or outside of vocab?)	No rules for using terms	Rules for using terms (eliminates interp. Synonymy)
4	Type of Coordination	Postcoordination	Precoordination
5	Composition of Vocabulary (warrant?)	Personal Information Warrant	LOC's warrant
6	Consistency of Vocabulary	Not consistent in coverage	Not consistent in coverage
7	Specificity of Descriptors	Not specific	Very specific
8	Levels of Hierarchy	Not present	Varies with precoordination
9	Lead-in Vocabulary	Not present	Partial
10	Syndetic Structure	None	Partial
11	Definitions/Scope Notes	None	Partial
12	Purpose(s)	Management and sharing	Collocation and Precision

Table 6. Characteristics and elements of indexing structures used to compare social tagging and subject cataloguing

17th SIG/CR Classification Research Workshop, November 4, 2006

	Discourse of Indexing for Indexers	Social Tagging	Subject Cataloguing
1	Authority	Personal	Institutional (two levels – local and national)
2	Authorship	Confessed	Occult
3	Technique	Generally unroutinized, matter of sense-making	Generally routinized, shaped by the institution
4	Links between texts (Intertextuality)	Collection of tags, other peoples' tags, and other web pages in individual's collection – explicit in interface	LCSH, other books in catalogue(s), other titles, user logs, user reference interactions – little is useful and it is transcendent – from an institution not from an interaction
5	Scope	Local Discourse	Totalizing Discourse
6	Language deployment	Personal informational tasks	Rational productive tasks

Table 7. Discursive components used to compare social tagging and subject cataloguing

	Components of Frameworks	Social Tagging	Subject Cataloguing
1	Purpose	Share, innovate organization	Fulfill cutter's objective #2
2	Predication	Tags, Profiles, Folksonomy Collections	Subject Headings Lists in an OPAC
3	Function	Share (social or accidental)	Find and Collocate (formal and intentional)
4	Context	The web	Library and its collection and users

Table 8. Comparison of the elements of social tagging and subject cataloguing

8. Discussion of Findings

Employing framework analysis we see that social tagging and subject cataloguing are quite different. This confirms our commonsense impressions of these two frameworks. However, the differences that obtain between these two, as set out in the rubric above, illuminate a set of discursive differences that once if at one time were assumed is laid bare by the boxes above.

Those differences pivot on the concept of *total representation* outlined by Day (2001). More than the Tennis and Jacob components, Day's discursive analysis serves us well in drawing a clearer line between social tagging and subject cataloguing. It also appears in the parts of the rubric attributed to the former two.

The discourse of total representation is a system that models work, work practices, and the language we use to discuss those work practices. Day outlines in his 2001 work, how knowledge management has made a shift to a post-Fordist view of work. In this shift conceptions of work practice and the language used to discuss it are cast in a different mode of economy – with different goals. The same can be observed in the discursive context of social tagging. It is a shift, in indexing, from a Fordist cataloguing environment, where every document is a Model-T, to a decentralized and creative craft of indexing that is not modeled on the assembly line. Social tagging is post-Fordist, to use Day's construction. And the discourse that presents it – is not a total representation discourse – it is a indexing of individual craft interaction. It is not dependent on anyone else's authorization or authority.

Furthermore, there is a reinvention of authorship and agency in the post-Fordist social tagging discourse. We no longer see a monolithic standard, we see individuals tagging personal collections, using ad hoc tools.

Finally, social tagging, in the context of the web, using links to personal collections and profiles, and with its focus on sharing highlights a novel kind of intertextuality in indexing. Intertextuality is present in much of indexing (Beghtol, 1986). However, the intertextuality that links personal collections and profiles in the web is different in kind.

Employing framework analysis to social tagging and subject cataloguing, we find a diversity of predications, functions, and contexts, and only a superficial similarity in purpose. If we probe

deeper into this perceived similarity of purpose shared among these frameworks, we see a complex diversity of collocation, personal information management, and attestation of conceptualizations (with no regard to retrieval based on search).

With these differences exposed, and from a theoretical vantage point, we can see indexing as an incipient and under-nourished framework that, as yet, only approximates a fulfillment of its intention. Indexing is incipient because it does not yet achieve what it could in the contemporary technological environment. Social tagging, as an organic activity, offers us insight into (1) the seemingly insufficient representation of authorship in indexing, (2) the lack of links to literary, user, and request warrant, and (3) points to a need for a more explicit intertextuality in indexing. We also see a wide diversity of purposes in indexing, and therefore a wide diversity of task fulfillment in indexing. We can look at the work practices of subject cataloguing and the analysis of types of social tags for evidence of this.

Indexing is incipient in another way. Although rich in conceptual development, indexing is incipient on the implementation level, because it does not exploit the current technological environment with adequate innovations from indexing theory. Social tagging has called this into question. Key to ameliorating this technological and conceptual divide is identifying the explicit links to intertextuality, authorship, and task.

Indexing is not only incipient, but also under-nourished. Indexing is languishing because as innovation moves on, many of indexing's prescriptions remain wedded to a modernist idea of mass production metaphors, monolithic or univocal concept markers (universal class marks and subject headings), and Fordist techniques and outcomes – a belief that we can index once and share – no matter what the context (Day, 2001). It was not until August of 2003 that IFLA removed the rhetorical of universality from its international standards work in bibliographic control and cataloguing (IFLA, 2006). The priority shifts that have accompanied the work in the networked environment have begun this change. Likewise, the phenomenon of social tagging shows us that the modernist concept of indexing is no longer desirable because we see a very personal and constantly evolving set of systems support a framework that works with profiles, personal collections, and novel tagging combinations.

Future research will help expand on these concepts and support them with further empirical research. And even here, we see many questions: is there a return to Fordist universalizing

discourse if we link a collection to an institution – as opposed to a personal collection? Is this inevitable? Indexing is what is, what’s the point in dismissing the totalizing discourse in indexing? Since its purpose is unified, it is presented, as a given. It is not an aporia requiring citation or situating. Finally, we are knocking at the door of professionalism and its knowledge base. In a Fordist environment we have some idea of a division of labour, and therefore a reification of valued professional knowledge. In a world of non-professionals indexing we are doing something else, we are critics or artists, using indexing techniques to make personal collections stand out. We are pop artists, not professionals.

9. Conclusion

This workshop establishes a stage on which we can ask what is the future of indexing. Does it have a rich future? Using a comparative framework analysis, we can begin to make claims about the next stages for indexing. The incipient and under-nourished state of indexing – made manifest by the rise in social tagging, and its similarities and contrasts to subject cataloguing – points not to the demise of indexing, but rather to the need for new design requirements; for more discursive and intertextual technologies – dependent on authored, personal, situational, and referential acts, or in the case of indexing as delegation (by professionals) – as authored, institutional, situational, and referential acts. Social tagging, as a phenomenon, has allowed us to reflect on what indexing can do better in this contemporary environment. It is neither Pandora’s box nor panacea. Social tagging highlights the interstices of authorship, intertextuality, and context in indexing, and asks us to fill in the gaps. It is a catalyst for improvement and innovation in indexing.

Acknowledgements

Thanks goes to Benjamin Good for reading a draft of this paper.

References

Beghtol, C. (2003). Classification for information retrieval and classification for knowledge discovery: relationships between “professional” and “naïve” classifications. *Knowledge Organization* 30: 64-73.

Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality, and the cognitive act of classifying documents. *Journal of Documentation* 42: 84-113.

Day, R. (2001). Totality and representation: A history of knowledge management through European documentation, critical modernity, and post-Fordism. *Journal of the American Society for Information Science* 52: 725-735.

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

IFLA. (2006). IFLA Core Activity: IFLA-CDNL Alliance for Bibliographic Standards (ICABS): Background. Retrieved October 30, 2006, from <http://www.ifla.org/VI/7/icabs-background.htm>.

Jacob, E. K. (2000). Unpublished notes on controlled vocabulary evaluation. Indiana University, Bloomington, IN.

Mai, J.-E. (2005). Analysis in indexing: Document and domain centered approaches. *Information Processing & Management* 41: 599-611.

Saupperl, A. (2004). Catalogers' common ground and shared knowledge. *Journal of the American Society for Information Science and Technology* 55: 55-63.

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science* 50: 1119-1120.

Tennis, J. T. (2005). Conceptions of subject analysis: A metatheoretical investigation. Ph.D. dissertation. University of Washington, Seattle, WA.

Vickery, B. C. (1997). Ontologies. *Journal of Information Science* 23: 277-286.