The influence of system designer intention over collaborative tagging. an analysis of tagging behaviour in CiteULike and Delicious

The Influence of System Designer Intention Over Collaborative Tagging

An Analysis of Tagging Behaviour in CiteULike and Delicious

Mariam Elhussein[1] and Keiichi Nakata[1]
[1]*Informatics Research Centre, University Reading, Reading,UK*
m.a.m.elhussein@pgr.reading.ac.uk, k.nakata@henley.ac.uk

Keywords: Social Tagging, CiteULike, Delicious, Semiotics, Activity Theory.

Abstract: Tagging provides support for retrieval and categorization of online content depending on users' tag choice. A number of models of tagging behaviour have been proposed to identify factors that are considered to affect taggers, such as users' tagging history. In this paper, we use Semiotics Analysis and Activity theory, to study the effect the system designer has over tagging behaviour. The framework we use shows the components that comprise the tagging system and how they interact together to direct tagging behaviour. We analysed two collaborative tagging systems: CiteULike and Delicious by studying their components by applying our framework. Using datasets from both systems, we found that 35% of CiteULike users did not provide tags compared to only 0.1% of Delicious users. This was directly linked to the type of *tools* used by the system designer to support tagging.

1 INTRODUCTION

Web 2.0 tools moved users from consumers to contributors of content in the web. Users provide different types of materials, such as pictures, text, videos and bookmarks. They also add tags, comments and reviews for their content and others' content as well. Tags were defined as lightweight keywords that are attached to content in order to provide support for retrieval and categorization (Trant, 2009). Tagging received much attention lately in research and was identified to provide services to Web contents. It facilitates self-retrieval and allows users to categorize their material, as well as opinion expression and content promotion (Rader & Wash, 2008). Tagging is also used to organize personal activities using keywords such as "to read" or "read later". Moreover, collaborative tagging systems such as CiteULike and Delicious leverage the collaborative efforts of users to share Web contents.

There are a number of research directions in tagging and collaborative tagging systems. One research direction focused on studying tagging models which were used to describe tagging behaviour. These models used data to explain tag generation. Golder & Huberman (2006) described the first tagging model using the urn model. They found that users tend to imitate each other while tagging by reusing already existing tags. Other models were based on this model and introduced other factors to imitation including the semantics of tags and the user background knowledge. All of these models used data retrieved from tagging systems such as Delicious in order to describe users' behaviour. Comparison between different tagging systems took place as well based on their tagging data.

This paper uses a framework that we generated using Semiotics Ladder and Activity Theory components to analyse tagging systems (Elhussein & Nakata, 2012). We use these theories to decompose the tagging system into smaller components in order to study their effect on certain tagging behaviours. This framework can be extended to study any phenomena that is linked to tagging systems, e.g. social norms generated within sub-communities. We choose here to study two tagging behaviours namely: number of tags per user and the usage of self-organizing tags.

The paper will continue as follows. First we provide a brief background of our framework. Second, we will use it to analyse CiteULike and Delicious to identify the factors that direct tagging

behaviour. Afterwards we use actual data from both systems to showcase the two behaviours we are focusing on with guide of the output from the theoretical analysis. Finally we use the factors identified from the theoretical analysis to explain the outcome of the data.

2 BACKGROUND

In a previous paper we described a systematic approach to analysing tagging systems using Activity Theory (AT) and Semiotics Analysis (SA), see (Elhussein & Nakata, 2012). The process uses a framework to identify the components that comprises a tagging system. According to Huang and Chuang (2009) a tag can be described as a sign using the triadic model of *representamen, object* and *interpretant*. Where the representamen is the tag itself, the object is the content tagged and the interpretant is the tagger who assigns the meaning to the tagged content using the tag. This facilitated the analysis of the tagging system using the Semiotic Ladder (SL), a six-level view of systems starting with their physical properties to their social effects (Charles Hartshorne & Weiss, 1935). Figure 1 shows the tagging system using the SL analysis. Each layer reveals a different property for the tagging system. In this analysis we focus on the human-information functions of the tagging system, as the IT platform levels is not directly relevant to this analysis. The *semantic layer* is about the meaning of adding tagging capabilities to a certain system. The *pragmatic layer* is about the intentions behind the communicated message. This refers to the intention the system designer had behind allowing users to tag. The *social world* reflects the social effects of tagging system.

The second level of analysis is based on Activity Theory (AT). Using the six components identified by Engestrom (1987), the tagging system is decomposed into the components that comprises the tagging activity. This includes the *tools, subject, object, rules, roles and community* of the tagging activity.

According to AT, the *object* of the tagging activity is the reason for including the tagging capability in the website. It answers the question of why users are allowed to provide tags. The *subject* of the system represents the internal understanding or motivations of the tagger who is performing the act of tagging. The subject uses a set of *tools* to perform the activity. Tools are sometimes referred to as "means" or "artefacts" (Blanton, Simmons, & Warner, 2001). In a tagging system, tools refer to the user interface components used to facilitate tagging, such as textboxes and buttons that allow users to provide tags for content. The community of the tagging activity includes taggers, tag consumers and moderators. They are the community members who are also governed by a set of *rules*. They can be the set of "terms of service" that the user agrees to when registering to use the sites. Users are also governed by the social norms that develop within the tagging

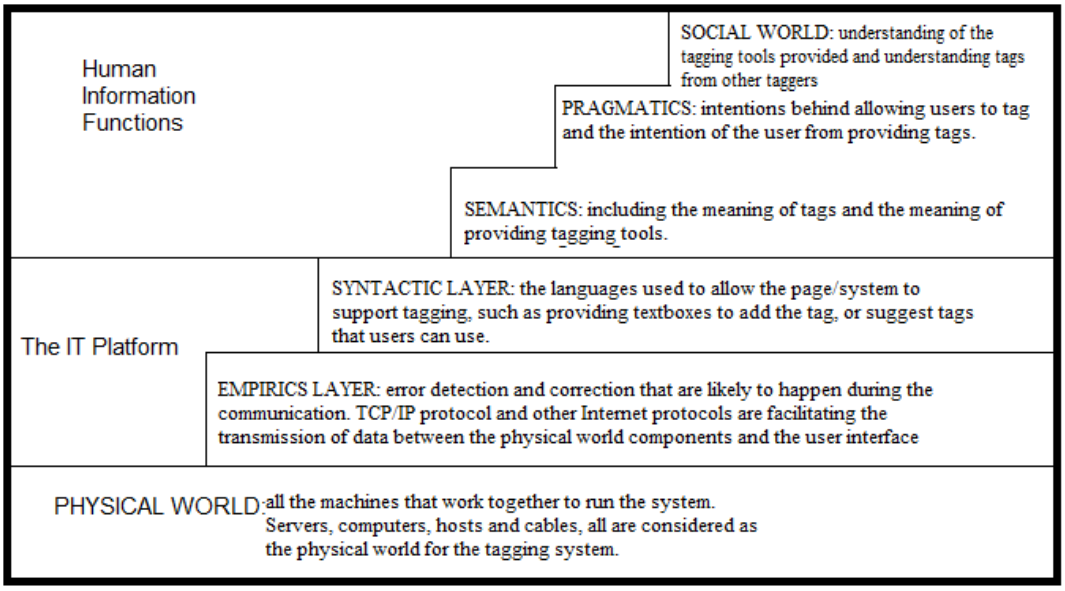

Figure 1: A semiotic analysis of tagging system using the semiotic ladder

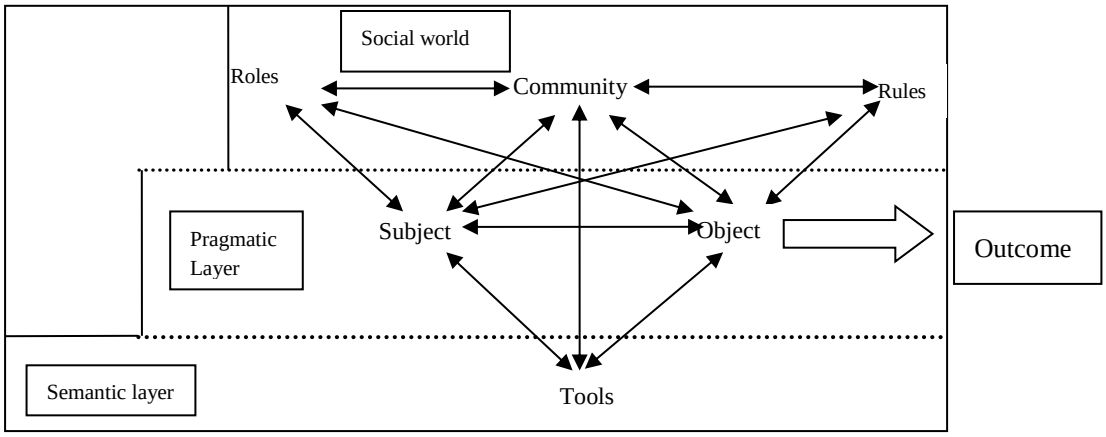

Figure 2: Tagging system as a system of signs and activity.

community. The *Division of Labour* (roles) is concerned with the role that each member of the community is supposed to play in the tagging activity.

The framework combines both theories in order to identify the factors that influence the tagging activity. The framework provides a two dimensional view for a tagging system as a system of signs and an activity. Figure 2 shows the framework developed as the result of combining SL and AT.

In our previous work, we used this framework to identify the factors that influences tagging behaviour (Elhussein & Nakata, 2012). These include *semantics components (tools), pragmatic components (subject and object)* and *social components (community, roles and rules).* In this paper we will explore the influence the *object* has over tagging behaviour. The object refers to the intention of the system designer when he included tagging capabilities. This affects all other components of the system. In the next section, an analysis for two tagging systems, namely CiteULike and Delicious, will be provided using the previous framework.

3 ANALYSIS OF TAGGING SYSTEMS

The system designer's intention behind adding tagging capabilities to a certain system can be revealed by the tools that are used, the rules that are set and the roles that he creates for users to play in the activity. The main focus of this paper is on the tools the system designer uses to express his/her intentions.

3.1 CiteULike

CiteULike (www.citeulike.org) is an online system that allows researchers to add research papers in an online repository that can be accessed from anywhere. The tagging tools provided by CiteULike are a textbox that allows users to write the tags. No tag recommendation is provided; the only help the user gets is by showing his list of previous tags (Figure 3).

In case the paper was copied from another library, the existing tags are posted as recommendation where the user can confirm them directly or edit them. The system does not provide any kind of keyword extraction to be used as tags. Multiple tag application to more than one paper is also allowed in CiteULike. When selecting papers, a textbox and an "Add" button appears and allows attaching the same tag to the selected papers.

CiteULike provides special types of tags, including private tags and the "for:" tags which are ones directed to other users. Private tags can only be seen by the user. Any tag starting with the character "-"is considered private. The "for:" tags allow users to tag other users in a paper. This type of tags is used to draw other user's attention to a certain paper. These tags can be added using the same textbox that the user uses to add regular tags. Another type of tags provided is called "priority". It is a drop-down list with specific values including "Top priority", "I really want to read it", "I will read it", "I might read it"," I don't really want to read it" and "I've already read it" (Figure 4).

Figure 3: Adding tags in CiteULike

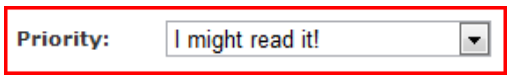

Figure 4: Priority options when adding a new paper in CiteULike.

From the previous analysis of the tools used in CiteULike, we can infer the following reasons for adding tagging capabilities to CiteULike with reference to the system designer's intention:

- Support link retrieval: this can be understood from adding tagging capabilities to the system and is also stated in the help section of the site.
- Categorization and grouping of similar papers: as stated in the help section.
- Signalling other users: using the "for:" tags.
- Self-organizing functions using the "priority" tags.

We will now move to analysing the second tagging system here, which is Delicious.

3.2 Delicious

Delicious (delicious.com) is an online bookmarking site that allows users to add their links and tag them. It acts like an online list of favourites that can be found from every computer that has internet access. Users can add their tags and descriptions to their links if they choose to. In Delicious, the tagging tools provided (Figure 5) consist of:

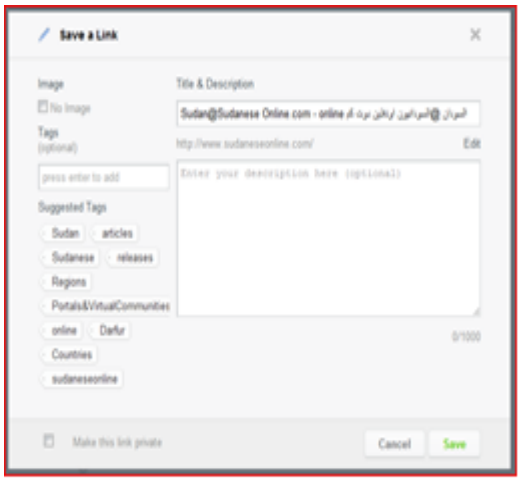

Figure 5: Saving a new link in Delicious.

- A toolbox where tags can be typed into
- A list of suggested tags that can be attached to the link. The list was clearly extracted from the website so an underlying modal must have been used to provide these keywords.

From the simple tagging tools provided by Delicious, we can infer the system designer's intention as to use tagging to support retrieval and categorization. It is also stated in the help section of Delicious.

In the next section we will analyse data from CiteULike and Delicious to show the effect the *object* (system designer intention) have over users' tagging behaviour. The actual data is used to explain how users were affected by the decisions made by the system designer.

4 DATA ANALYSIS

Data from CiteULike was downloaded from their website on August 13th, 2012. The data captures users' activities starting from May 30th, 2007 onwards. It consist of 17, 622,158 paper along with the tags that were assigned to each of them. Delicious dataset used is DeliciousT140 Dataset (Zubiaga, 2009), an XML file consisting of 144,574 URLs tagged with 67,104 different tags and tagging occurrences of 2,015,059 tags. Table 1 provides a description of the data sample we used from both datasets.

From table 1 we can see that the average number of tags per user in CiteULike is less than Delicious. It also shows that the number of untagged items in CiteULike and the number of users who did not tag was significantly less than Delicious where 35% of CiteULike users did not provide tags. This can be due to the fact that CiteULike provided tools that took some of the functions that can be provided by tags, such as self-organizing function (the "priority" tag), leaving free tagging for categorization and retrieval purposes. This was clearer when we analysed the data sample to find the number of times users of both systems used a self-organizing tag. Table 2 shows the result of this analysis.

Table 1: Breakdown for tagging data in CiteULike and Delicious.

	CiteULike	Delicious
#users	2,500	2,500
#tags	195,655	515,325
#items	200,469	261,428
Avg #tags/user	120	206
Avg #tags/item (per user)	1	2
#untagged items	4814 (2.4%)	895 (0.3%)
#user who did not tag	874 (35.0%)	3 (0.1%)

Table 2: The number of times some selected self-organizing tags were used in CiteULike and Delicious

Tags	CiteULike	Delicious
Important	4	34
Read	114	3836
Later	0	477
To do	14	499
Temp	1	822
Test	622	1013
Total	755	2883

The list of tags in table 2 is an example of the popular self-organizing tags usually used. We can see from the numbers that CiteULike users used fewer self-organizing tags than those in Delicious. This can be explained by the "priority" tag provided by CiteULike which reduced the use of the tag such as "later" to zero.

5 DISCUSSION

The analysis framework developed in this paper combines the semiotic framework and activity theory. This was based on the observation that tags can be treated as signs that stand for Web contents, and tagging is a collaborative activity that is embedded in a community of users. In this attempt, we have established that tools are assigned a meaning (semantic), which is interpreted in the context of an activity. In the two cases we examined, this was evident from the tool features provided that Delicious encouraged tagging behaviour through a range of tagging support compared to CiteULike, which resulted in significantly larger proportion of users actively providing tags. This indicates that features of the tool influenced the interpretation of purpose (object) of the activity – which in turn affected the social level factors of rules, communities and roles.

Mapping the semiotic ladder and activity theory components is by no means uncontroversial and unique. Tools can be a syntactic artefact to which meanings (semantics) are assigned according to the context of use (pragmatics). Nevertheless, this attempt has shown that by combining the two

methods of analysis, each benefitting from the other – the semiotic analysis from the identification of clearer activity components, and activity theory analysis from the semiotic dependencies between the layers.

In this paper, we only compared two popular tagging sites, CiteULike and Delicious. While these were effective in contrasting the level of user tagging contributions, a further analysis of other tagging sites would enable us to generalise our findings. Moreover, there could be other factors, such as difference in the communities they serve, i.e., CiteULike for academic communities and Delicious for public in general.

6 CONCLUSION AND FUTURE WORK

In this paper we applied a theoretical framework that was extracted from two theories: Semiotics and Activity Theory. The paper shows how the framework can be used to analyse tagging systems in order to identify the activity components that may have effect over the tagging behaviour. The main focus of the paper was to show the effect the system designer have over the tagging behaviour of the users.

We showed how the system designer intention was manifested through the tools chosen to support tagging. In CiteULike, the system designer intended to separate the self-organizing function of tags from other types of tags. This was done by specifying a list of tags that allow users to organize their papers according to their reading priority. This reduced the number of self-organizing tags in CiteULike. This can also be a reason for those who did not provide tags at all. In Delicious, the use of a recommender system supported tag generation and resulted in a very low percentage of untagged items compared to CiteULike.

The framework identified tools as a way to reflect the system designer intention. Our future work will be focused on other factors that affect tagging behaviour. The factors are identified using the components of our earlier framework. Our goal is to gain an understanding of how do these factors affect tagging behaviour and ultimately use them to design a tagging system that is tailored to meet predefined set of tagging behaviours.

REFERENCES

Blanton, W., Simmons, E., & Warner, M. (2001). The Fifth Dimension: Application of Cultural-Historical Activity Theory, Inquiry-Based Learning, Computers, and Telecommunications to Change Prospective Teachers Preconceptions. *Journal of Educational Computing Research*, *24*, 435–463.

Charles Hartshorne, & Weiss, P. (1935). *Collected Papers of Charles Sanders Peirce*. *Nature* (Vol. 135). Cambridge, MA: Harvard University Press. doi:10.1038/135131a0

Elhussein, M., & Nakata, K. (2012). Analysing the Factors that Influence Tag Choice based on Semiotic Analysis and Activity Theory. *Internation Conference on Social Informatics*. Washington, DC.

Engestrom, Y. (1987). *Learning by Expanding: An Activity-theoratical Approach to Developmental Research.* Helsinki: Orienta-Konsultit.

Golder, S., & Huberman, B. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, *32*(2), 198–208. doi:10.1177/0165551506062337

Huang, A. W., & Chuang, T. (2009). Social Tagging , Online Communication , and Peircean Semiotics : A Conceptual Framework. *Journal of Information Science*, *35*(3), 340–357. doi:10.1177/0165551506nnnnnn

Rader, E., & Wash, R. (2008). Influences on tag choices in del.icio.us. *Proceedings of the ACM 2008 conference on Computer supported cooperative work CSCW 08*, 239. doi:10.1145/1460563.1460601

Trant, J. (2009). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, *10*(1).

Zubiaga, A. (2009). Content-based clustering for tag cloud visualization. *Network Analysis and Mining*, 316–319.