Dynamics of human categorization in a collaborative tagging system: How social processes of semantic stabilization shape individual sensemaking

CrossMark

Tobias Ley [a,*], Paul Seitlinger [b]

[a] Tallinn University, Tallinn, Estonia
[b] Graz University of Technology, Graz, Austria

ARTICLE INFO

ABSTRACT

We study how categories form and develop over time in a sensemaking task by groups of students employing a collaborative tagging system. In line with distributed cognition theories, we look at both the tags students use and their strength of representation in memory. We hypothesize that categories get more differentiated over time as students learn, and that semantic stabilization on the group level (i.e. the convergence in the use of tags) mediates this relationship. Results of a field experiment that tested the impact of topic study duration on the specificity of tags confirms these hypotheses, although it was not study duration that produced this effect, but rather the effectiveness of the collaborative taxonomy the groups built. In the groups with higher levels of semantic stabilization, we found use of more specific tags and better representation in memory. We discuss these findings with regard to the important role of the information value of tags that would drive both the convergence on the group level as well as a shift to more specific levels of categorization. We also discuss the implication for cognitive science research by highlighting the importance of collaboratively built artefacts in the process of how knowledge is acquired, and implications for educational applications of collaborative tagging environments.

1. Introduction

Because of the ubiquity of social web technologies, there has been a recent growing interest in how people make sense of large quantities of information when they browse the web (Pirolli & Russell, 2011). In this paper, we focus on the sensemaking process which results from people using a collaborative tagging system (Golder & Huberman, 2006; Marlow, Naaman, Boyd, & Davis, 2006). In systems like *Delicious* (websites), *Flickr* (photos) or *Soundcloud* (music), people describe different types of resources they discover on the web by assigning freely chosen keywords (called *tags*) to store them for later use. Fu (2008) describes this process as an iterative exploratory search-and comprehend cycle which leads to a close interaction between internal and external representations of concepts, tags, and resources as a user searches the web. The tags that a user applies are a result of his or her mental categorization processes. Over time and as more resources are tagged, a user's understanding of a particular topic increases, and his internal categories change and become more refined.

Because the resources and tags collected by one user can be seen by others, it is usually assumed that individuals are influenced by tags as social cues. In the social web, there is considerable influence of collective information on individual behavior (Li & Sakamoto, 2014). In collaborative tagging, tags function as primes in activating prior knowledge (Cress, Held, & Kimmerle, 2013; Fu, 2008), users imitate each other's tag assignments to a certain degree (Fu, Kannampallil, Kang, & He, 2010; Seitlinger & Ley, 2012; Seitlinger, Ley, & Albert, 2015; Sen et al., 2006), and learning processes can take place when a user browses the tag collection of another user, thereby discovering resources and tags that influence that user in his/her future tag assignments (Held, Kimmerle, & Cress, 2012; Nelson et al., 2009). As a consequence, collaborative tagging allows for studying how individual sensemaking is shaped by social processes, namely by artefact-mediated collaboration.

Gaining an understanding of the development of categories and the corresponding tag assignments is important for improving information access in the social web. Research on improving information access with tags has been seeking different routes. For example some researchers have suggested enhancing semantic

* Corresponding author at: Center for Educational Technology, Tallinn University, Narva Mnt 29, 10120 Tallinn, Estonia. Tel.: +372 6409 355; fax: +372 6409 422.
  E-mail address: tley@tlu.ee (T. Ley).

search technologies with social tags (e.g. Gayo, de Pablos & Lovelle, 2010; Specia & Motta, 2007). Another direction has been to employ recommender services that are based on tags, or that suggest suitable tags (Bollen & Halpin, 2009; Jäschke, Marinho, Hotho, Schmidt-Thieme, & Stumme, 2007). Finally, several tools have been suggested that attempt to enhance the collaboration processes around tags and their corresponding categories (e.g. Braun, Schmidt, Walter, Nagypal, & Zacharias, 2007; Tolosa, Gayo, Prieto, Núñez, De Pablos, 2010; Yew, Gibson, & Teasley, 2006).

However, to fully exploit this potential for semantic technologies or recommenders, it is important to understand the process by which shared terminological patterns can emerge without an explicit coordination (Gayo et al., 2010). This is because retrieval of information may actually become more difficult both for experts and for novices when resources are described on varying levels of specificity. While for the former, the information value of a basic level category is too low, for the latter the specific categories are not sufficiently well represented in memory, and, hence, their labels difficult to comprehend (Rogers & Patterson, 2007).

The particular contribution of our paper is twofold. First, and rather than studying the information access, we examine the underlying mechanisms of how people develop and extend their categorization in collaborative tagging system. Our assumption is that phenomena of categorization can only be studied by looking at how internal categories and tags people use are coupled. Therefore, we looked both at internal categories students learned, as well as the tags they used.

Secondly, we study how the development of categories is mediated by an important social process in collaborative tagging, namely the development of shared language that results from semantic stabilization (e.g., Baronchelli, Felici, Loreto, Caglioti; Crokidakis & Brigatti, 2015; Steels, 2006). Here, we hypothesize that individual sensemaking will be more successful if it is built on a shared use of tags on the group level. While in previous studies, the impact of tags as social cues on individual learning and browsing behavior was examined under lab conditions (e.g. Held et al., 2012), none of these previous studies has actually looked at how semantic stabilization influences individual learning. This is because semantic stabilization is a social process that develops over time and therefore needs to be studied in a setting that allows users to interact with each other over a more extended period of time. Others have examined the development of categories when using a tagging system (Fu & Dong, 2012), but have not focused on semantic stabilization as a mediating factor.

We have therefore conducted a field experiment in which students used a collaborative tagging system in the context of a university course over a more extensive period of time. By giving some groups more time for their task, we realized a condition in which semantic stabilization should be more likely as compared to those groups that spent less time working on the task. As a measure of individual learning, we then looked at the categories and their increasing specificity that students developed as a result, in terms of both, the tags students used and their internal memory representation.

2. Dynamics of human categorization in collaborative tagging

When students interact with the Collaborative Tagging system by searching for web resources, assigning tags and sharing these with other students, we assume a dynamic coupling between the students and their shared artifacts forming a cognitive ecosystem (Hutchins, 2010). The artifacts (e.g. tags) influence students' categorization and – depending on the artifacts' emerging shape and structure – support or exacerbate the collaborative process of

exploring and deepening the understanding of a given topic. Similar to individual learning progress, where internal categories do not emerge out of nowhere but are differentiated representations of existing ones (e.g., Rogers & Patterson, 2007), artifacts emerge over time from a collective artifact-mediated activity (Hutchins & Hazlehurst, 1995). Here, we use the term artifact to refer to some type of inscription (Latour & Woolgar, 1986) that is a malleable means of representation of things (e.g., Web resources) that can be changed and improved continuously by members of a community (e.g., students of a university course) (see also Schreiber, 2006). The continuous development of artifacts, such as tags assigned to a Web resource, reflects the ongoing development of underlying processes of understanding of things within the community and thereby provides an empirical unit of analysis. The relationship between artifacts and things to which they refer, becomes a matter of a dynamic, social practice (e.g., Roth & McGinn, 1998; see also Schreiber, 2006), leading to a shared understanding of this relationship.

In case of tagging, one simple mechanism behind the emergence of a shared artifact-thing-relationship is semantic priming: existing tags prompt a particular user to activate related memory content resulting in converging categorization and verbalization processes among the users (e.g., Fu & Dong, 2012; Fu et al., 2010; Seitlinger et al., 2015). Over time, these mutual influences create positive feedback loops (Hutchins & Johnson, 2009) that result in semiotic dynamics (e.g., Steels, 2006) giving rise to semantic stabilization (Wagner, Singer, Strohmaier, & Huberman, 2014), i.e., a consensual use of tags for a resource. In terms of distributed cognition, the mechanism of a positive feedback loop underlying the spreading adoption of tag-resource pairs can be described as follows: "Once a behavior enters the repertoire of one agent, …, it is likely to enter the repertoires of others, which makes it even more likely to enter the repertoires of still others, and so on." (Hutchins & Johnson, 2009, p. 526).

The assumed semiotic feedback loops leading to semantic stabilization imply that external artifacts and peoples' interpretations co-shape each other, as also proposed by approaches towards cognition distributed across extended systems of human and non-human actants (e.g., Fu, 2008; Hutchins, 2010; Hutchins & Hazlehurst, 1995; Latour, 1996): Artifacts introduced by preceding individuals augment the experience of subsequent individuals; the augmented experience may influence its interpretation and hence, the creation of new artifacts.

To summarize, our main assumption is that different levels of semantic stabilization that form in a group observable in the use of tags will mediate individual understanding in a sensemaking task. In the following two sections, we will first discuss how semantic stabilization has been studied in collaborative tagging and define a way to measure it. Secondly, we will introduce the basic level shift in categorization as a way to define a deepening of understanding in individual sensemaking. We will then summarize our hypotheses in more detail and describe a field experiment that was designed to test the influence of semantic stabilization on individual sensemaking.

2.1. Semantic stabilization in collaborative tagging: the emergence of a shared language

Although all users of a collaborative tagging system are free to use whichever keyword they want to describe the resources they collect, research in the tradition of semiotic dynamics (e.g., Steels, 2006) has shown that the development of a consistent tag vocabulary can usually be observed (Golder & Huberman, 2006). In the present study, we measure an emerging coherence in tag assignments by observing the time evolution of the number of unique tags N_u (see also Baronchelli, Felici, Loreto, Caglioti, &