The use of social tagging to support the cataloguing of learning objects

Anderson Roque do Amaral
and Luciana A M Zaina
Computer Science Department
Federal University of São Carlos
Sorocaba, Brazil
Email:{anderson.amaral, lzaina}@ufscar.br

José F Rodrigues Jr
ICMC - Computer Science
University of São Paulo
São Carlos, Brazil
Email: junio@icmc.usp.br

Renan de Aguiar
and Marcos Cavalcante Barboza
Computer Science Department
Federal University of São Carlos
Sorocaba, Brazil
Email:renan.lord.de.aguiar@gmail.com,
mcavalcanteb@hotmail.com

Abstract—**Social networks have become the main media for information dissemination in the so-called Web 2.0. The core of these networks is social tagging, the act of annotating what users see in their social space. In the education domain, social tagging is potentially a useful resource to improve the organization (cataloguing) of large repositories of learning objects. To the present moment, however, many questions are open about social tagging in e-learning. In this work, hence, we proceed to answer three questions: (1) Can social tagging successfully catalog e-learning objects? (2) How do students behave according to Körner's classification: categorizers or describers? and (3) Does social tagging converge to a well-defined descriptive vocabulary of tags? We performed a large experiment with 336 technician students that marked 218 electronic learning objects for about 4,985 times. Our results show that social tagging is a promising practice for e-learning; however some issues must be addressed to prevent an excessive number of categorizer students and, also, a premature convergence of the vocabulary of tags. Our conclusions are specific for the setting of our experiment, but we generalize them as much as possible suggesting guidelines of how to use social tagging in e-learning.**

Keywords—social tagging, folksonomy, learning objects.

I. Introduction

In the last years, a notorious increase in the number of applications related to the Social Web has been observed; applications that promote the interrelationship of people and of knowledge through the Web. Accordingly, academic and commercial e-learning systems have adapted to the characteristics of these applications, becoming more appealing to students [1].

In the social Web, a remarkable feature is the possibility of tagging online content, what allows the users to create vocabularies that categorize the resources – or learning objects [2] – they interact with.

Formally, learning objects refer to entities used in the teaching-learning process; videos, images, simulator software, and text, among other possibilities. In the electronic-learning domain, it is desirable that learning objects be reusable for different learning objectives, or be combined to build up more complex objects [3]. To this end, the objects must carry metadata that contextualize and describe their use in a standard manner [4].

The possibility of tagging standard learning objects is converging to what is known as folksonomy (folk, as for

people; sonomy, as for taxonomy), which manifests in the form of social tagging systems [5], a self-evolving classification system that emerges from user interaction without *a priori* rules for what is created [6]. In the practice of social tagging, users collaboratively use tags to annotate and make sense of content, a valuable source of information that has the potential of bringing order (indexing and classification, or cataloguing) to vast volumes of information.

Social tagging systems assume that users will express their impressions by means of tags that they use to classify the content they use [7]. Social tagging fulfills the impracticable classification that would be performed by specialists [8], its main features include: flexibility, as the users use their everyday dynamic vocabulary; pattern identification, as the users spontaneously choose the words that best describe the content; and collaboration, as predicted by Social Web applications.

In the context of e-learning, social tagging supports the cataloging of learning objects based on the tags provided either by students or by professors. In the case of students, the tagging process concerns a reflection experience in which students tag the objects based on their own experience [9]. Then, the repositories of tagged learning objects can be searched by the very students or by other people in the course of learning [10].

However, for a good cataloging to occur, it is necessary to have an adequate tagging process; this is observed in repositories of commercial objects professionally tagged, but seldom verified in academic self-organizing repositories. In such systems, the users may behave following distinct patterns – categorizers or describers, according to Körner's classification [11], impacting the process in different ways; in either case, the vocabulary of tags must be suitably heterogeneous in order to extensively describe the objects. These are relevant issues that we address in the present work; more specifically, we aim at answering the following research questions:

1) Can social tagging successfully catalog e-learning objects?
2) How do students behave according to Körner's classification: categorizers or describers?
3) Does social tagging converge to a well-defined descriptive vocabulary of tags?

In order to gather evidence and hypothesize about these questions, we performed a large experiment with 336 techni-

cian students that marked 218 learning objects 4,985 times. We, then, analyzed the results raising conclusions about how the process of social tagging must be carried in an e-learning setting. In our experiment, we did not use a particular pedagogical learning model. But we believe that our proposal can be extended to different pedagogical theories. The focus is the cataloging of learning objects which facilitates the access to such objects.

The rest of the paper is organized as follows: Section II introduces basic concepts related to educational social tagging systems, also presenting specific works on the topic; Section III presents the methodological and experimental processes followed by discussions in Section IV. Section V concludes the paper.

II. RELATED WORKS

This section is divided in two parts, in the first one we review basic concepts necessary as basis for this research; in the second one we present works related to this research, that is, on social tagging over the educational domain.

A. Basic and related concepts

Social tagging systems grew in popularity in the last years due to their simplicity to categorize and retrieve content based on tags. The increase in the number of users that provide information to such mechanisms caused the emergence of systems that assume the users express their preferences by means of the tags they create and use [7].

The main features of social tagging are: communication and immediate feedback, fast adaptation to vocabulary alterations, single or collective organization of objects, potential of cataloging, and assistance in the recommendation of content, among others. The tags entered in the system allow users to freely explore objects and other users' profiles without having to follow a rigid predefined hierarchy of concepts [12]. Users, objects, and tags define a tri-dimensional space in which one can analyze and discover similarities, tendencies, and users' interests [7]. Along this work, we refer to these desirable features when analyzing the decisions and results of our experiments.

According to Wal [5], folksonomies are of two kind: narrow, in which the author is the first one to tag an object that cannot have the same tag associated to it more than once; and broad, in which the objects may receive the same tag more than once from different users. In narrow folksonomies, the number of tags is smaller, what confers a stronger relationship between tags and objects; in wide folksonomies, there is a stronger relationship between tags and users, who can freely tag the objects again and again. For this work, we use broad folksonomy because we rely exclusively on the tags provided by users, not relying on authors.

Körner [11] classifies the users that associate tags to the objects as categorizers and describers. Categorizers create tags to be used during their own usage (search and exploration) of objects – their vocabulary is related to their knowledge and interests. Describers, in turn, create tags to be used by other users, therefore their tags tend to be more general and of common sense. These characterizations have been targets of intense debate in the literature [8], therefore, one of our research questions, as presented, hypothesizes about how our objective users behave following the work of Körner.

López et al. [13] and Shih and Tseng [14] introduce algorithms for classification and indexing, respectively, of tagged learning objects. They discuss the challenges of these problems proposing innovative algorithms that demonstrate good potential. Although we do not work on the same research, our conclusions relate directly to the use of classification and indexing techniques as such methods depend on abundant tagging activity.

B. Works on designing and analyzing social tagging over the educational domain

Ahn and Dabbish [15] explain the design of EspGame, a computer game that motivates the students to take part of collaborative tagging. The action of the game is to show images that must be tagged by two students at the same time within a limited time period. After that, the game engine works to find correspondence between the tags of both users; in case of success, the tag is accepted and a new image is presented. The results of Ahn and Dabbish demonstrated that the correct motivation can be an important factor in amplifying the will to create tags; they discuss the use of their idea in many applications that rely on the correct description of digital content. Overall, they demonstrate that designing systems in which users are interested in tagging the content is not only viable, but also recommended.

Bateman et al. [9] analyze social tagging applied to e-learning by using tags collected from the interaction of students and professors with learning objects. In their study, they observe that the professors use a more specialized terminology than the students, and that an initial set of tags (a seed) must be provided during the earlier stages of the system. We follow their advice providing our students with such a set during the experiments; differently, we further discuss the behavior of the students, and the resulting vocabularies of tags, tracing some relevant hypotheses.

Dahl and Voussen [6] argue that in the narrow social tagging, each user collaborates more intensely with the group of users since they must choose which tags to use in a limited fashion. Dahl and Voussen discuss their thoughts raising the question of whether the vocabulary of tags used in e-learning systems converges to a limited set of tags. We conduct experiments to answer this same questioning.

In a recent work, Zervas and Sampson [16] evaluate how the motivation affects the enlargement of tagged learning repositories. Although they discuss some interesting issues about the influence of the profiles of the students, they do not put conclusive considerations about this interesting topic. In our research, we follow a similar investigation to settle, as much as possible, revealing remarks about how social tagging occurs in the educational domain.

Trant [17] present an extensive survey of works on social tagging and folksonomy, raising questions about the use of these techniques, criticism, potential, and open problems. Among the things they point out are the uncertainty about the applicability of social tagging and folksonomy, and about the
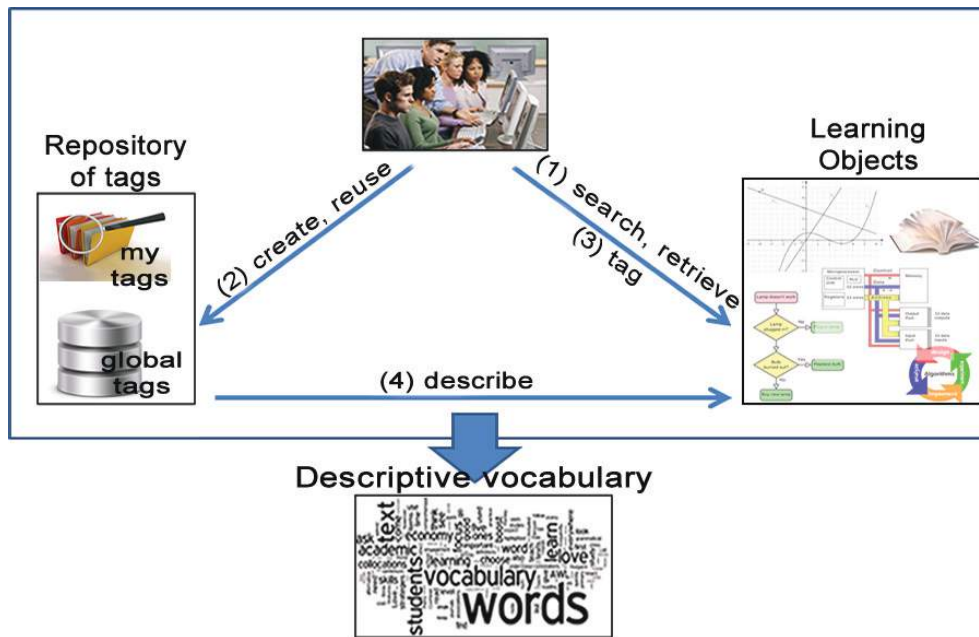
Fig. 1. The elements that constitute the triplet our of social tagging methodology: users (students), learning objects, and tags.

consistency about the vocabulary; these are some issues that we approach in the current research.

III. METHODOLOGY: SOCIAL TAGGING AND LEARNING OBJECTS CATALOGUING

Based on the presented concepts and related works, we have defined a methodology to answer the questions posed in section I. We conceived the tagging system as a triplet made of students, tags, and learning objects that, together, interact to form a descriptive vocabulary.

Figure 1 shows this triplet, which starts from the search and retrieval of learning objects (1); proceeds to the creation and reuse of tags (2); and evolves through the refinement achieved by the association of tags to the objects (3). These three steps cycle for some indefinite time, during which a repository of tags is built. The repository is organized as two sets: "my tags", the tags created by a specific user who is logged in the system; "global tags", the universe of all the distinct tags created by all users. Along time, some tags are recurrently used for describing the learning objects. The convergence of the tags to a stable descriptive set defines a vocabulary (4).

We noticed that previous works did not offer an adequate (open source and accessible) environment for experimentation according to our methodology and to our research questions. For this reason, our research group designed and developed the system TagLink, which is able to (1) retrieve learning objects, (2) create/reuse tags, (3) support the tagging of objects, and (4) manage a descriptive vocabulary of tags.

A. TagLink tool

TagLink was designed following the scheme presented in Figure 1. It allows students to search and retrieve learning objects from the Web using Google's Custom Search API (https://developers.google.com/custom-search/?hl=en). To do so, the students provide search terms and TagLink returns the objects and their corresponding links. Each link is processed as a learning object to which students can associate tags. TagLink is accessible at http://200.133.238.124/tag/.

TagLink tool is configured with a set of relevant repositories of learning objects. The repositories are registered in TagLink together with a priority indicator that specifies in which order they are to be searched. The results returned by TagLink – see Figure 2 – correspond, each, to one learning object that can be tagged with a new or with an existing tag – see Figure 3.

Fig. 2. Search interface of TagLink.

In TagLink, it is possible to register users and to organize them in groups – classes, or workgroups, for instance – so that it becomes possible to observe the behavior of specific sets of students. It also supports the retrieval of data about the tags: creation date, how many times it was used, who created it, corresponding objects, and so on; and the retrieval of data about the learning objects: which tags, times of use, origin, and so on.

Fig. 3. TagLink - creation and association of tags.

B. Experiment

We carried out an experiment with 168 students from the vocational education level (information technology technicians) at a country side school in Sorocaba, São Paulo, Brazil, called Fernando Prestes, belonging to Paula Souza Center. The students were instructed to search and retrieve learning objects related to their current courses and to tag these objects – we used broad social tagging. The activity of the students was recorded for analysis, as we report in the next sections.

A preliminary poll revealed that most of the students were regular users of social networks; and that they were familiar with tagging, but they had never used such functionality for educational purposes. The results of this preliminary poll led us to provide the users an initial training before using the system.

Two groups of students were set: Group A consisting of the older students, and Group B consisting of the younger students. We also defined two sets of search terms: **Ta** and **Tb**, both with 5 terms each – these sets answer for the initial seed necessary at earlier stages of tagging systems, as pointed out in Section II-B. The first set, **Ta**, contained generic terms for information technology, like logic, C#, Databases, Windows, and Linux. The second set, **Tb**, contained web programing terms, like JQuery, PHP, XNA, Android and Flash. We conducted the experiment in two phases, in Phase 1, Group A used set **Ta** and Group B used set **Tb**; in Phase 2, Group A used set **Tb** and Group B used set **Ta**.

The experiment took 2 months. Prior to the experiment we had an introductory period when the students learned about how to use TagLink and about the importance of the experiment. In the first month, we conducted Phase 1; and in the second month, we conducted Phase 2. At the end we had 2,019 distinct tags for 218 learning objects selected by 336 students.

IV. ANALYSIS

We analysed the data collected during the experiments to answer to the research questions.

A. Can social tagging successfully catalog e-learning objects?

For this question, we assume that the more number of times and the more students take part of the experiment, the better tagging catalog we shall have for our specific setting. This is a reasonable assumption since a tagging process is supposed

to satisfy the expectations of the very students who create the tags. Therefore, we analyze two aspects: the number of times and of students that created new tags – Figure 5; and the number of times and of students that reused existing tags – Figure 6.

In Figure 5, we can verify one first concentration in the data defining a nearly Gaussian-peaked distribution around 2 tags per student; and one main second concentration defining a nearly Gaussian-smooth distribution center around 12 tags per student. The first concentration is expected; since the participation in the experiment was optional, a significant fraction of the students created no more than 3 tags. The second main concentration revealed that 4 times as much students took part of the experiment creating from 4 up to 18 tags each. The participants of the second main concentration created 2,420 tags – not necessarily distinct – or nearly 11 tag creations per learning object.

In Figure 6, we can verify two peaks, one around 5 reused tags per student, and another one around 15 reused tags per student. The distribution now is shifted if compared to the distribution of new tags per student – Figure 6. In the place where there was a peak, now there is a valley; the events indicate that there was an increase in the participation of the students who did not create tags before – more to the left in the distribution. Meanwhile, a smaller set of students engaged even more in the experiment and increased the expected value – more to the right in the distribution. This is a curious observation, it shows three behaviors for the students, as derived from Table I: a set with students that only created new tags (11.3%), a set with students that only reused tags (8.9%), and a set of students that did both things (79.7%).

By considering Figures 5 and 6, we can affirm that, for our specific setting, the students satisfactorily participated in the tagging process by defining enough tags for the description and latter retrieval of objects. Our argument is based on the students navigation on the tag globe of Taglink (see Figure 2). After the experiment we asked to other students' group (200 students of 3 different course) to elaborate a study evolving the terms described on Ta and Tb (see Section III-B. The students used the Taglink's tag globe. During the navigation the students reported they found or not the learning object according to the terms they searched. We had a 98% of positive feedback. Therefore we suggest, with significant evidence, that social tagging can successfully be used in e-learning.

TABLE I. SUMMARY OF CREATION AND REUSE OF TAGS PER GROUP, AND PER PHASE.

Group	Phase 1		Phase2		Total	
	Created	Reused	Created	Reused	Created	Reused
A	1,126	648	144	303	1,270	951
B	1,329	933	139	363	1,468	1,296
Total	2,455	1,581	283	666	2,738	2,247

B. How do students behave according to Körner's classification: categorizers or describers?

For this question, we tracked the tags used by the students – in Table II, we present the tags the most used by the students. It is possible to notice that, except for tag "interesting", the tags are of general meaning and of common sense. This fact
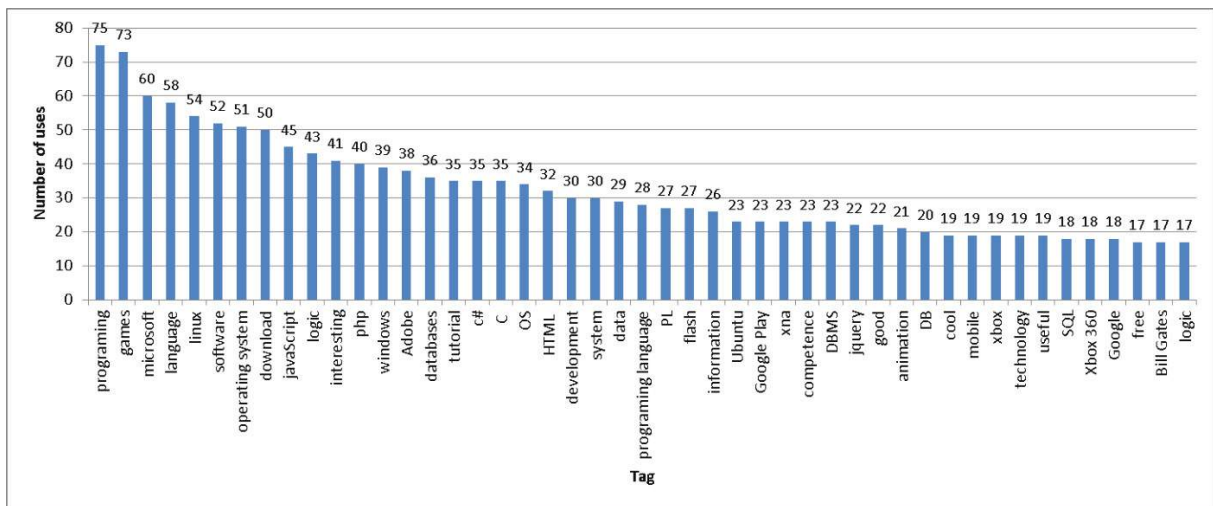
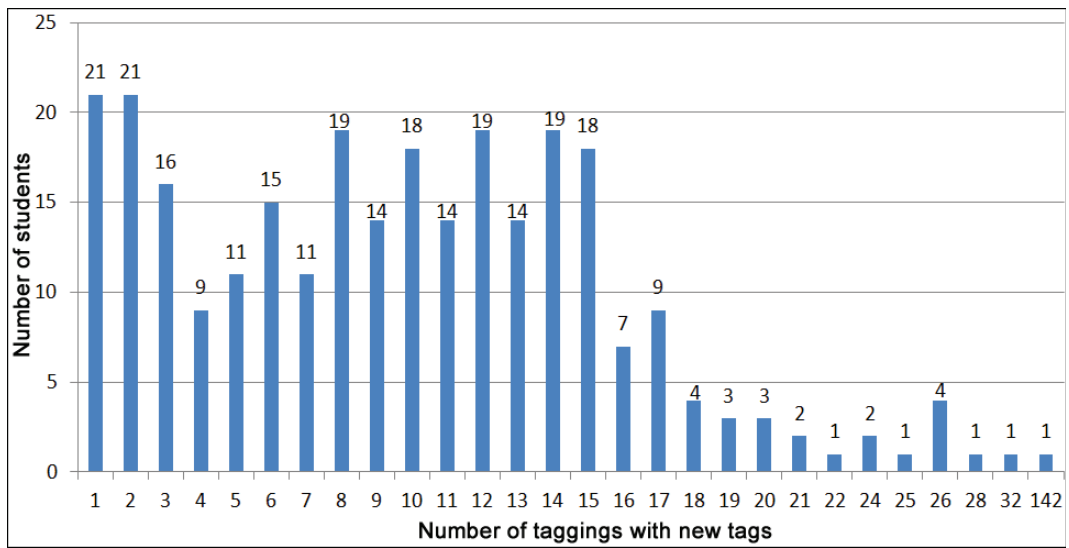Fig. 4. Distribution of the number of times each tag was used.

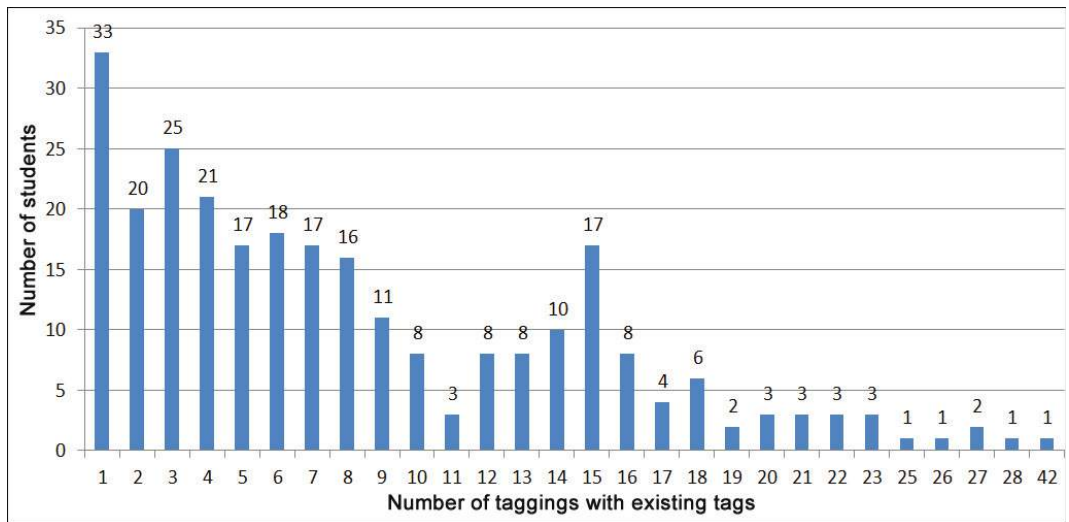Fig. 5. Distribution of the number of students per number of tags created.

Fig. 6. Distribution of the number of students per number of tags reused.

TABLE II. TAGS USED BY EACH GROUP IN EACH PHASE OF THE EXPERIMENT.

	Tag x Group x Phase				
	Phase 1		Phase 2		
tag	Grup A	Grup B	Grup A	Grup B	Total
programming	30	7	6	11	54
Microsoft	20	17	6	8	51
Linux	23	13	2	2	40
logic	27	0	1	5	33
software	13	10	7	10	40
language	16	19	3	9	47
operating system	14	10	7	10	41
JavaScript	0	28	9	0	37
SO	10	3	2	9	24
interesting	1	19	3	0	23
Total	154	126	46	64	390

indicates that the students had a *describer* profile, rather than a *categorizer* profile. For the sake of our research, this is a significant conclusion because it points that social tagging in e-learning tends to produce catalogs of general use, rather than catalogs of customized information. This is important for the following aspects:

- the catalogs are prone to be reusable by students in subsequent courses, who will use general terms for searching documents;

- the catalogs can be indexed more efficiently because the recurrence of tags can generate indexes guided by relevance;

- the e-leaning objects can be grouped (clustered) using term-frequency analysis, dismissing less meaningful and specific tags;

- with general common sense tags, it is possible to combine the e-learning search system with commercial search engines (educational and of general use), improving the retrieval of content with complementary objects.

We conclude that, although students tend to be categorizers in their virtual social activity – tagging what they see with personal perceptions and feelings, they are more likely to behave like describers when the social tagging occurs in the educational domain. Therefore we advocate, with relevant basis, that students behave according to Körner's *describer* profile.

C. Does social tagging converge to a well-defined descriptive vocabulary of tags?

For this question we had to analyze the number of new tags created in the system along the time of the experiment (two months). In Figure 4, we can see that the number of new tags behaves according to a Normal distribution with a peak close to the middle of the period – in the 28th day. The Normal distribution suggests that after the peak, the students create just a few more new tags, a number tending to zero as we get far from the expected value.

The Normal distribution, while valid, is not as strong as a descending power-law distribution would be; nevertheless, this fact is also interesting. Why did not the process behave like a power-law in which most of the new tags were created at the beginning of the period? The answer is quite straight when we consider that the experiment goes over a human-computer interface. In such systems, the user goes through a learning curve with three phases: slow beginning, steep acceleration, and plateau. We speculate, hence, that the left side of the Normal distribution – until nearly the 20th day – was affected by the learning period and that, only after, the users were able to fully work on TagLink and demonstrate their tagging profiles.

While Figure 7 shows that the tags are to stabilize after the 28th day, Figure 4 shows that a small subset of tags dominates the usage in the system. More precisely, there were 2,019 different tags; among them, 1,972 of them were used less that 17 times – not shown in the figure; and only 47 were used 17 times or more – shown in the figure. These characteristics describe a long heavy-tailed distribution indicating a strong imbalance as the students concentrate on a very small subset.

By comparing this dominant set of terms with the set of seed terms provided in the beginning of the experiment – see Section III-B – it is possible to observe a great intersection. This fact indicates that the seed of terms strongly influenced the vocabulary and that, possibly, this seed accelerated the process as suggested in other works – see Section II-B. Therefore, based on the evidences of Figures 7 and 4, we argue that social tagging is supposed to converge to a well-defined set of tags. We also argue that an appropriate set of seed terms may provide some control over this process, influencing the definition of the set of most frequent tags and, consequently, influencing how descriptive they will be.

V. CONCLUSION

We analyzed the practice of social tagging in the domain of e-learning aiming at answering questions related to applicability, profiling of users' behavior, and tagging vocabulary.

In a real setting with 336 students over a prototype named TagLink, we found that: (1) the use of social tagging is viable in the sense that students are inclined to build extensive catalogs over the learning objects; (2) despite their colloquial experience with tagging content in social networks, students will tag learning objects using descriptive (formal and general)
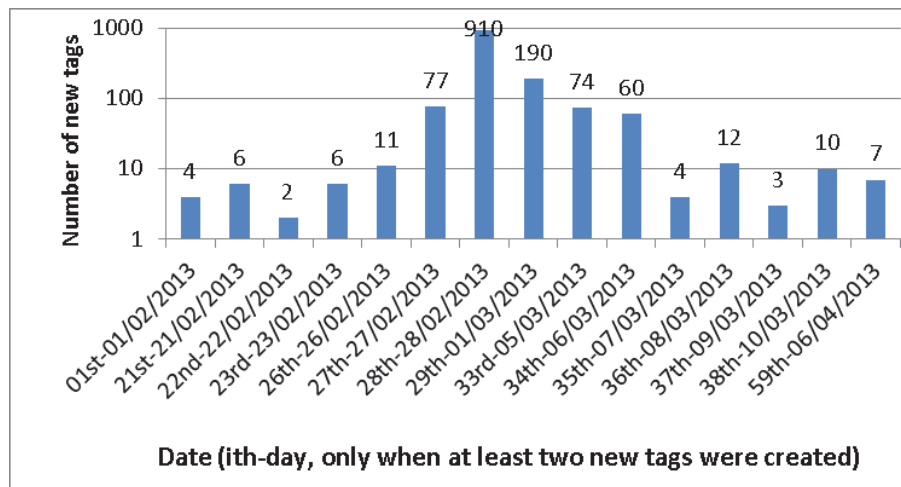
Fig. 7.   Distribution of the number of new tags created along time.

terms that aid the posterior use of objects catalogs; and (3) the vocabulary of terms converges to a "rich" subset of terms that answers for over 95% of the tags crated and/or reused by the students, indicating that guidance (an initial set of terms) can lead to a faster convergence and to an improved control over the process.

Although we used a specific setting in our experiments, we argue that our findings can lend judicious generalizations that should be investigated in further works. In any case, our experience brings insights that could be used as first assumptions in motivating the construction of social tagging learning systems. This possibility is the more general contribution of our work that, by paving the way for e-leaning endeavors, can stimulate the emergence of social-educational systems of broad usage.

REFERENCES

[1] Glavinic, Vlado and Granic, Andrina. "Hci research for e-learning: Adaptability and adaptivity to support better user interaction,". In *In: HCI and Usability for Education and Work. LNCS 5298*, 2008, pp. 359–376.

[2] Jaschke, Robert, Hotho,Andreas, Mitzlaff, Folke and Stumme, Gerd. "Challenges in Tag Recommendations for Collaborative Tagging Systems,". *Recommender systems for the social web*, pp. 65–87, 2012.

[3] McGreal, Rory. "Learning objects: A practical definition.". *International Journal of Instructional Technology and Distance Learning. 1(9)*, Vol. 1 (9), 2004.

[4] Vazquez, Arturo R. and Ostrovskaya, Yulia A. "Analysis of open technological standards for learning objects.". In *In Proceedings of 4th LA-Web, Puebla Cholula, Mexico*, 2006, pp. 105–108.

[5] Wal,   Thomas   V.   "Explaining   and   showing   broad and   narrow   folksonomies,".   *Off   the   Top   [blog]*, http://www.vanderwal.net/random/entrysel.php?blog=1635,   2005, Accessed: 13 December 2013.

[6] Dahl,Daniel and Vossen, Gottfried "Evolution of learning folksonomies: social tagging in e-learning repositories,". *International Journal of Technology Enhanced Learning*, Vol. 1, pp. 35–46, 2008.

[7] Milicevic,Aleksandra K., Nanopoulos, Alexandros and Ivanovic, Mirjana. "Social tagging in recommender systems: A survey of the state-of-the-art and possible extensions,". *Artificial Intelligence Review*, Vol. 33(3), pp. 187–209, 2010.

[8] Gupta, Manish, Li, Rui, Yin, Zhijun and Han, Jiawei. "Survey on social tagging techniques,". In: *Newsletter ACM SIGKDD Explorations*, Vol. 12(1), pp. 58–72, 2010.

[9] Bateman, Scott, Brooks,Christopher and Mccalla, Gordon. "Applying collaborative tagging to e-learning,". *In Proceedings of ACM WWW*, 2007.

[10] Garcia-Penalvo, Francisco J., Morales, Erla and Barron, Angela. "Learning objects for e-activities in social web," *WSEAS Transactions on Systems*, Vol. 6 (3), pp. 507–513, 2007.

[11] Korner, Christian. "Understanding the motivation behind tagging," *ACM Student Research Competition*, Vol. 9, 2009.

[12] Shepitsen, Andriy, Gemmell, Jonathan, Mobasher, Bamshad and Burke, Robin. "Personalized recommendation in social tagging systems using hierarchical clustering,". In *Proceedings of the 2008 ACM conference on Recommender Systems*, 2008.

[13] Lopez, Vivian F., Prieta, Fernando de la, Ogihara, Mitsunori and Wong, Ding D. "A model for multi-label classification and ranking of learning objects," *Expert Systems with Applications*, Vol. 39(10), pp. 88788884, 2012.

[14] Shih, Wen-Chung and Tseng, Shian-Shyong. "Folksonomy-based indexing for retrieving tutoring resources,". In *2012 Seventh IEEE International Conference on Wireless, Mobile and Ubiquitous Technology in Education*, 2012, pp. 97–101.

[15] Ahn, Luis von and Dabbish, Laura. "Labeling images with a computer game,". In *CHI '04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2004, pp. 319–326.

[16] Zervas, Panagiotis and Sampson, Demetrios G. "The effect of users' tagging motivation on the enlargement of digital educational resources metadata,". *Computers in Human Behavior*, Vol. 32, pp. 292–300, 2014.

[17] Trant, Jennifer. "Studying social tagging and folksonomy: A review and framework," *Journal of Digital Information*, Vol.10(1), 2009.