

Types of Tags for Annotating Academic Blogs

Lei Li¹, Daqing He², Danchen Zhang², Yu Chi², Chengzhi Zhang¹

¹Nanjing University of Science and Technology

²University of Pittsburgh

Abstract

Academic blog sites are popular academic information exchange platforms, and they have been widely used in recent years. Blogs in those sites are often annotated with tags, and the tags can help to describe, organize and retrieve these blogs. However, it is still unknown what types of tags are frequently adopted for annotating academic blogs. In this poster, we present survey results for detecting the usage of tag types, and its changes with the bloggers' demographic information. We believe that our study can benefit users in their access to academic blogs and help the academic blog websites improve their services.

Keywords: tag types; blog tagging; social tag; academic blog; academic social media

DOI: Citation info is to be added.

Copyright: Copyright is held by the authors.

Acknowledgements: This work is supported by the Major Projects of National Social Science Fund (No. 16ZDA224).

Contact: lileiwelldone@gmail.com; dah44@pitt.edu; daz45@pitt.edu; yuc73@pitt.edu; zhangcz@njust.edu.cn (corresponding author. +86-84315963)

1 Introduction

Social tags have been widely used for annotating various kinds of online resources, such as blogs, pictures, books, videos, and music. The tags can help users in online resources retrieval and organization. Previous work showed that tags can be divided into many types, and they proposed plenty of tag type taxonomies for online resources of different kinds, e.g., book (Golder & Huberman, 2006; Wu, He, Qiu, Lin, & Liu, 2012), movie (Sen et al., 2006), TV program (Melenhorst & Setten, 2007) and URL (Xu, Fu, Mao, & Su, 2006). And these studies also confirmed that knowing tag types can enhance the effect of the tag recommendation (Xu et al., 2006) and resources search (Bischoff, Firan, Nejdil, & Paiu, 2008).

Academic blog website, which supports scholars to share and to acquire academic related information through the posted blogs, has emerged in recent years (Li & Zhang, 2016). Few studies, however, have investigated the tag types of the academic blogs, which are created by academic bloggers. Moreover, the tag type taxonomy is context-sensitive, which makes it inappropriate to apply the existing generic tag type taxonomy to the tags of academic blogs. Furthermore, it is still unknown what kind of information the academic blog tags convey. In this study, we combine the tag type taxonomies proposed by previous studies and the unique characteristics of blog tags to propose the tag types that are applicable to academic blogs. To achieve this goal, a survey instrument is adopted to explore the types of tags that are preferred by the bloggers for annotating academic blogs.

Through understanding the tag types used by bloggers for academic blogs, a tagging system could recommend proper tags to the blogs that still need tags. The users of the blog site thus has opportunities to more effectively access the relevant blogs, which could be a significant factor that influences scholars to be more actively engaging in academic social networks (Bik & Goldstein, 2013).

2 Method

2.1 Research site

We chose blog.sciencenet.cn¹ to send the questionnaires. [Blog.sciencenet.cn](http://blog.sciencenet.cn) is one of the most popular academic blog sites in China, and its registered users all have to provide real names, research institutions or

¹<https://en.wikipedia.org/wiki/ScienceNet.cn>

universities, and research fields. The majority of its users posted academic blogs to communicate with others, and the author of each blog assign tags to his/her blogs.

2.2 Sampling

Before sending out our questionnaire, we conducted a sampling on the users. Our target users are those with many tagged blogs because they have rich tagging experience, and they log in the website more frequently thus have a greater chance to reply our questionnaire. We crawled each blogger’s webpage, which includes the blog visited times, number of friends (#friend), number of blogs (#blog), registration time, last visited time, and last published time. In total, we collected 44,509 bloggers’ information. Then based on the bloggers’ number of published blogs, we selected top 5,000 bloggers. After further manual checking on whether these bloggers indeed tagged academic blogs, we finally identified 4,111 bloggers to send our questionnaire.

2.3 Questionnaire design

The questionnaire aims to investigate bloggers’ demographic information and their blog tagging behaviors. The demographic information is used to detect the association between the blogger’s basic information and the types of tags she uses.

Guided by the tag type taxonomies in previous work(Xu et al., 2006; Golder & Huberman, 2006; Melenhorst & Setten, 2007; Sen et al., 2006; Bischoff et al., 2008; Heymann, Paepcke, & Garcia-Molina, 2010), we classified the tags into four types, namely content-based tags, context-based tags, subjective tags and organizational tags. Then according to the characteristics of the blog, survey items of each tag type were proposed. Table 1 shows the survey items belonging to the above four tag types.

In the questionnaire, firstly, bloggers were asked to answer the question about their demographic information which includes the blogger’s discipline, education, gender and age. Then base on their tagging experience,the bloggers were required to rate their preference of using the tag types (showed in table 1) for annotating the blog. They can rate each tag type with a Likert scale from 1 to 5.

Tag types	ID	Items
Content-based tags	T1.1	Describing the blog topic, and existing in the blog’s content
	T1.2	Describing the blog topic, but not existing in the blog’s content
	T1.3	Describing the blog topic, and existing in the blog’s title
	T1.4	Describing the categories of the blog
Context-based tags	T2.1	The blog’s publisher
	T2.2	The blog’s published time
	T2.3	The blog’s published location
	T2.4	The source of the blog
Subjective tags	T3.1	The opinion to the blog
Organizational tags	T4.1	Self-organization tags

Table 1: Survey Items for the Tag Types

3 Results

3.1 Participants

Launched from 09/16/2013 to 05/30/2014, 499 questionnaire responses were received, and of which 444 questionnaires were completed. The following analysis is based on these 444 questionnaires. Table 2 summarizes the demographic information of the participants.

Demographic information	Options	N/%
Gender	Male	387 (87.2/%)
	Female	57(12.8/%)
Age	21-30	211(47.5/%)
	31-40	172(38.7/%)
	41-50	51(11.5/%)
	Above 50	10(2.3/%)
	Disciplines	Engineering and technology science
	Natural science	178(40.1/%)
	Humanities and Social Sciences	43(9.7/%)
	Medical Science	28(6.3/%)
	Agricultural science	17(3.8/%)
Education	Master candidate	55(12.4/%)
	Master	94(21.2/%)
	Doctor candidate	102(23.0/%)
	Doctor	193(43.5/%)

Table 2: Demographic Information of the Participants

3.2 Usage of tag types

Table 3 shows the 444 participants' usage of tag types for academic blogs. The top three tag types all belong to the content-based category, which are the tags (1) existing in the blog's content, (2) existing in the blog's title and (3) describing the categories of the blog. Meanwhile, the context-based tags that annotate the blog's author, published time and published location are less used.

Tag types	Items	1(low)	2	3	4	5(high)	M
Content-based tags	T1.1	8.6%	14.9%	21.4%	32.9%	22.3%	3.45
	T1.2	18.9%	28.8%	32.2%	16.2%	3.8%	2.57
	T1.3	5.9%	14.6%	31.5%	35.6%	12.4%	3.34
	T1.4	7.7%	16.4%	34.2%	33.3%	8.3%	3.18
Context-based tags	T2.1	35.6%	27.7%	23.0%	9.9%	3.8%	2.19
	T2.2	44.1%	22.7%	20.3%	9.2%	3.6%	2.05
	T2.3	43.9%	27.3%	17.3%	9.7%	1.8%	1.98
	T2.4	18%	20.0%	22.7%	23.9%	15.3%	2.98
Subjective tags	T3.1	20.3%	22.1%	30.0%	21.6%	6.1%	2.71
Organizational tags	T4.1	12.4%	20.0%	30.4%	28.6%	8.6%	3.01

Table 3: The Usage of Tag Types

3.3 Comparison the usage of tag types on different bloggers

The association between the blogger's demographic information and the usages of the tag types is reported in this section.

To find the gender's effect, ANOVA test was conducted. We find that female bloggers have the significantly stronger will to use the self-organizational tags to annotate academic blogs ($F=7.098$, $p<.01$). But there is no significant difference on the usage of other tag types between genders.

ANOVA test also shows that the younger bloggers with under 41 years old, in comparison with older bloggers, have the significantly higher preference to use the tags that describe the blog topic but do not exist in the blog's content ($F=8.782$, $p<.01$). Meanwhile, these younger bloggers also have a significantly higher probability to use the tags for annotating the blog's published time ($F=7.652$, $p<.01$). Other types show no significant difference among age groups.

Furthermore, through conducting ANOVA test, we find that bloggers with master degree or being master candidates, in comparison to those with doctor degree or being doctor candidates, are more likely to use subjective tags ($F=10.009$, $p<.01$) and the tags for annotating the blog's published time($F=9.268$, $p<.01$). No other statistical difference was found across user's education levels. Similarly, ANOVA test shows no significant difference across the user's disciplines.

4 Discussion and Conclusion

In this work, we explored what types of tags bloggers use for annotating academic blogs. To our knowledge, this is the first study that detects the tag types employed in academic blog platform. The results imply that tag types that describe the blog content, including those existing in the content and the title of a blog are more preferred by the blogger to annotate the academic blog. Tags that annotate the blog context information, such as the blog's author, published time and published location, are less used. Based on the tag types' preference for academic blogs, the academic blog website can improve the bloggers' satisfactory with the tags recommendation.

This study also give us the evidence that not all tags are equally useful for different users. For example, the female bloggers are more willing to use self-organizational tags. The bloggers under 41 years old have the significant higher preference to use the tags that describe the blog topic but do not exist in the blog's content, and the tags for annotating the blog's published time. These findings imply that the tag recommendation should distinguish the bloggers demographic information to be conducted.

Future work includes studying the other factors that influence the academic blog tag usage, such as the time effect and the platform effect. Finally, we will try to apply our findings to the academic tagging systems to provide the bloggers with better service.

References

- Bik, H. M., & Goldstein, M. C. (2013). An introduction to social media for scientists. *PLoS Biol*, 11(4), e1001535.
- Bischoff, K., Firan, C. S., Nejd, W., & Paiu, R. (2008). Can all tags be used for search? In *Proceedings of the 17th acm conference on information and knowledge management* (pp. 193–202).
- Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(32), 198-208.
- Heymann, P., Paepcke, A., & Garcia-Molina, H. (2010). Tagging human knowledge. In *Proceedings of the third acm international conference on web search and data mining* (pp. 51–60).
- Li, L., & Zhang, C. (2016). Characterizing users tagging behavior in academic blogs. In *Proceedings of the 16th acm/ieee-cs on joint conference on digital libraries* (pp. 215–216).
- Melenhorst, M., & Setten, M. V. (2007). Usefulness of tags in providing access to large information systems. In *Professional communication conference, 2007. ipcc 2007. ieee international* (p. 1 - 9).
- Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., . . . Riedl, J. (2006). Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on computer supported cooperative work* (pp. 181–190).
- Wu, D., He, D., Qiu, J., Lin, R., & Liu, Y. (2012). Comparing social tags with subject headings on annotating books: A study comparing the information science domain in english and chinese. *Journal of Information Science*, 0165551512451808.
- Xu, Z., Fu, Y., Mao, J., & Su, D. (2006). Towards the semantic web: Collaborative tag suggestions. In *Collaborative web tagging workshop at www2006, edinburgh, scotland*.