Keck, Helena; Heck, Tamara

Improving tagging literacy to enhance metadata and retrieval for open educational resources

Proceedings of the Conference on Learning Information Literacy across the Globe, Frankfurt am Main, Germany, 10th of May 2019. Frankfurt am Main : DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation 2019, 10 S.

Mitglied der
Leibniz-Gemeinschaft

Improving tagging literacy to enhance metadata and retrieval for open educational resources

Helena Keck[1] and Tamara Heck[1\[0000-0001-5519-6395\]]

[1] DIPF | Leibniz Institute for Research
and Information in Education, 60323 Frankfurt, Germany
`helena.keck@stud.h-da.de|heck@dipf.de`

Abstract. The growing amount of open educational resources and the diversity on learning and teaching makes social tagging attractive for the educational field. Social tagging services become valuable in contexts where users can support the enrichment, sharing and management of relevant resources. Potential benefits are the enrichment of incomplete metadata, which is crucial to offer effective retrieval services. However, user tagging skills need to be fostered if users shall effectively contribute to the idea of collaboratively sharing and creating educational resources. We aim at fostering user tagging literacy. We analysed tags and user behaviour from a German referatory for educational resources. Our results show that users apply specific tags for their learning and teaching resources that we tried to assign to additional tag categories. Based on our results, we suggest improving such services with a more user-centric approach that supports the development of user competencies on social tagging. We will contribute to a better understanding of user tagging behaviour in services focusing on educational resources. On the one hand, this will help us to improve current services. On the other hand, we are able to build services that foster tagging literacy. This will be beneficial for users, which will be able to better manage their digital resources, and for infrastructure providers, which can apply user-generated data to improve their services.

Keywords: tagging literacy, open educational resources, user behaviour

1 Introduction

Digital educational resources allow easy access and storage for relevant learning and teaching material. Openly licenced – for example with a creative commons licence – those open educational resources (OER) allow every user to retain, reuse, revise, remix and redistribute them [1]. OER include all kinds of resources, like learning, tools and implementation resources, which have diverse granularity levels, i.e. OER can be single learning objects like an open textbook or whole learning courses like MOOCs [2]. A major aspect often underestimated is searching and finding OER. Repositories or referatories for OER offer basic search functions [3–5]. More advanced systems could improve OER retrieval and usability to allow users to find proper resources for their diverse needs. To further improve retrieval, digital resources need to have complete and structured metadata. An OER search service would profit from rich and properly-described metadata fields, which are the basis to establish effective filter

functions to retrieve OER. LOM and LRMI are two common standards for OER, but there exist many variations. For some services, editors professionally add and manage metadata while applying common vocabulary standards. However, considering the financial capacities needed and the growing amount of digital resources, getting support from web users might be crucial. A challenges is that those actors assign any terms to describe their relevant resources, e.g. via tagging, and most of them are not aware of any standard. If users become more competent in tagging, they can contribute to completing missing OER metadata that improves retrieval.

The following paper aims at improving OER retrieval while fostering users tagging literacy to make users able to contribute to OER metadata enrichment. A first step is to get insights into user tagging behaviour. Based on the results from tagging data in a German OER referatory, we suggest options to improve an OER tagging and retrieval system that fosters positive user tagging behaviour to be valuable for the proposed goals. The research questions are:

- Can we classify tags of educational resources according to existing tag categories to distinguish relevant tags for enriching metadata and improving search?
- In which way can a tagging service effectively support users in improving their tagging to support metadata enhancement and retrieval?

We will first relate to literature on social tagging and tagging literacy, before we introduce our methodological approach and referatory we refer to. Section 3 presents the results. In the discussion section, we give recommendations for enhanced tagging literacy within an OER tagging service, before we conclude on our research.

Social tagging or collaborative tagging is the process to enrich digital resources with the help of web users: users add relevant resources (bookmarks) to a service and describe them with freely chosen keywords (tags). Tagging happens in a social environment, usually shared and open to others. Those social bookmarking services have a folksonomy structure with user-bookmark-tag relations [6], i.e. "folksonomy is created from the act of tagging by the person consuming the information" [7]. Tagging is „an act of organizing through labelling, a way of making sense of many discrete, varied items according to their meaning" [8]. Users benefit in different ways. They describe their resources with tags to make them findable and distinguishable from other resources. As well, they can search further relevant resources via tags from all users of a system. Thus, all users collaboratively support each other in the management of their resources. Social tagging can be beneficial when there is too much content to classify or there are no editors that take the 'librarian' role [8]. Besides, web services use tags to automatically index resources.

Enabling users to freely index their resources without having to care about guidelines or terminologies causes problems because tag collections lack common vocabulary. Moreover, many tags show grammatical or typing errors, or seem meaningless for those who have not applied them. Services that want to use tags for resource data enrichment or user recommendations need to handle various language requirements such as multilingualism, semantic and morphological problems [9].

To overcome those challenges and benefit from social tagging, users need to become competent in tagging. They need to learn to use tags that describe resources in effective and complete ways. Moreover, studies showed that there are different types of users, describers and categorisers [10, 11]. Describers generally have a broad tag vocabulary, including many synonyms. Categorizers rather apply tags for structural purposes, like navigating through their resources. They generally do have fewer synonyms on their tag vocabulary. Identifying those user groups could help in pre-determine tags for possible metadata enrichment, for example describers' tags might be relevant to find new formal vocabulary terms [11].

We see tagging literacy as part of information literacy [12]. The Association of College & Research Libraries describes information literacy skills within six frames [13]. They include dispositions like "see themselves as contributors to scholarship rather than only consumers of it;", and "value user-generated content and evaluate contributions made by others" [13]. The DigComp 2.0 framework formulates similar digital competencies including "to use digital tools and technologies for collaborative processes, and for co-construction and co-creation of resources and knowledge" and "to share data, information and digital content with others through appropriate digital technologies" [14]. Social tagging is a process of co-creation, in which all users support each other in their resource and data management and beneficially contribute to a folksonomy structure that supports the retrieval of relevant resources.

As such, there exist diverse strategies to improve tagging literacy and support users in their tagging activities. For example, a system may allow users to re-edit or delete tags, and offer formatting guidelines or checklists [15]. A tagging interface could hint users to tags that are informationally powerful [16]. A service may as well propose tags used by others in similar contexts [11]. Finally, studies investigated the use of tag clouds and found them to be a positive support for finding suitable resources: "The user can search and browse science education resources using an appropriately formatted tag cloud produced by the tags that all users of the tool have offered" [11].

To help users to become tag literate, we first need to get deeper insights in how users apply tags. Categories help to distinguish between different meanings of tags and their purpose. There are different approaches on how to categorise social tags. One approach [8] differs between seven several functional tags: Identifying what (or who) it is about, identifying what it is, identifying who owns it, refining categories, identifying qualities or characteristics, self-reference and task organizing. Other categories are be "foreign language" [17] or "location" and "time" [18]. Lawson differs between objective and subjective social tags [19] and Wu between topical and non-topical tags [19]. Gupta et al. [20] base their tag categories on those approaches and come up with eleven categories. In our study, we refer to those categories.

OER services can benefit from user tagging activities in different ways. First, the growing amount of OER requires cost-efficient solutions to enrich OER metadata. Metadata is inevitable when OER services want to offer effective search function to users to be able to find their most relevant resource. Second, tagging OER allows users to describe their relevant resources according to their own purposes, which supports individual learning and teaching. Standard vocabularies might be insufficient to describe the diverse user needs. Third, tags allow a better description of new and in-

novative resources as structured vocabularies might become archaic. Regarding the innovativeness of OER and learning and teaching concepts in general, tags might allow for an enrichment of those vocabularies. So far, a few studies analysed user tagging for educational resources [11, 21], mainly with the purpose to enhance metadata and give resource recommendations [22]. This paper does not focus not on the techniques of a system, but wants to analyse user tagging behaviour to suggest a concept to improve user literacy.

2 Method

We categorized user tags and had a closer look at user behaviour considering describers and categorizers. We used tags from our service *Edutags* described below, where we have direct data base access. We collected our data via SQL queries on October 19th, 2018 and December 13th, 2018. We have two data sets, a) all tags, and their number of times used by any user of the service, b) users-resource-tag sets, where a single user bookmarked a resource for the first time, i.e. excluding OER automatically collected by the system (see 2.1) and excluding bookmarks that other users bookmarked before to not influence tagging behaviour.

Data set *a* was the basis for categorising our tags. Within *a*, we did minor data cleaning and deleted the automatically generated tags *import*, *Import*, *WONG-Import* and *import_delicious*. We aimed for categorising 10 % of the most used tags within the service and checked the 1196 mostly used tags. We merged synonyms and corrected minor grammar errors. We also checked on acronyms, words that occur in German and English, in singular and plural forms, different spelling and duplicates. That resulted in 984 unique tags and this makes up about 10% of the total number of tags that have been used at least twice (Table 1).

Table 1. Number of tags and categorized tags from data set *a*.

Description	#Number
Total number of tags	27084
Number of tags used > 1	9843
Total number of categorized tags	984

Table 2. Unique numbers of data set *b*.

Description	#Number
Bookmarks	14793
Tags	885
Users	955

Data set b was the basis to assign a larger tag set to categories and look for user describers and categorisers. We deleted the automatically generated tags import, Import, WONG-Import and import_delicious. In addition, we deleted user-bookmark-tag sets,

where a user or tag was missing. This resulted in 73425 tags (not cleaned), 17803 single bookmarks, and 1194 single users. Applying the manually assigned categories from data set a, we received a sub set of b (Table 2) that includes all resources with tags that could be assigned to categories. Note, we did not have any information on the tagging order of a single resource.

2.1 Edutags – a digital resource referatory

Edutags is a reference service for digital educational resources, established 2011 [4]. *Edutags* focuses on teachers and wants to support them in seeking teaching materials and teaching ideas. Users can search for resources and assign them as individual bookmarks in their profiles. *Edutags* applies two ways of collecting OER references. A crawler automatically collects OER and their metadata stored by services from ten cooperating partners. Additionally, users are able to add their own resources to the service. When users add a resource, they must add at least one tag to describe their resource. Hereby, they see a list of their own tags used before. Additionally, users can comment and rate bookmarked resources. Users can search for a resource via a tag cloud, where the colour of the tag terms refers to their assignment frequency.

2.2 Problems with tag categories

We could not explicitly assign every tag to a single category. In most cases, we lacked context and could not identify any tag meaning. In some cases, we could identify the exact meaning by looking more closely at the attached resource. Another occurring problem is two-word tags like *biology lesson* or *English lesson* that are subject tag and domain tag at the same time. In our study we classified those tags as subject tags (Table 3). Other tags included a name, date, place or a resource type and we were unsure if those describe the context or the content of a resource. We categorised these kinds of tags as factual tags. Subject tags (like school subjects) are helpful as our results indicate, but the granularity of subject distinction is arguable. In our study, we assigned tags like *biochemistry* as subject tags and applied the rule that they are subject tags if the tag consists of only one word. We as well had several acronyms, for which we could not find the right meaning and put them into non-classifiable tags. At last, there exist tags that compound two or more words and meanings, like *math primary school*. It is a combination of a subject tag and a domain tag. So, we decided that one tag could be a member of more than one tag category.

3 Results

Our used tag categories are based on the tag classification by [20]. However, our 984 tags represented only four of the eight categories (first four in table 3). Instead, we realised that users applied more explicit descriptions for educational resources. Many tags referred to school subjects, school types, types of materials and licenses. We did not assign them to content-based tags because we assume that users want to be more

6

specific on their resource content. Thus, we added four additional categories, which are subject tags, license tags, resource-type tags and domain tags (Table 3).

Table 3. Tag categories based on [20], with four additional OER categories.

Tag category	Description
Content-based tags	These tags identify the concrete content of the resource
Ownership tags	The ownership tags determine who owns the resource. We merged this category with context-based tags [19] that describe the context in which the object was saved or created. Example: *coer13* (workshop).
Factual tags	The factual tags identify facts about an object, for example people, concepts or objects. Most of our tags are geographical.
Attribute tags	These tags are inherent attributes of an object, which might describe qualities or characteristics, like *current*, *mobile*.
Subject tags	They characterise school subjects and disciplines like *maths*, *biochemistry*.
Domain tags	They describe the domain of the resource, which include the education level (*primary*, *secondary*) and class level.
Resource-type tags	They describe all kind of learning material and media type. For example: *worksheet* or *video*. Some of the resource type tags are more specific than others.
License tags	This tag describes the creative common license of the resource.

Fig. 1 shows the distribution of tag categories from data set *a* (left chart). 0.79 % of existing tags could not be assigned because we could not determine the meaning. Of the 984 unique tags, 27 tags were assigned to two categories and one tag to three. Thus, the distribution of categories is based on 1012 tags. Fig. 1 as well shows the distribution of tag categories from data set *b* (right chart). There are no noticeable differences in the overall distribution of categories. However, the newly applied cate-
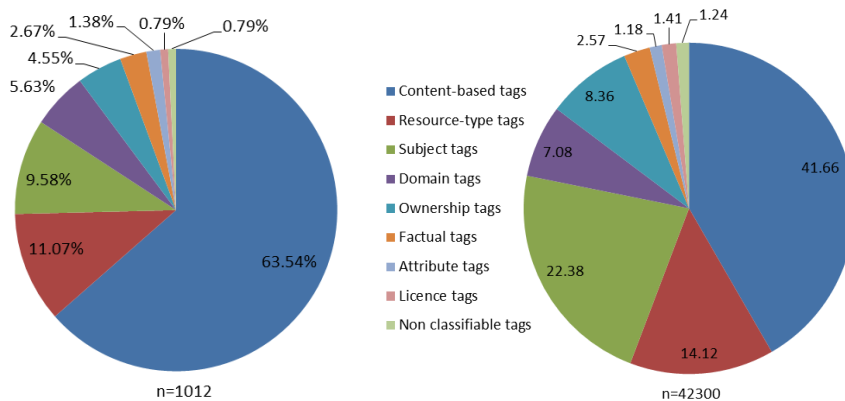

Fig. 1. Tag category distribution from data set *a* (left) and data set *b* (right).
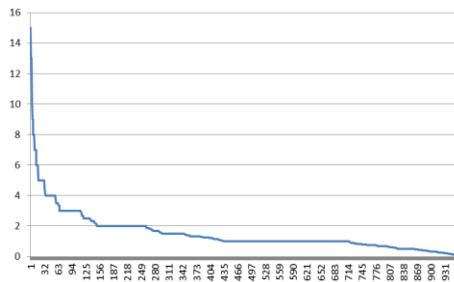
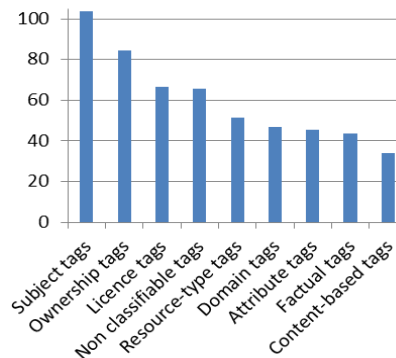**Fig. 2.** Tag/resource ratio per user, data set *b*.

Fig. 3. Average number of users assigning a common tag in a category, set *b*.

gories subject tags and resource-type tags have a higher percentage, which might stress their importance for users.

On average, users assigned 2.8 tags to a resource. Fig. 2 shows the tag/resource ratio of a single user, which can be an indicator of describer users, who would have a higher score according to the variety of tags they use [10]. That means, users on the left hand side of the diagram in Fig. 2 could be identified to look for new metadata vocabulary [11]. Another hint of finding relevant metadata vocabulary is the number or average number of users that assign a single tag. Here, results show that on average many users assign tags from our additional OER categories (Fig. 3). Tags applied by a high number of users can be considered for new metadata vocabulary, e.g. to expand determined values in specific metadata fields.

4 Discussion

Tags for OER show classic categories from the research literature. However, users seem to need more additional categories that specifically describe the purpose of OER within learning and teaching. We added four more tag categories that describe educational purposes and our data shows that users often apply tags from those categories. Those categories allow for a finer distinction between resources than having a single content-based category, and users are quicker in finding relevant educational material.

We tried to assign our tag categories to the LOM-CH metadata standard (Table 4). Not all tag categories exactly match a unique field, e.g. factual tags might fit into three fields. However, it is possible to assign tag categories to current standard metadata fields. Current OER repositories already offer some relevant search filters, e.g. Elixier (bildungsserver.de/elixier), a service for educational resources, which offers filters like keywords, education level, resource type and license. This shows a great match with our tag categories. A challenge for users might be the different field values applied by services, like e.g. in the resource-type category or with regard to values for media formats. Here, it would be desirable to agree to common values. Improving tagging literacy is a process that needs continuous fostering. Users need to be aware of the benefits that come with appropriately assigned tags.

Table 4. Tag categories assigned to existing LOM-CH fields.

LOM-CH	Tag Category	Example
1.4 Description	Content-based tags	Fractional arithmetic
5.2 Learning resource type	Resource-type tags	Worksheet
5.8 Difficulty	Attribute tags	simple
1.6 Coverage 2.1 Version 2.3 Contribute	Factual tags	Brazil
9.1 Purpose 9.2 Taxon Path 10.2.1.4 Scope	Subject tags	Biology
5.6 Context 9.1 Purpose 9.2 Taxon Path 10.2.1.4 Scope	Domain tags	Primary school
6.2 Copyright and other restrictions	License tags	CC-BY-SA
2.3 Contribute	Ownership tags	DBS-Wiki-KW

As such, we suggest implementing and evaluating the following functions for an OER service that allows tagging.

- Recommend OER related tag categories to users to make them aware of tagging options. A system could ask users to assign concrete categories during the book-marking process.
- Guide users through their search process. A service could hint users to relevant filters to make them aware of search options.
- Show users their own tags and allow them to edit those. Users would be able to correct typing errors, or merge synonym tags [12].
- Show users the most relevant metadata fields of OER. This makes users aware of missing metadata and motivates them to add additional tags.
- Show users the user network and its activities. This might foster user collaboration and awareness of the benefits of collaborative tagging.

Those suggestions focus on a user-oriented approach that supports user needs and

Fig. 4. Example of a tagging interface.

intentions. This will help users developing relevant competencies to be able to self-manage OER sharing and retrieval. We will do further research on evaluating those proposals via designed web interfaces (Fig. 4). Other studies as well show promising results with regard to hedonic features [23].

5   Conclusion

The growing amount of OER as well as their diverse context-based and subjective purposes makes social tagging attractive for the OER field. However, to overcome challenges and get the most beneficial user support, we argue that OER services need to foster user tagging competencies. We analysed tags to identify user tagging behaviour. Users applied specific categories to better describe the educational purpose and context of their resources. We therefor proposed service functions to foster user tagging and improve tagging competencies. We aim at designing an exemplary service interface and first evaluate it with regard to its design and usability aspects. Such an established system would as well allow assessing changes in user tagging literacy.

6   References

1. Hylén, J., van Damme, D., Mulder, F., D'Antoni, S.: Open Educational Resources. Analysis of Responses to the OECD Country Questionnaire. OECD Education Working Papers 76 (2012). doi:10.1787/19939019
2. Kerres, M., Heinen, R.: Open Informational Ecosystems. The missing link for sharing resources for education. The International Review of Research in Open and Distributed Learning 16 (2015). doi:10.19173/irrodl.v16i1.2008
3. Zervas, P., Alifragkis, C., Sampson, D.G.: A quantitative analysis of learning object repositories as knowledge management systems. Knowledge Management & E-Learning: An International Journal 6, 156–170 (2014)
4. Heinen, R., Blees, I., Kerres, M., Rittberger, M.: Open educational resources and social bookmarking: Connecting Users and Editors. In: International EIF / LINQ Conference 2014 (2014), https://learninglab.uni-due.de/user/3/biblio?page=3
5. Santos-Hermosa, G., Ferran-Ferrer, N., Abadal, E.: Repositories of Open Educational Resources. An Assessment of Reuse and Educational Aspects. IRRODL 18, 84+ (2017). doi:10.19173/irrodl.v18i5.3063
6. Peters, I.: Folksonomies. Indexing and Retrieval in Web 2.0. De Gruyter Saur, Berlin, Germany (2009)
7. Vander Wal, T.: Folksonomy coinage and definition (2007), http://vanderwal.net/folksonomy.html
8. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32, 198–208 (2006). doi:10.1177/0165551506062337
9. Siever, C.M.: Multimodale Kommunikation im Social Web. Forschungsansätze und Analysen zu Text-Bild-Relationen. Peter Lang Edition, Frankfurt am Main (2015)

10. Körner, C., Kern, R., Grahsl, H.-P., Strohmaier, M.: Of categorizers and describers. In: Chignell, M. (ed.) Proceedings of the 21st ACM conference on Hypertext and hypermedia, p. 157. ACM, New York, NY (2010). doi:10.1145/1810617.1810645

11. Zervas, P., Sampson, D.G.: The effect of users' tagging motivation on the enlargement of digital educational resources metadata. Computers in Human Behavior 32, 292–300 (2014). doi:10.1016/j.chb.2013.06.026

12. Dennen, V.P., Bagdy, L.M., Cates, M.L.: Effective Tagging Practices for Online Learning Environments: An exploratory study of approach and accuracy. Online Learning 22, 103–120 (2018)

13. ACLR: Framework for information literacy for higher education (2016), http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/infolit/Framework_ILHE.pdf

14. Vuorikari, R., Punie, Y., Carretero, S., van den Brande, L.: DigComp 2.0. The digital competence framework for citizens. Publications Office, Luxembourg (2016)

15. Weller, K., Peters, I.: Seeding, Weeding, Fertilizing - Different Tag Gardening Activities for Folksonomy Maintenance and Enrichment, 110–117 (2012)

16. Farooq, U., Kannampallil, T.G., Song, Yang, Ganoe, Craig H., Carroll, J.M., Glies, C.L.: Evaluating Tagging Behavior in Social Bookmarking Systems: Metrics and design heuristics. Association for Computing Machinery, New York New York (2007)

17. Thomas, M., Caudle, D.M., Schmitz, C.M.: To tag or not to tag? Library Hi Tech 27, 411–434 (2009). doi:10.1108/07378830910988540

18. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: Shanahan, J.G. (ed.) Proceedings of the 17th ACM conference on Information and knowledge management, pp. 203–212. ACM, New York, NY (2008)

19. Lawson, K.G.: Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers. The Journal of Academic Librarianship 35, 574–582 (2009). doi:10.1016/j.acalib.2009.08.020

20. Gupta, M., Li, R., Yin, Z., Han, J. (eds.): An overview of social tagging and applications. Springer Science+Business Media (2011). doi:10.1007/978-1-4419-8462-3_16,

21. Zervas, P., Sampson, D., Pelliccione, L.: Studying tag vocabulary evolution of social tagging systems in learning object repositories. Smart Learn. Environ. 3, 93 (2016). doi:10.1186/s40561-016-0037-z

22. Klašnja-Milićević, A., Vesin, B., Ivanović, M.: Social tagging strategy for enhancing e-learning experience. Computers & Education 118, 166–181 (2018). doi:10.1016/j.compedu.2017.12.002

23. Allam, H., Bliemel, M., Spiteri, L., Blustein, J., Ali-Hassan, H.: Applying a multi-dimensional hedonic concept of intrinsic motivation on social tagging tools. A theoretical model and empirical validation. International Journal of Information Management 45, 211–222 (2019). doi:10.1016/j.ijinfomgt.2018.11.005