Unsupervised Auto-tagging for Learning Object Enrichment

Ernesto Diaz-Aviles, Marco Fisichella, Ricardo Kawase, Wolfgang Nejdl, and Avaré Stewart

L3S Research Center, Leibniz University Hannover, Germany
{diaz,fisichella,kawase,nejdl,stewart}@L3S.de

Abstract. An online presence is gradually becoming an essential part of every learning institute. As such, a large portion of learning material is becoming available online. Incongruently, it is still a challenge for authors and publishers to guarantee accessibility, support effective retrieval and the consumption of learning objects. One reason for this is that non-annotated learning objects pose a major problem with respect to their accessibility. Non-annotated objects not only prevent learners from finding new information; but also hinder a system's ability to recommend useful resources. To address this problem, commonly known as the cold-start problem, we automatically annotate specific learning resources using a state-of-the-art automatic tag annotation method: *α-TaggingLDA*, which is based on the Latent Dirichlet Allocation probabilistic topic model. We performed a user evaluation with 115 participants to measure the usability and effectiveness of *α-TaggingLDA* in a collaborative learning environment. The results show that automatically generated tags were preferred 35% more than the original authors' annotations. Further, they were 17.7% more relevant in terms of recall for users. The implications of these results is that automatic tagging can facilitate effective information access to relevant learning objects.

Keywords: Metadata Generation, User Study, LDA, Cold-Start, Recommender Systems.

1 Introduction

Learning strategies have shifted from a solitary activity to a collaborative web-based one [2]. In collaborative learning systems, digital collections of educational materials or Learning Objects (LOs), such as, lecture videos, notes and presentations, are made available in online repositories. Online learners are not only able to browse or search for LOs, but also enrich this content with value-added metadata.

Learning object enrichment is crucial within a collaborative setting. For example, consider a scenario in a collaborative environment where a user wants to retrieve specific documents related to their interests and uses tags to navigate to the associated resources. Ideally, if the system can effectively provide good tag coverage over the resources, the user can better navigate through document objects and be steered to the relevant resources in the system. On the contrary, if tags are either unclear, not specific for the resource, noisy, or ambiguous, then users cannot retrieve or easily locate resources. Unfortunately, the latter situation is all too common. Since users typically only

C. Delgado Kloos et al. (Eds.): EC-TEL 2011, LNCS 6964, pp. 83–96, 2011.

tag a small fraction of the documents, most of the other documents have no associated metadata. Furthermore, newly added resources, which have not yet been tagged, are hard to be located or associated with related objects. This dilemma is well known and referred to as the *new item* cold-start problem.

As outlined in the aforementioned scenario, an important prerequisite for realizing the benefit of tags in a collaborative learning setting, is that a LO actually has to have at least a minimum number of tags associated with it. When a learning resource has no associated tags, the collaborative learning system cannot provide a recommendation, nor does any descriptor exist in the tag cloud to help support navigation to the orphan resource.

One way to address the cold-start problem in collaborative systems is by using automatic tagging, which associates tags with untagged resources. State-of-the-art work in this area relies upon latent data models to make explicit, some hidden, underlying "context" (i.e., set of keywords or tags). The untagged resources are treated as a new resource, for which inferencing can be performed to bring the new resource into a known context where it can inherit an appropriate set of tags. Little has been done in the area of latent model based automatic tagging for learning objects repositories. Moreover, the usability and effectiveness of such automatic tagging has not been assessed by the learners themselves.

In this work, we propose to automatically associate tag annotations to untagged LOs by exploiting the content from different, but similar resources, found outside the boundaries of a single content repository and using a state-of-the-art method, *α-TaggingLDA*, which is based on the probabilistic topic model Latent Dirichlet Allocation [4]. In doing so, we will address the following research questions:

Q1: To what extent do the LO's annotations assigned by the authors agree with the ones assigned automatically?
Q2: From the user perspective, how relevant are the automatic generated tags comparing to the ones provided by experts?
Q3: In a social tagging recommendation scenario for LOs, are the automatic annotations better candidate terms for assisting users in the tagging process than the keywords assigned by the LO's author?

The contributions of this work are:

– an automatic tagging approach to efficiently address the cold-start problem in collaborative learning environments by relying on content from resources in an auxiliary domain, i.e., one that lies outside of the content repository of the untagged LO.
– an evaluation of our approach through a user study involving 115 participants in a online setting, with real-world data.

The rest of this paper is organized as follows. In Section 2 we will first present related work on the areas of automatic tagging and learning objects enrichment. In Section 3 we describe our approach to automatically tag learning objects. In Section 4 we present a three part evaluation to measure the effectiveness of our method and in Section 5 we describe the evaluation results, which are discussed in Section 6. Finally, in Section 7 we conclude and discuss the directions for future work.

2 Related Work

Cold-start is a common problem in many user-centric systems that seek to improve information access. Specifically for the collaborative learning environment setting, a new LO is introduced into the domain, but it has no (or incomplete) associated user-defined metadata or annotations. Automatic enhancement of LO metadata is an alternative to address this problem. In this section we present related work in the areas of automatic tagging and learning objects enrichment.

2.1 Learning Object Enrichment

Lohmann et al. [13] demonstrate the importance of additional metadata to learning resources visibility and reusability and suggest design guidelines for automatic tagging approaches. The authors suggest (i) the use of a stable set of tags for agreed description of resources, (ii) to guide the tagging process (e.g., with tag recommendations), (iii) to use text extracted from resources for starting set of tags, and (iv) the use of a small set of selectable tags for tags convergence. The two user evaluations we present in this work align with these guidelines.

Another recent system, namely, ReMashed [7,6] takes advantage of tagged and rated data of combined Web 2.0 sources, integrating the metadata from decentralized sources of content. Their work addresses the *new user* cold-start scenario and shows that a recommender system that exploits already tagged resources can mitigate the lack of user information. Our work, on the other hand, focuses on the *new item* cold-start problem, and aims to annotate untagged LOs. Once objects are tagged, it is possible to improve the performance of a recommender system [19].

Abel et al. [1] introduce the LearnWeb 2.0 environment,which supports sharing, discovering, and managing learning resources, which are spread across different Web 2.0 platforms, between learners and educators. LearnWeb 2.0 aggregates resources and enhances their metadata using functionalities from ten different Web 2.0 services. Furthermore, in order to support collaborative searching, the authors are provided with an automatic resource annotation service. Once a search result is displayed in the environment, it is automatically tagged with the corresponding query terms. This mechanism assumes that the system has enough information to retrieve the item as relevant for a given query. Furthermore, it requires an initial user interaction, i.e., search, in order to be able to annotate the resource. This is not necessarily the case for resources with sparse text, or multimedia resources with little or no metadata available. Our approach is content based and does not require any user interaction to automatically annotate the LOs.

2.2 Automatic Tagging

Numerous forms of enrichment can be used to annotate a LO such as ratings, comments and tagging, to name but a few. Tagging, in particular, has proven to be an intuitive and flexible mechanism for improving the access to information or personalizing the users' online experience. Tags are capable of facilitating search [3,10] and improving recommendations [19,18]. Tags have also been used for personalization: improving information access in collaborative tag recommendations [16] and facilitating personalized information access across disparate media types [19].

State-of-the-art methods for automatic annotation and tag recommendation rely on dimensionality reduction and are based on tensor factorization [17,16,20] or on probabilistic topic models, in particular on Latent Dirichlet Allocation (LDA), for example, Krestel et al. in [11,12] exploited resources annotated by many users, thus having a relatively stable and complete tag set, to overcome the cold-start problem. They build an LDA model from tags that have been previously assigned by users. In this way, a resource in the system is represented with tags from topics discovered by LDA. For a new resource with few or no annotations, they expand the latent topic representation with the top tags of each latent topic.

In these auto tagging systems, the performance of the aforementioned approaches highly relies on the assumption of a dense set of data upon which the model can be built. To overcome this issue, Diaz-Aviles et al. [5] introduced *α-TaggingLDA*, a method for automatic tagging resources with sparse and short textual content. In the presence of a new resource, an *ad hoc* corpus of related resources is created, and then the method applies LDA to elicit latent topics for the resource and the associated corpus. This is done in order to automatically tag resources based on the most likely tags derived from the latent topics identified. In our work this method was chosen to annotate LOs in a cold-start scenario.

In contrast to previous work, we aim to evaluate the usability and effectiveness of this automatic tagging approach, and address its potential to generate metadata for novel resources in the context of a collaborative learning environment.

3 Automatically Enhancing Learning Objects with Tags

In order to enhance the learning object with tags we use *α-TaggingLDA*, a state-of-the-art approach for automatic tagging introduced by Diaz-Aviles et al. [5]. *α-TaggingLDA* is designed to mitigate new item cold-start problems by exploiting content of resources, without relying on collaborative interactions.

The approach is based on the probabilistic topic model: Latent Dirichlet Allocation (LDA) [4], which is a generative probabilistic model for collections of discrete data such as text corpora. The basic idea of LDA is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over terms.

For example, an LDA model might have topics that can be labeled as EDUCATION and ENTERTAINMENT[1]. Furthermore, a topic has probabilities of generating various words such as *school*, *students*, and *teacher*, which can be classified and interpreted as EDUCATION. Naturally, the word *education* itself will have high probability given this topic. The ENTERTAINMENT topic likewise has high probability of generating words such as *film*, *music*, and *theater*.

[1] Please note that these labels are arbitrary. The algorithm does not automatically assign any particular label to the latent topics.

More formally, assume that a text collection consists of a set of documents D. Furthermore, consider the set of topics Z, the distribution $P(z \mid d)$ over topics $z \in Z$ in a particular document $d \in D$, and the probability distribution $P(t \mid z)$ over terms $t \in T$ given topic $z \in Z$, where T is the set of terms. Each term $t_i \in T$ in a document (where the index refers to the ith term token) is generated by first sampling a topic from the topic distribution, then choosing a term from the topic-term distribution. We write $P(z_i = j)$ as the probability that the jth topic was sampled for the ith term token and $P(t_i \mid z_i = j)$ as the probability of term t_i under topic j. The model specifies the following distribution over terms within a document:

$$P(t_i \mid d) = \sum_{z} P(t_i \mid z_i = j)P(z_i = j \mid d) \tag{1}$$

where $|Z|$ is the number of topics.

In LDA, the goal is to estimate the distribution topic-term $P(t \mid z)$ and the document-topic distribution $P(z \mid d)$. These distributions are sampled from Dirichlet distributions (e.g., using Gibbs sampling [8]) and indicate which terms are important for which topic and which topics are important for a particular document, respectively.

α-TaggingLDA

An overview of the *α-TaggingLDA* method is shown in Figure 1. In order to illustrate the method with an example, consider a novel LO entitled *Knowledge Technologies in Context*, this resource is new to the collaborative learning system and does not have any tag annotations assigned. The absence of tags makes it difficult for the system to consider it as candidate for recommendations, for instance.

α-TaggingLDA first extracts relevant LO's *textual content*, such as the title, description or metadata (e.g., author) and creates a document denoted as d_{LO}. Then, the LO is associated to a set of 'similar' documents, which we refer to as an *ad hoc corpus* for the LO, represented as $corpus_{LO}$.

Note that the *α-TaggingLDA* method does not impose any restriction on the similarity measure used to associate the corpus with the LO. The similarity measure could be specified based on the nature of the resources, (e.g., text documents, multimedia items) and the textual content or metadata available. For example, a particular implementation might rely upon a computationally inexpensive similarity measure or on a more complex clustering algorithm.

In our particular example, the title of the LO is used to query an Internet search engine in order to retrieve the title and snippets of the n relevant results ($n = 4$, in this case). This subset corresponds to $corpus_{LO}$.

The LO's textual content is extracted and the subset of the top n results constitute the text collection $D = \{d_{LO}\} \cup corpus_{LO}$, which is input to LDA, together with the number of topics required. In this example, the number of topics is set to two, i.e., $|Z| = 2$. The set of tags to be used to annotate the LO is denoted as $TopN_{tags}(LO)$, and its size is set to six for this particular case, i.e., $|TopN_{tags}(LO)| = 6$.

Table 1 presents an example of the output produced by LDA according to the setting described above. Topics are ordered based on the document-topic distribution $P(z \mid d)$, and within each topic, terms are ranked based on the topic-term $P(t \mid z)$ distribution.

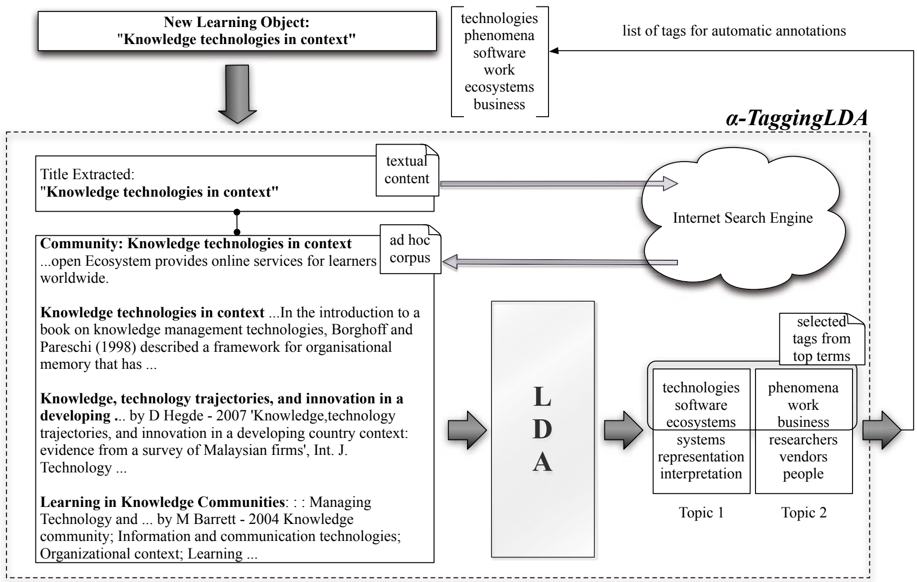

Fig. 1. *α-TaggingLDA* is applied to annotate a new LO, *Knowledge Technologies in Context*, with a list of six tags: $TopN_{tags}(LO)$ = { *technologies, phenomena, software, work, ecosystems, business* }, based on two LDA topics

For the construction of the final set of tags $TopN_{tags}(LO)$, *α-TaggingLDA* selects the first candidate tag from $Topic_1$'s top terms, the second tag from $Topic_2$'s top terms, the third tag, again from $Topic_1$'s top terms, and so forth. The final list of tag annotations for the LO in our example corresponds to $TopN_{tags}(LO)$ = { *technologies, phenomena, software, work, ecosystems, business* }. For the details of this strategy, we refer the reader to the work done by Diaz, et.al. [5].

Table 1. Example of two topics output by LDA. Topics are ordered based on the document-topic distribution $P(z \mid d)$, and within each topic, terms are ranked based on the topic-term $P(t \mid z)$ distribution.

$Topic_1$		$Topic_2$	
$P(z = 1 \mid d_{LO}) = 0.70$		$P(z = 2 \mid d_{LO}) = 0.30$	
Term t	$P(t \mid z = 1)$	Term t	$P(t \mid z = 2)$
technologies	0.45	phenomena	0.33
software	0.25	work	0.28
ecosystems	0.16	business	0.19
systems	0.11	researchers	0.15
representation	0.03	vendors	0.04
interpretation	0.01	people	0.01

4 Evaluation

In this section we measure the benefits of automatic tag annotations for a collaborative learning environment. To answer the research questions presented in Section 1, we conducted three distinct evaluations, first, an experimental evaluation followed by two user studies. The rest of the section describes each evaluation settings.

4.1 Experimental Setting

Dataset. We based our experiments on a dataset sampled from the OpenScout project collection [15]. The project gathers metadata information from learning resources located at different learning content repositories. For our evaluation, we selected learning objects whose language is English and have at least five keywords added by their author. In total, 563 learning objects, 1692 unique keywords and 3150 keywords assignment were considered for the experiments.

Metadata Enrichment. The method used for automatic metadata enrichment in our experiments, *α-TaggingLDA*, is implemented in Java. The corpus builder is based on the search results obtained by querying Yahoo!'s open search web services platform (BOSS)[2]. The titles and short text summaries (snippets) of the ten most relevant results returned are used to create ten different textual documents. The final *ad hoc* corpus for the learning object consists of these and the textual content of the resource. Then, by applying LDA on this corpus we extract the desired number of latent topics, and from them, the needed tags are inferred. The default number of topics considered was two, according to the optimal setting specified in [5].

We use the LDA with Gibbs sampling implementation provided by the Machine Learning for Language Toolkit (MALLET) [14].

4.2 Evaluation I: Author's Keywords and Automatic Tags

Our first evaluation consists of an offline study that measures the agreement between the keywords assigned to the LOs by its author and the tag annotations provided by the method. In order to quantify such agreement, we consider a recommender system setting, where the author's keyword assignments constitute our test set. The task of the collaborative learning environment is to recommend $TopN_{tags}$ relevant tags for a given LO.

We use recall, precision and f1, three widely used metrics in recommender systems [9], to assess the performance. The metrics are defined in Equations 2 and 3.

– Recall for a given author u and a learning object i is defined as:

$$recall(u,i) = \frac{|Keywords(u,i) \cap TopN_{tags}(i)|}{|Keywords(u,i)|} \quad , \; and \tag{2}$$

[2] http://developer.yahoo.com/search/boss/

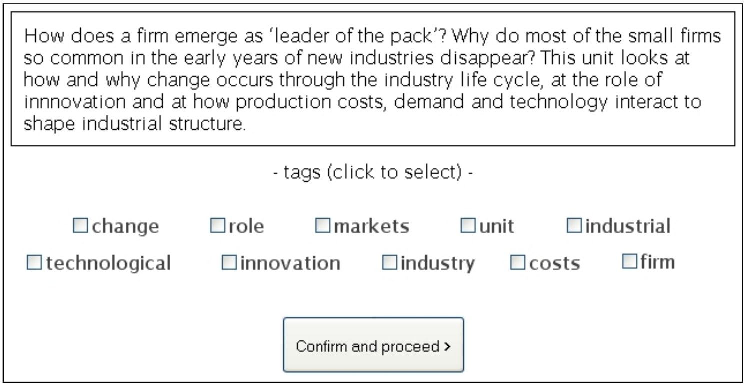

Fig. 2. Evaluation II: Guided Choice User Study Interface. Each participant was instructed to choose at least three tags from the set of ten suggested tags. Five tags were originally added by the expert/author of the content, while the remaining five were automatically generated. The tags were presented in a random order and their origin was not disclosed to the participants.

– Precision for a given author u and a learning object i is defined as:

$$precision(u, i) = \frac{|Keywords(u, i) \cap TopN_{tags}(i)|}{|TopN_{tags}(i)|} \quad , \tag{3}$$

where $Keywords(u, i)$ is the set of keywords assigned by the author u to the learning object i and $TopN_{tags}(i)$ is the set of size N corresponding to the tags automatically assigned by *α-TaggingLDA* to the given LO. In this experiment we set $N = 10$.

For the dataset, we averaged these values over all the authors. The aggregated values of recall and precision are then used to compute their harmonic mean or f1 measure as defined according to Equation 4.

$$f1 = 2 \cdot \frac{recall \cdot precision}{recall + precision} \tag{4}$$

4.3 Evaluation II: Guided Choice User Study

The goal of this experiment was to compare the automatically generated tags against the ones provided by experts.

This evaluation is a user study in which each participant was presented with basic information regarding a learning object, namely, the title and an abstract that varies from 20 up to 200 words (see Figure 2). The format of the original resource (e.g. video, image, presentation or document) was not made known to the participant in order to align the nature of the evaluation and to avoid biased judgments of the tag relevance based on non computer-understandable information. In addition to that, the participants were presented with ten tags to be evaluated. From the ten tags presented, five tags were originally added by the expert/author of the content, while the remaining five were the top ranked automatically generated ones. The tags were presented in a random order and their origin was not disclosed to the participants.

Each participant was then instructed to read the title and the description of the learning object and finally choose at least three tags from the set of ten suggested tags. Once the submission of the form is completed the participant was presented with a new object to be evaluated. We kindly asked for each participant to repeat the process for at least ten objects, however, we did not limited the maximum of their contribution to the study.

In order to compare the automatically generated tags against the ones provided by experts, we designed this experiment as a recommendation task, and used the recall measure, which is widely used to assess the recommendation quality [9] of recommender systems.

In this case recall is defined in Equation 5, as follows:

– Recall for a given author u and a learning object i is defined as:

$$recall(u,i) = \frac{|Tags(u,i) \cap TopN_{tags}(i)|}{|Tags(u,i)|} \quad , \tag{5}$$

where $Tags(u,i)$ is the set of tags assigned by the user u to the learning object i and $TopN_{tags}(i)$ is the set of size N corresponding to the tags recommended to the user, either based on *α-TaggingLDA* or on the author's keywords. In this experiment we set $N = 5$.

As in Evaluation I, we averaged the values over all the participants. Using a fixed number of recommendations, precision is just the same as recall up to a multiplicative constant and thereby there is no need to evaluate precision.

4.4 Evaluation III: Free Choice User Study

The goal of this study was to collect evidence to evaluate if the automatic annotations are better candidate terms for assisting users in the tagging process than the keywords assigned by the LO's author.

Similarly to Evaluation II, in this user study, each participant was presented with the title and an abstract of a learning object. Once again, due to same reasons as presented before, the format of the original resource (e.g. video, image, presentation or document) was not disclosed to the participants.

Each participant was then instructed to read the title and the description of the learning object and finally input five tags she thinks to be relevant for describing the object, as depicted in Figure 3. Once the submission of the form was completed, the participant was presented with a new object to be evaluated. Each participant was asked to repeat the process for at least ten objects.

As in Evaluation II, in other to evaluate this experiment we cast it as a recommendation task and evaluate the recall measure. In this case $Tags(u,i)$ corresponds to the set of tags that would be recommended to the participant u for the given LO i. Note that, even though, the set of tags is not presented to the participant, it helps us to measure which terms are better for assisting users in the tagging process.

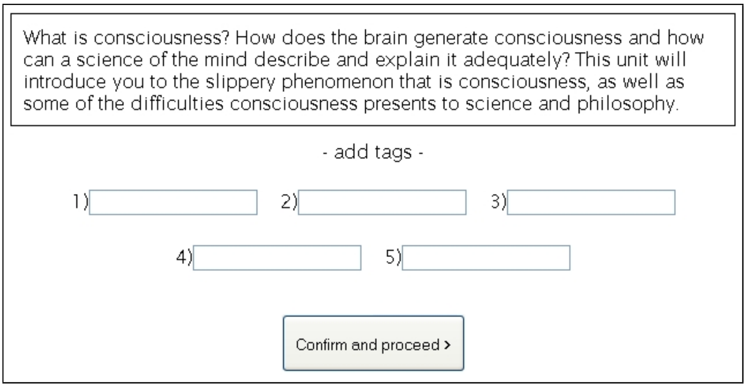

Fig. 3. Evaluation III: Free Choice User Study Interface. Participants were instructed input five tags they think to be relevant for describing the LO.

5 Results

Evaluation I aims to answer the research question *'Q1: At what extent do the LO's annotations assigned by the authors agree with the ones assigned automatically?'*. As an outcome for $|TopN_{tags}| = 10$, we obtained the following results, recall=0.26 precision=0.13 and f1=0.18. Table 2 shows the f1 measure for different sizes of $TopN_{tags}$.

Table 2. f1 measure for different sizes of $TopN_{tags}$

$\boldsymbol{TopN_{tags}}$	1	2	3	4	5	6	7	8	9	10
f1	0.04	0.07	0.09	0.10	0.10	0.11	0.12	0.13	0.16	0.18

From the user study in Evaluation II we collected the feedback of 115 participants (43 female and 72 male), 100 of them explicitly stated to be students. Their average age was 24, ranging from 20 to 53 years old. In total the participants evaluated 1,134 objects covering 478 unique ones.

Also, in total 4,035 tags were chosen to represent the documents, in average each participant picked 3.56 tags per document. As explained in the setup of this evaluation, the tags exposed to the participants were originated from two different sources, the expert who created the learning material and a second set from the automatic tagging method.

It's important to note that the tags from each group were always presented to the participants in an equal distribution to preserve the fairness of the study. Additionally, when a participant chose a tag that was in both *α-TaggingLDA* set and the Experts' set we computed this choice as two tag assignments as outlined in Section 3. Although the participants chose 4,035 tags, in our experiments, we used a total of 4,939 tag assignments.

Out of the 4,939 tag assignments, 67.5% of them were originated by *α-TaggingLDA* and 32.5% by the experts.

The most straightforward analysis of these results shows a clear preference of the participants for the tags that were automatically added. Thus, it is also reasonable to conclude that these tags are more relevant to the participants, which answers our second question: '*Q2: From the user perspective, how relevant are the automatic generated tags comparing to the ones provided by experts*?'. The main reason is that the underneath approach generates tags that represents better the learners tagging behavior. Through the outcomes of this evaluation we interpret that the automatic generated tags are, in general, more descriptive and more useful for the learners than experts' tags. Additionally, it is reasonable to assume that these learners, when searching for one of these documents would (with a higher probability) use a tag that was automatically generated rather than the experts' tags. The same assumption is valid for the case of browsing resources in a hierarchical classification or in a facet browsing interface.

To validate the significance of the results achieved, for each participant, we took the averages of the distribution of *α-TaggingLDA* and Expert's tag sets. With two groups of 115 samples we performed a two-tailed t-test that confirmed a statistically significant difference of *α-TaggingLDA* mean (68.2%) and Expert's keywords mean (31.8%).

For Evaluation III, the participants evaluated 832 objects covering 454 unique one. In this phase, where the participants were instructructed to freely choose terms that best classify the objects, 4,745 tags were generated (1,868 unique tags).

Using these data we now have three different sets of tags: *α-TaggingLDA* tags, experts' tags and learners' tags. By validating the learners' generated tags against the other sets we found an overlap of 38.4% with automatic generated *α-TaggingLDA* tags and 20.7% with the experts' tags. Additionally, in only 8.9% of the cases, a tag occurs in all three sets. At this point we are just considering the whole sets of tags and not the precision of them regarding each resource. These results complement Evaluation II by firmly stating that on average the automatic generated tags are closer to the ones used by learners.

By considering only the results from those 100 participants that stated to be students, the numbers do not change significantly. The overlap with automatic generated tags increases slightly to 39.04% while the overlap with the experts' tags remains on the same levels (20.3%). These results help us to answer the third question – Q3, as the automatic annotations turned out to be the best candidate terms for assisting users in the tagging process.

Table 3. Tag Assignment (TAS) results for the user evaluations. The *Sets* rows show the number of TAS that were chosen by the participants that overlapped with TAS given by the experts, or with *α-TaggingLDA* method and the respective recall measure.

		Experiment II: Guided Choice		**Experiment III: Free Choice**	
	Participant's TAS	4939	-	4745	-
Sets	*α-TaggingLDA*	3336	67.5%	1824	38.4%
	Experts	1603	32.5%	983	20.7%

6 Discussion

The values of recall amd precision of Evaluation I (Section 5) suggest that the information captured by the automatic tag annotations partially agree with the expert keywords assigned to the LO. The values are not exceptionally high, which suggest that the automatic tag annotations tend to capture different information than the expert keyword assignments. The user studies conducted in Evaluation II and III exposed how additional information captured by the automatic annotations are perceived by the learners, and explore the usability improvements of the collaborative learning system. The results from the first user study setup (Evaluation II) clearly demonstrate the preference of the participants for tags produced by the automatic tagging method. This means that, the produced tags reflects better the participants' preferences in comparison to the experts' keyword assignments. The most probable reasons are, first, the aforementioned problem that a tag assignment is not always clear to users other than its creator. Second, learners usually have a viewpoint that differs from the experts, thus they are more prone to avoid terms that are too specific or that they would probably not remind later. Finally, the terms given by the automatic tagging, extracted from search results' snippets, represent better the wisdom of the crowd since these results are originally extracted from multiple resources. It is also important to remark that the search results themselves are consequence of ranking algorithms that exploit collective knowledge and preferences.

In principle, the results of the second user study (Evaluation III) support the same benefits. The goal of this phase was to prevent any possible biases in the first evaluation. We hypothesize that, when asking a participant to tag a learning object, we are implicitly observing which tags the participants would use in a collaborative social learning environment, and indirectly potential terms to query or browse for a learning object.

Bearing in mind the overall results obtained in the experiments, the most important consideration to highlight is the potential benefits produced by the information delivered by the automatic tagging method evaluated.

7 Conclusion

We have empirically demonstrated through a series of evaluations that the proposed *α-TaggingLDA* method produces quality metadata enhancement for the learning objects. First, by experimentally comparing against existing authors' tag annotations. Second, by running a user study comparing the participants' preference for automatically produced tags against the authors' tags. Finally, a last user study that demonstrated that *α-TaggingLDA* tags are the best candidate terms for assisting users in the tagging process.

The additional metadata that was automatically generated by our method can improve personal recommendations of learning objects and most notably it overcomes the 'cold-start' for objects that are not tagged, consequently isolated from the rest of the folksonomy.

Our approach faces some limitations, since we depend on the external resources provided by a search engine. One implication of this shortcoming is that the collection of the documents retrieved may not contain enough meaningful text for good topics to be generated. Another implication is that in some cases, there may be valuable documents for enriching the learning resources, but the documents may not be available if they are buried in the "Hidden-Web", i.e. documents that are not indexed by search engines. One potential solution to, at least, mitigate this limitation is to use multiple and heterogeneous sources for building the topic model. Heterogeneity would include the use of multiple search engines, and open information sources such as wikipedia. Future experiments are needed to examine heterogeneous sources, and we consider this in future work.

Additionally we plan to evaluate how this method could be further refined to assist authors in tasks of keyword assignment by recommending them relevant terms for the LO. Additionally we are interested in explore how automatically added tags can be incorporated in user's profiles and to what extent it can improve recommendations and discovery of new items. Although our work focuses on approaching the cold start problem, we are also interested in running an evaluation with learning objects that have already been enriched by an active community. This would provide us valuable insights to compare the automatic generated tags based on the general wisdom of the crowd and the focused learning community.

Acknowledgement. The authors would like to thank Katrina Maxwell at INSEAD for her support on the user evaluation conducted. This research has been co-funded by the European Commission within the eContentplus targeted project OpenScout, grant ECP 2008 EDU 428016 (cf. http://www.openscout.net).

References

1. Abel, F., Marenzi, I., Nejdl, W., Zerr, S.: Sharing distributed resources in learnWeb2.0. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 154–159. Springer, Heidelberg (2009)
2. Atkins, D.E., Seely, B.J., Allen, H.: A review of the open educational resources (oer) movement: Achievements, challenges and New Opportunities. Review Literature And Arts Of The Americas, 84 (2007)
3. Bischoff, K., Firan, C.S., Nejdl, W., Paiu, R.: Can all tags be used for search? In: CIKM 2008: Proceeding of the 17th ACM conference on Information and knowledge management, pp. 193–202. ACM Press, New York (2008)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. Journal of Machine Learning Research 3, 993–1022 (2003)
5. Diaz-Aviles, E., Georgescu, M., Stewart, A., Nejdl, W.: Lda for on-the-fly auto tagging. In: Proceedings of the fourth ACM conference on Recommender systems, RecSys2010, pp. 309–312. ACM Press, New York (2010)
6. Drachsler, H., Pecceu, D., Arts, T., Hutten, E., Rutledge, L., van Rosmalen, P., Hummel, H., Koper, R.: ReMashed – recommendations for mash-up personal learning environments. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 788–793. Springer, Heidelberg (2009)
7. Drachsler, H., Rutledge, L., van Rosmalen, P., Hummel, H.G.K., Pecceu, D., Arts, T., Hutten, E., Koper, R.: Remashed - an usability study of a recommender system for mash-ups for learning. iJET 5(S1), 7–11 (2010)

8. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proc. Natl. Acad. Sci. U S A 101(suppl. 1), 5228–5235 (2004)
9. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. 22, 5–53 (2004)
10. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: Search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006)
11. Krestel, R., Fankhauser, P.: Language models and topic models for personalizing tag recommendation. In: 2010 IEEE WIC ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 82–89 (2010)
12. Krestel, R., Fankhauser, P., Nejdl, W.: Latent dirichlet allocation for tag recommendation. In: Proceedings of the third ACM conference on Recommender systems, RecSys 2009, pp. 61–68. ACM Press, New York (2009)
13. Lohmann, S., Thalmann, S., Harrer, A., Maier, R.: Learner-generated annotation of learning resources - lessons from experiments on tagging. In: International Conference on Knowledge Management (I-KNOW 2008), Graz, Austria (September 2008)
14. McCallum, A.K.: Mallet: A machine learning for language toolkit (2002), `http://mallet.cs.umass.edu`
15. Niemann, K., Schwertel, U., Kalz, M., Mikroyannidis, A., Fisichella, M., Friedrich, M., Dicerto, M., Ha, K.-H., Holtkamp, P., Kawase, R., Parodi, E., Pawlowski, J., Pirkkalainen, H., Pitsilis, V., Vidalis, A., Wolpers, M., Zimmermann, V.: Skill-based scouting of open management content. In: Wolpers, M., Kirschner, P.A., Scheffel, M., Lindstaedt, S., Dimitrova, V. (eds.) EC-TEL 2010. LNCS, vol. 6383, pp. 632–637. Springer, Heidelberg (2010)
16. Rendle, S., Balby Marinho, L., Nanopoulos, A., Lars, S.-T.: Learning optimal ranking with tensor factorization for tag recommendation. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD 2009, pp. 727–736. ACM Press, New York (2009)
17. Rendle, S., Lars, S.-T.: Pairwise interaction tensor factorization for personalized tag recommendation. In: Proceedings of the third ACM international conference on Web search and data mining, WSDM 2010, pp. 81–90. ACM Press, New York (2010)
18. Sen, S., Vig, J., Riedl, J.: Tagommenders: Connecting users to items through tags. In: International World Wide Web Conference, Madrid, Spain, April 20. ACM Press, New York (2009)
19. Stewart, A., Diaz-Aviles, E., Nejdl, W., Marinho, L.B., Nanopoulos, A., Schmidt-Thieme, L.: Cross-tagging for personalized open social networking. In: Cattuto, C., Ruffo, G., Menczer, F. (eds.) Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (Hypertext 2009), Torino, Italy, pp. 271–278. ACM Press, New York (2009)
20. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: Tag recommendations based on tensor dimensionality reduction. In: Proceedings of the 2008 ACM conference on Recommender systems, RecSys 2008, pp. 43–50. ACM Press, New York (2008)