Raising up Annotations In Pedagogical Resources by Human-Computer Collaboration

Armelle Brun, Hala Skaf-Molli, Anne Boyer

RAISING UP ANNOTATIONS IN PEDAGOGICAL RESOURCES BY HUMAN-COMPUTER COLLABORATION

Armelle Brun, Hala Skaf-Molli, Anne Boyer

LORIA-Nancy Université, 615, avenue du jardin botanique – 54506 Vandoeuvre les Nancy

{armelle.brun,hala.skaf,anne.boyer}@loria.fr

Abstract

Huge repositories of pedagogical resources such as the French initiative of digital thematic libraries are now accessible by both students and professors. This high quantity of data makes their access difficult to the students, as they cannot get easily pertinent pedagogical resources that fit their needs. One way to make this access easier is to add annotations to these resources and to exploit these annotations to find pertinent answers. Of course, these annotations can be semantic. Semantic wikis are a new approach that automatically processes semantic annotations and that can be used to find the adequate resources given the requests of students. However, annotating semantically resources is not an easy task for human. Despite of their high potential, semantic wikis suffer from a lack of human provided semantic annotations, resulting in a loss of their efficiency. We propose a system (called HCA) that suggests automatically computed annotations to users in semantic wikis. In this paper users are students and professors. Users only have to validate, complete, modify, refuse or ignore these suggested annotations. Therefore, the annotation task becomes easier, and we assume that more users will provide annotations, leading to an improvement of the system and a facilitated access to pertinent pedagogical resources. The HCA system is based on collaborative filtering recommender systems, it does not exploit the content of the pages but the usage made on these pages by the students and the professors. The resulting semantic wikis contain several kinds of annotations, with different status: human, computer or human-computed provided annotations.

Introduction

Web 2.0 encourages the production of new pedagogical resources and the collaboration of students and professors to increase the quality of available resources by using for instance, wikis and blogs. This results in a huge number of resources broadly available on the Internet via dedicated repositories, such as the French initiative of digital thematic libraries [15]. However, these facilities lead to a too huge number of pedagogical resources accessible. Therefore students, and professors as well, may be lost in this sea of resources, they cannot easily find the resources they search.

Of course, these pedagogical resources are often annotated or tagged by the professors who produce them. However, these annotations are often different from those the students or other professors use in their requests. Students and professors may never find pertinent resources. Moreover, these annotations are put *a priori* and may not reflect the real usage of the resources. The objective of this paper is to learn about the real usage made on these resources to provide *a posteriori* annotations that reflect better the reality of usage of these resources, compared to the *a priori* annotations. As a result, the use of these annotations makes the access to pertinent resources easier.

Semantic wikis are one of the most promising approaches to automatically process information. Compared to classical wikis, they allow users to add *semantic annotations* in the wiki pages. In this paper the wiki pages will be pedagogical resources such as lessons or exercises. When using semantic wikis, users do not only collaborate for writing the content of the wiki pages but also for writing the semantic annotations that will allow a better and easier usage of these pages. It will then be possible to answer user queries by exploiting semantic annotations. However, adding semantic annotations is not an easy task and is time consuming. In the frame of repositories of pedagogical resources, the consequence is that only few users annotate pages semantically. As it is not easy to motivate students or professors to provide with annotations, many existing semantic wikis contain only few

annotations; their potential added value is thus decreased. One possible solution is to propose a system that suggests pertinent annotations to users (students and professors).

This paper proposes a new kind of collaboration between users and machine to produce pertinent, useful and reliable annotations in semantic wikis. The production of these annotations is based on a recommender system. Many works about recommender systems [5,1,9] and semantic wikis [6,3] can be found in the literature. However, few existing works [4] propose to use recommender systems in the frame of semantic wikis. This paper proposes to suggest new annotations to users by mining the automatically collected observations about the real usages of wiki pages. These annotations represent better the real interest of users for the pedagogical resources and will provide additional information about them. These "computer recommended" annotations are suggested to students and professors who can validate (and/or complete), modify, refuse or ignore them. Based on these suggested annotations, the task of annotating is made easier as they do not have to create the annotations from the scratches, they only have to decide about proposed annotations.

The first section gives a brief overview of the proposed system. The following section summarizes the required background in term of semantic wikis and recommender systems. Then we describe how we use recommender systems to automatically determine possible additional semantic annotations. It presents also how the users and machines collaborate to provide reliable semantic annotations. The last section concludes and points out the perspectives of this work.

Overview of the System

The efficiency of semantic wikis relies among others on the quantity and the quality of the available semantic annotations. As semantic wikis often face a lack of user provided annotations, the question that we raise in this paper is: how can we encourage and support users of the repository of pedagogical resources to provide semantic annotations?

This paper addresses this question by introducing the machine as a partner of the collaborative process of annotation in semantic wikis. The community of users becomes a community gathering students, professors and computers working together to produce semantic annotations. Students, professors and computers do not have equivalent roles within the community: computers can only suggest new semantic annotations but they do not have any influence on the final decision to keep, modify or discard them. In contrast, students and professors can add new semantic annotations and participate to the final decision of integrating or not a given annotation provided by computers. The tool we propose can be viewed as a semi-automatic annotation tool. Concretely, a specific recommender suggests possible annotations to students and professors. Each of them can decide to ignore, accept (and/or complete) or refuse them. If agreed, the status of the annotation will change to become a Human Computer Annotation. If refused, the annotation will be discarded but memorized by the system as Refused Annotation to avoid to suggest it again. The validated annotations will be added to the usual annotations resulting from human collaboration. An annotation can thus have four different status:

1. Human Annotation (HA): this annotation results from human collaboration;
2. Computer Annotation (CA): this annotation is suggested by the recommender but not reviewed by humans;
3. Human Computer Annotation (HCA): this is a computer annotation that has been validated (and/or completed) by the users;
4. Refused Annotation (RA): this is a computer annotation that has been discarded by humans.

Both HA and HCA annotations are used to support navigation and answer semantic queries, the CA annotations can only be used to support navigation when no other annotation is available.

This work uses collaborative recommendation based on usage mining to compute annotations. The recommender exploits the usages by the students and professors of the wiki pages, here pedagogical resources, to determine pertinent suggestions of annotations. It exploits implicit collaboration between users of the wiki: we consider that students or professors who share usages collaborate implicitly. This approach does not take into account the content of the pedagogical resources to perform recommendations, only their usage is considered. The system integrates three levels of collaborations.

- Level 1: Explicit human collaboration when users write annotations, HA.
- Level 2: Implicit human collaboration when the system suggests CA.
- Level 3: Human-computer collaboration when users validate and complete CA that become HCA.

Suppose we apply Human-computer collaboration to a repository about "Computer Science". Let two pedagogical resources be "C programming" and "Principles of compilation", these two resources are not directly linked by semantic annotations. The "C programming" page may contain the following semantic annotation: "DedicatedTo" as "Beginners". However, "C programming" and "Principles of compilation" may be often altogether as students who want to write a C program may be interested in knowing how programs are compiled. Therefore, the semantic wiki pages of these two pedagogical resources should be semantically linked. As many students interested in "C programming" often also consult the "Principles of compilation", our HCA system will automatically discover this link, based on the usage made by students and it this link will be recommended to users as a CA annotation. The semantic annotations will be enriched and the usability of the semantic wiki will be improved, students will access easier pertinent resources.

Background

This section presents backgrounds about semantic wikis and recommender systems.

Semantic wikis

Semantic Wikis are an extension of wiki systems, they embed semantic annotations in the wiki content. These annotations allow to better organize and structure the wiki contents. They guide the users from informal knowledge contained in texts to more formal structures. Many semantic wikis are being developed such as Semantic MediaWiki (SMW) [6], SweetWiki [3] and Swooki [12]. In SMW, links between wiki pages are typed. For instance, a link between the wiki pages "France" and "Paris" may be annotated by a user as "has Capital". Annotations express semantic relationships between wiki pages. They are usually written in a formal syntax so they are processed automatically by machines and are exploited by semantic queries. In semantic wikis, semantic annotations are added by users, they can be viewed as Human Annotations, they correspond to the Level 1 of collaboration as presented in the previous section. Semantic wikis, as classical wikis, suffer from scalability, availability and performance problems and they do not support offline works. To overcome these limitations, peer to peer extensions [12] [13] for semantic wiki have been proposed . SWooki [12] is a peer to peer (P2P) semantic wiki that follows the same annotation principles as SWM. It is a P2P network of autonomous semantic wiki servers, every server hosts a copy of all semantic wiki pages and the semantic data. Every peer can autonomously offer all the services of a semantic wiki server. When a peer updates its local copy of data, it generates the corresponding operation. This operation 1) is executed immediately against the local replica of the peer, 2) it is broadcasted to all other peers, 3) it is received by the other peers, 4) and it is integrated to their local replica. If needed, the integration process merges this modification with concurrent ones, generated either locally or received from a remote server.

Recommender systems

Recommender Systems provide personalization to users to cope with the well-known problem of overload of information [1]. Among the possible approaches in recommender systems are content-based [9] and collaborative filtering approaches [5]. The first approach uses the content of the resources to compute recommendations for users, leading to accurate recommendations. However, the content of all types of resources cannot be automatically analyzed (videos, audio, etc.), thus this analysis often requires human interventions. Moreover, only resources directly linked to those the user has consulted can be suggested: no "novelty" can be recommended to users, users may thus be frustrated. Collaborative filtering (CF) approaches do not take into account the content of the resources. They only consider the usage of these resources to compute recommendations. The usage can be the consultation made by users, the votes given by users, etc. A CF-based recommender system exploits the traces of usage to deduce information about the resources. CF-based recommenders can either compute similarities between resources [11] or exploit data mining techniques to learn relationships between the resources [13]. As in content-based recommender systems, given a user, his previously consulted resources are used and

are linked/compared to all possible resources. The comparison is no more made in terms of content but on the similarities or relationships computed between resources, based on their usage. This approach allows to recommend "original" resources: resources that are not semantically linked to the past resources consulted by the user (but that are similar in terms of usage) can be recommended.

Recommender Systems for Annotation Suggestion

Existing recommender systems for Semantic Wikis directly transpose recommenders to suggest wiki pages to users as in [4]. In our work, we go a step further by suggesting annotations to wiki pages based on usage traces. We use CF-based recommender systems to provide automatically pages with additional annotations. We exploit the usage of wiki pages: which users have consulted which wiki pages and which page(s) is(are) frequently consulted after a given page? To deduce the links/relationships between pages, by using approaches similar to those presented in [11,1]. Given the relationships between pages of pedagogical resources and the annotations given by users (HA), the recommender system will suggest additional annotations. These annotations can be made on pages that either already have HA or not. Suggested annotations correspond to the implicit human collaboration level (level 2).

Suggestion of Semantic Computer Annotations

We propose two algorithms to compute annotations to suggest to students about a given page *Pj*. The first one is similar to those used in item-based approaches and classification of pages based approach [8], the second one is based on data mining techniques for recommendations [7].

Item-based approach. This algorithm first computes a similarity matrix of wiki pages [11]. This matrix is computed based on the traces of usage of the wiki pages. This approach is based on the hypothesis that two similar pages may have similar semantic annotations. Thus, given two similar pages or pedagogical resources, the recommender suggests the semantic annotations of the first one to the second one and suggests the semantic annotations of the second one to the first one. These suggested annotations are called CA. Figure 1 presents how the recommender computes candidate annotations. Given a page *Pj*, the recommender searches the pages that are similar to *Pj*, those with a high similarity value in the similarity matrix. The recommender collects the HA from the pages similar to *Pj*. The suggested CA to *Pj* can be computed by several policies:

1. The set of HA from the similar pages are suggested. The recommender may propose too many annotations, and some of them may be not pertinent.

2. The set of HA present in at least *n* similar pages are suggested. An appropriate value of *n* has to be fixed. This policy proposes a lower number of annotations and more "reliable" as they are present in several pages. However, according to the value of *n*, some pages may have no annotation.

3. Given a HA, the similarity values of the pages containing that HA are summed up. The HA with a sum of similarity values above a given threshold are suggested. This policy is more accurate than the second one as the similarity values are considered.

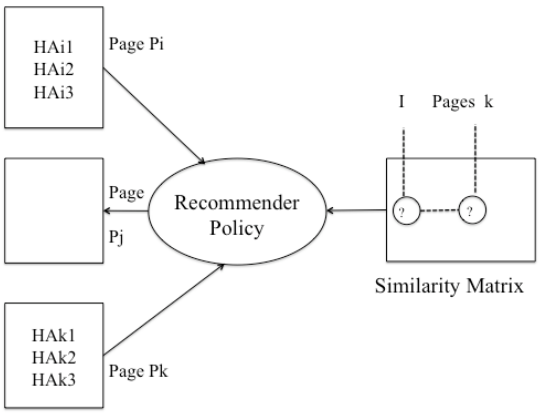

Figure 1: Annotation suggestion of HCA with the item-based approach

Classification-based approach. In this approach, given a set of pages, the recommender first computes a classification of these pages to create classes of "similar" pages (see Figure 2). As in the item-based approach, the classification is based on the usage of the pages by the users, not on the content of the pages. Then, given a page *Pj*, the recommender searches its corresponding class, and exploits all the pages in this class. All the policies presented in the item-based approach can be used to compute annotations.

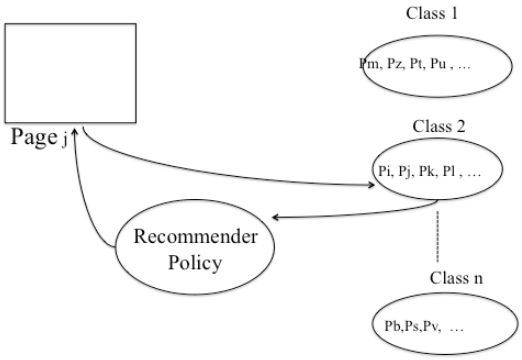

Figure 2: Automatic suggestion of HCA with the classification-based approach

The set of pages used to compute annotations is predefined (the pages in the class) whereas it is dynamically computed in the item-based approach, that can take time.

Data Mining based approach. It exploits data mining techniques to extract information about the usage of the resources. They study the sequences of consultation of resources by using, for example, association rules or Markov models to discover frequent patterns [2]. The algorithm is based on the hypothesis: if two or more pages are frequently consulted in sequence, then the links used to traverse these pages are useful and it is important to annotate semantically these links. The recommender suggests CA to the users to annotate semantically the frequently passed links. The provided annotations are not semantic annotations, they are just annotations, as the type of the annotation cannot be automatically discovered by the system. The recommender suggests that an annotation at a given place should be important. It can also specify the label of the annotation. For instance, suppose in the example given in the previous section, there is no semantic annotation between the pages "C programming" and " Principles of compilation", the recommender can learn that this link is highly passed, thus annotating semantically this link may be useful. The recommender suggests to the users the link "C programming". The user can accept, modify, refuse or ignore this suggested annotation and type it with [ApplicationDomain], for example. The annotation task is made easier as the system suggests to the user where useful annotations should be, the user is guided.

The HCA approach

As presented in the previous sections, each semantic wiki page, i.e. pedagogical resource, has three sources of annotations: first, those added by humans (HA), they correspond to the explicit human collaboration, in our framework these annotations are proposed by students and professors; second, those suggested by the recommender system (CA), they come from implicit human/student interaction. These annotations are a new kind of annotations, they do not exist in classical semantic wikis, they will be used to encourage the users to annotate semantically pages. Third, those reviewed by the students, they result from the Human computer collaboration (HCA).

To integrate these annotations to a peer to peer semantic wiki, several questions are raised: how the recommender suggests annotations (CA) to students? how to make the students validate (and/or complete), modify, refuse or ignore a CA? How to make the other students know these annotations? The way we propose to answer these questions is presented in Figure 3.

When new annotations for pedagogical resources are suggested by the recommender, they are broadcasted to all the local replica with the status Computer Annotation (CA). They are then proposed to the students in a pop-up box, for example, to be differentiated from Human Annotations HA.

A student (a user) can choose among four possible operations : 1) Ignore the CA annotation. 2) Accept it, 3) modify (and) type it or 4) refuse it. If he does nothing, the annotations remain with a CA status. If the annotations are typed (from item-based or classification-based approaches), he can either validate, modify or refuse each of them. If the annotations have no type (from data mining approach), he can complete (and thus validate), or modify each of them. When a validation, completion or modification action is made, the local replica of the user is modified and the CA is updated to a HCA. The information about this action is broadcasted to the other local replica and the annotation is suggested to users as a HCA. Let us notice that when a user does not agree on a received HCA, he can modify it as in the case of classical HA in semantic wikis. If the user refuses the annotation, it is discarded from the set of CA and is included in the set of RA and is also broadcasted to other users. This set is used by the recommender to avoid re-suggesting annotations that have already been refused by users. These RA can however be re-suggested in the case the content of the wiki page has been highly modified.

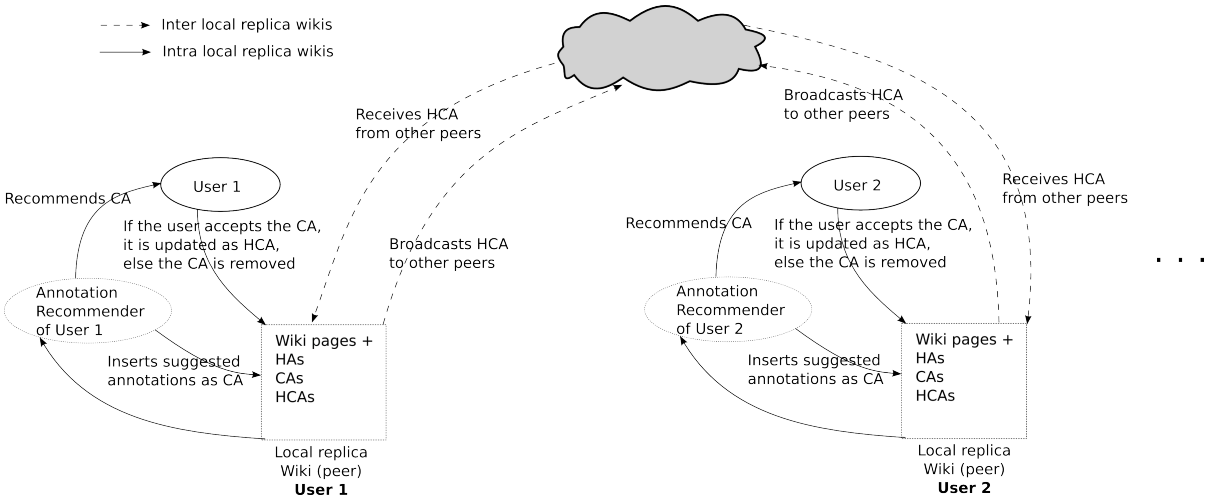

Figure 3: The way CA are proposed, validated and propagated in a P2P semantic wiki

Conclusion

This paper proposes an original system to encourage semantic annotations in semantic wikis of repositories of pedagogical resources. This ongoing work is based on the observation that users do not often semantically annotate pages in semantic wikis, as this task is not easy. This leads to semantic wikis with few annotations, they are therefore less efficient than they could be and students and professors cannot find resources that fit their requests. Our HCA system suggests to users annotations on the wiki pages. To annotate pages semantically, the users can either create semantic annotations as in a normal semantic wiki or can use the annotations suggested by the system. The HCA system exploits the usage the students and the professors make of the wiki pages and is based on classical collaborative filtering recommender systems; it uses item-based, classification-based and data mining-based approaches. These approaches learn the similarities/links, in terms of usage, between the wiki pages. The HCA system exploits both these similarities/links and the semantic annotations provided by humans that are present on the pages to suggest annotations to wiki pages. These annotations are not directly stored in the pages, they are suggested to users that can validate, modify or refuse these annotations. The resulting annotations in the semantic wiki are based on several levels of collaboration: explicit human collaboration when writing HA, implicit human collaboration to suggest CA and human-machine collaboration to produce HCA.

We are currently conducting experimentations and user studies to validate our approach; the preliminary results are very encouraging. As a future work, we intend to refine the way the HCA are obtained; we aim at not making a CA become directly a HCA when a user validates or modifies it, or being a RA if a user refuses it, but by waiting to have a consensus about the actions of the users. This approach, that will be based on a server of traces, will have the advantage that annotations will not change constantly.

References

1. Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art. IEEE transactions on knowledge and data engineering, 17(6):734-749.
2. Bonnin, G., Brun, A., and Boyer, A. (2009). A low-order markov model integrating long-distance histories for collaborative recommender systems. In Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI'09), pages 57-66, Sanibel Islands, USA.
3. Buffa, M., Gandon, F. L., Ereteo, G., Sander, P., and Faron, C. (2008). Sweetwiki: A semantic wiki. Journal of Web Semantics, 6(1):84-97.
4. Durao, F. and Dolog, P. (2009). Tag-based recommendation in kiwi. In Fourth Workshop on Semantic Wikis (SemWiki2009) at the 5th European Semantic Web Conference (ESWC 2009).
5. Goldberg, D., Nichols, D., Oki, B., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12):61-70.
6. Krotzsch, M., Vrandecic, D.,Vôlkel, M., Haller, H., and Studer, R. (2007). Semantic wikipedia. Journal of Web Semantic,5(4):251-261.
7. Mobasher, B. (2007). Data Mining for Web Personalization, chapter 3, pages 90-135. LNCS 4321 - Brusilovsky, P. and Kobsa, A. and Nejdl, W.
8. O'Connor, M. and Herlocker, J. (1999). Clustering items for collaborative filtering. In Proceedings of the 22th Annual International ACM SIGIR Conference.
9. Pazzani, M. and Billsus, D. (2007). The Adaptive Web, chapter Content- Based Recommendation Systems, pages 325-341. Springer Berlin / Heidelberg.
10. Reeve, L. and Han, H. (2005). Survey of semantic annotation platforms. In Proceedings of the 2005 ACM Symposium on Applied Computing.
11. Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In World Wide Web, pages 285-295.
12. Skaf-Molli, H., Rahhal, C., and Molli, P. (2009). Peer-to-peer semantic wikis. In Bhowmick, S. S., Kung, J., and Wagner, R., editors, DEXA, volume 5690 of LNCS, pages 196-213. Springer.
13. Rahhal, C., Skaf-Molli, H., Molli, P. and S. Weiss. Multi-synchronous Collaborative Semantic Wikis. 10th International Conference on Web Information Systems (Wise 2009), LNCS 5802.
14. Yong, W., Zhanhuai, L., and Yang, Z. (2005). Mining sequential association-rule for improving web document prediction. In Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05).
15. Boyer, A. and Dib, K. and Capul, J.Y. Providing French Universities with a Powerful Environment to Support E-learning development. In Proceeding of the 19th annual EDEN conference, June 9-12 2010.