HUMAN COMPUTER COLLABORATION TO IMPROVE ANNOTATIONS IN SEMANTIC WIKIS

Anne Boyer, Armelle Brun and Hala Skaf-Molli
LORIA-Nancy Université, 615, avenue du jardin botanique, 54506 Vandœuvre les Nancy, France

Keywords: Recommender systems, Semantic wikis, Automatic annotation.

Abstract: Semantic wikis are promising tools for producing structured and unstructured data. However, they suffer from a lack of user provided semantic annotations, resulting in a loss of efficiency, despite of their high potential. We propose a system that suggests automatically computed annotations to users in peer to peer semantic wikis. Users only have to validate, complete, modify, refuse or ignore these suggested annotations. Therefore, the annotation task becomes easier, more users will provide annotations. The 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 users. The resulting semantic wikis contain several kinds of annotations with different status: human, computer or human-computed provided annotations.

1 INTRODUCTION

Web 2.0 provides an easy way to produce new content and to make it broadly available on Internet. Users collaborate all together to increase quantity and quality of available contents by using for instance, wikis and blogs. This large production increases the difficulty to access the right information at the right moment. The question is no more to determine if a specific content exists but to find where it is available and how to access it. Therefore, it is mandatory to develop tools to help users finding the pertinent information on the web. Semantic wikis are one of the most promising approach to overcome this problem. Compared to classical wikis, they allow users to add *semantic annotations* in the wiki pages. These semantic annotations are not necessarily mapped to an existing ontology compared to the approaches detailed in (Reeve and Han, 2005). In semantic wikis, semantic annotations allow the emergence of "lightweight" ontologies in a cooperative way. 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 wiki pages. It will be possible to answer queries by exploiting information from different wiki pages, based on the semantic annotations. However, adding semantic annotations is not an easy task and is time consuming. The consequence is that only few users annotate pages semantically. As it is not easy to motivate users to provide annotations, many existing semantic wikis contain only few annotations, that decreases their potential added value. Having only manual semantic annotations is a bottleneck for semantic wikis. One possible solution is to propose a system that suggests pertinent annotations to users.

This paper proposes a new kind of collaboration between users and machine to produce pertinent, useful and reliable annotations in semantic wikis based on a recommender system. Many works about recommender systems (Goldberg et al., 1992; Adomavicius and Tuzhilin, 2005; Pazzani and Billsus, 2007) and semantic wikis (Krötzsch et al., 2007; Buffa et al., 2008) can be found in the literature. However, few existing works (Durao and Dolog, 2009) propose to use recommender systems for navigation personalization in 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 "computer recommended" annotations are suggested to users that can either validate (and/or complete), modify, refuse or ignore them. Based on these suggested annotations, the task of the users is made easier as the users do not have to create the annotation from the scratches.

Section 2 gives a brief overview of the proposed system. Section 3 summarizes the required background in term of semantic wikis and recommender systems. Section 4 describes how we use recommender systems to automatically determine possible

Boyer A., Brun A. and Skaf-Molli H.
HUMAN COMPUTER COLLABORATION TO IMPROVE ANNOTATIONS IN SEMANTIC WIKIS.
DOI: 10.5220/0002797900890094
In *Proceedings of the 6th International Conference on Web Information Systems and Technology (WEBIST 2010)*, page
ISBN: 978-989-674-025-2

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.

2 OVERVIEW OF THE SYSTEM

Semantic wikis allow users to add semantic annotations in the wiki pages. These semantic annotations are used for reasoning and finding pertinent answers to complex queries. The efficiency of semantic wikis relies among others on the quantity and the quality of the available semantic annotations. An essential problem still remains: how can we encourage and support users to provide semantic annotations? This paper addresses this question by introducing the machine as a partner of the collaborative process of annotation in peer to peer semantic wikis. The community of users becomes a community gathering humans and computers working together to produce semantic annotations. Humans 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, humans can add new semantic annotations and participate to the final decision of integrating or not a given annotation provided by computers. We propose a semi-automatic annotation tool for semantic wikis. A specific recommender suggests possible annotations to humans. Every human 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 :

- Human Annotation (HA) : this annotation results from human collaboration;
- Computer Annotation (CA) : this annotation is suggested by the recommender but not reviewed by humans;
- Human Computer Annotation (HCA) : this is a computer annotation that has been validated (and/or completed) by the users;
- 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 of the wiki pages by the humans to determine pertinent suggestions of annotations. It exploits implicit collaboration between users: we consider that users who share usages collaborate implicitly. This approach does not take into account the content of pages to perform recommendations, only the 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 semantic wiki about ”France”, let two semantic wiki pages ”Eiffel Tower” and ”Montmartre”, these two pages are not directly linked by semantic annotations. The ”Eiffel Tower” page has the semantic annotation : ”Gustave Eiffel” as a ”Designer”. However, ”Eiffel Tower” and ”Montmartre” are two well known touristic buildings in ”Paris”, therefore, the semantic wiki pages of these two buildings should be semantically linked. And usually users interested in ”France” often consult both pages. Our HCA system will automatically discover this link based on the users usage and it 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.

3 BACKGROUND

This section presents backgrounds on semantic wikis and recommender systems.

Semantic Wikis are an extension of wiki systems, they embed semantic annotations in the wiki content. These annotations allow to better organise and structure the wiki contents. Semantic wikis allow mass collaboration for creating and emerging ontologies. 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) (Krötzsch et al., 2007), SweetWiki (Buffa et al., 2008) and Swooki (Skaf-Molli et al., 2009). 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”. The following example shows of a wiki page and its corresponding semantic wiki page.

```
France is located in [Europe]
The capital of France is [Paris]
```

```
France is located in [located In::
 Europe] The capital of France is
[has Capital:: Paris]
```

Annotations express semantic relationships between wikis 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 so they are *Human Annotations* they correspond to the *Level 1* of collaboration as presented in section 2. Semantic wikis, as classical wikis, suffer from scalability, availability and performance problems and they do not support offline works (Weiss et al., 2007). To overcome these limitations, peer to peer extensions for semantic wiki are proposed. SWooki (Skaf-Molli et al., 2009) 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 provide personalization to users to cope with the well-known problem of overload of information (Adomavicius and Tuzhilin, 2005). Among the possible approaches in recommender systems are content-based (Pazzani and Billsus, 2007) and collaborative filtering approaches (Goldberg et al., 1992). The first approache uses the content of the resources to compute recommendations for users, these approaches are accurate. 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 the resources 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 consider only 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 exploits the traces of usage to deduce information about the resources. CF-based recommenders can either compute similarities between resources (Sarwar et al., 2001) or exploit data mining techniques to learn relationships between the resources (Yong et al., 2005). 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 relastionships 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.

4 RECOMMENDER SYSTEMS FOR ANNOTATION SUGGESTION

Existing recommender systems for Semantic Wikis directly transpose recommenders to suggest wiki pages to users as in (Durao and Dolog, 2009). 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 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 the ones presented in (Sarwar et al., 2001; Yong et al., 2005). Given the relationships between pages 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) .

4.1 Suggestion of Semantic Computer Annotations

We propose two algorithms to compute annotations to suggest to a given page P_j. The first one is similar to those used in item-based approaches and classification of pages based approach (O'Connor and Herlocker, 1999), the second one is based on data mining techniques for recommendations (Mobasher, 2007).

Item-based Approach. The algorithm first computes a similarity matrix of wiki pages (Sarwar et al., 2001). 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

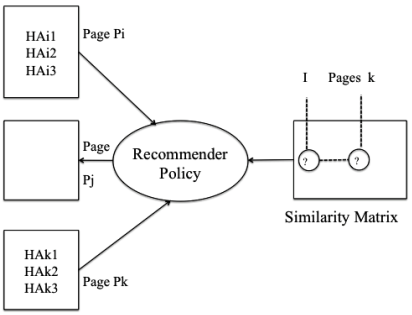

Figure 1: Automatic suggestion of HCA with the item-based approach.

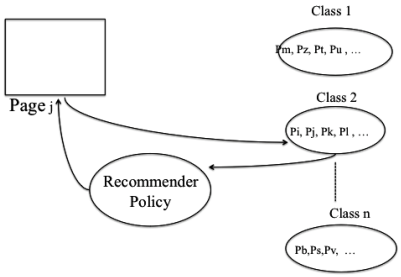

Figure 2: Automatic suggestion of HCA with the classification-based approach.

pages, 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 P_j, the recommender searches the pages that are similar to P_j, those with a high similarity value in the similarity matrix. The recommender collects the HA from the pages similar to P_j. The suggested CA to P_j can be computed by several policies: • 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. • 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. • 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.

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 P_j, 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.

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 (Bonnin et al., 2009). 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 section 3, there is no semantic annotation between the pages "France" and" Paris", 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 "Paris". The user can accept, modify, refuse or ignore this suggested annotation and type it with [HasCapital], for example. The annotation task is made easier as the system suggests to the user where useful annotations should be, the user is guided.

4.2 HCA Approach

Every semantic wiki page has three sources of annotations. Those added by humans (HA), they correspond to the explicit human collaboration, those suggested by the recommender system (CA), they come from implicit human interaction. These annotations are original annotations, they do not exist in classical semantic wikis, they will be used to encourage the users to annotate semantically pages. Those reviewed by the user, Human computer collaboration HCA. To integrate these annotations to a peer to peer semantic wiki, we have to answer several questions : how the recommender suggests annotations (CA) to users? how to make the users validate (and/or com-

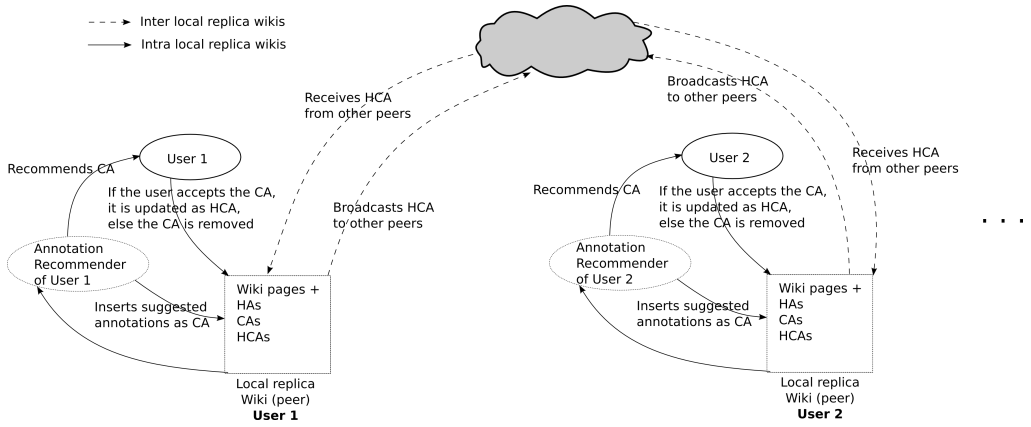

Figure 3: The way CA are proposed, validated (HCA) and propagated in a P2P semantic wiki.

plete), modify, refuse or ignore a CA? How to make the other users know these annotations? The way we propose to answer these questions is presented in Figure 3. When new annotations are suggested by the recommender, they are broadcasted to all the local replica with the status Computer Annotation (CA). They are proposed to the users, in a pop-up box, for example, to be differentiated from Human Annotations HA. The 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 the user does nothing, the annotations remain with a CA status. If the annotations are typed (from item-based or classification-based approaches), the user can either validate, modify or refuse each of them. If the annotations have no type (from data mining approach), the user 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 resuggesting annotations that have already been refused by users. These RA can however be resuggested in the case the content of the wiki page has been highly modified.

5 CONCLUSIONS

This paper proposes an original system to encourage semantic annotations in semantic wikis. 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. The HCA system suggests annotations on the wiki pages to users. The users can create semantic annotations as in a normal semantic wiki and can also use the annotations suggested by the system to annotate semantically pages. The HCA system exploits the usage of the wiki pages by the users 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, and suggests 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 is 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

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.

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.

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.

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)*.

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.

Krötzsch, M., Vrandecic, D., Völkel, M., Haller, H., and Studer, R. (2007). Semantic wikipedia. *Journal of Web Semantic*, 5(4):251–261.

Mobasher, B. (2007). *Data Mining for Web Personalization*, chapter 3, pages 90–135. LNCS 4321 - Brusilovsky, P. and Kobsa, A. and Nejdl, W.

O'Connor, M. and Herlocker, J. (1999). Clustering items for collaborative filtering. In *Proceedings of the 22th Annual International ACM SIGIR Conference*.

Pazzani, M. and Billsus, D. (2007). *The Adaptive Web*, chapter Content-Based Recommendation Systems, pages 325–341. Springer Berlin / Heidelberg.

Reeve, L. and Han, H. (2005). Survey of semantic annotation platforms. In *Proceedings of the 2005 ACM Symposium on Applied Computing*.

Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *World Wide Web*, pages 285–295.

Skaf-Molli, H., Rahhal, C., and Molli, P. (2009). Peer-to-peer semantic wikis. In Bhowmick, S. S., Küng, J., and Wagner, R., editors, *DEXA*, volume 5690 of *LNCS*, pages 196–213. Springer.

Weiss, S., Urso, P., and Molli, P. (2007). Wooki: a p2p wiki-based collaborative writing tool. In *Web Information Systems Engineering*, Nancy, France. Springer.

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)*.