New Tagging Paradigms for Content Recommendation in Web 2.0 Portals

Andreas Nauerz[1], Fedor Bakalov[2], Martin Welsch[1], and Birgitta König-Ries[2]

[1]IBM Research and Development, 71032 Böblingen, Germany,
{andreas.nauerz | martin.welsch}@de.ibm.com
[2]Friedrich-Schiller University of Jena, 07743 Jena, Germany,
{fedor.bakalov | birgitta.koenig-ries}@uni-jena.de

Abstract. With this paper we tie in with what we presented during last year's workshop [5] where we illustrated how to analyze users' tagging and rating behavior to construct user- and context models that can be used to perform adaptations and to issue recommendations in order to create more user-tailored Web Portals. This time we want to present more sophisticated tagging paradigms and their influence on users collaboration behavior and the construction of user- and context-models. The concepts presented are currently been prototypically implemented within IBM's WebSphere Portal and can be presented in a live demo at the workshop.

1 Introduction

In recent years Enterprise Information Portals have gained importance in many companies. As a single point of access they integrate various applications and processes into one homogeneous user interface. Today, typical Portals are comprised of a huge amount of content. They are no longer exclusively maintained by an IT department, instead, Web 2.0 techniques are used increasingly, allowing user generated content to be added. These systems grow quickly and in a more uncoordinated way as different users possess different knowledge and expertise and obey to different mental models. The continuous growth makes access to really relevant information difficult. Users need to find task- and role-specific information quickly, but face information overload and often feel *lost in hyperspace*. Thus, users often miss out on resources that are potentially relevant to their tasks, simply because they never come across them. On the one hand, users obtain too much information that is not relevant to their current task, on the other hand, it becomes cumbersome to find the right information and they do not obtain all the information that would be relevant.

The recent popularity of collaboration techniques on the Internet, particularly tagging and rating, provides new means for both semantically describing Portal content as well as for reasoning about users' interests, preferences and contexts. It can add valuable meta information and even lightweight semantics to web resources.

In our previous work [5] we proposed a framework which allowed arbitrary annotators, e.g. human users or analysis components (for automated tagging), to annotate any of these resources. Analysis of the tagging behavior allowed us to model interests and preferences of users as well as semantic relations between resources, and thus to perform reasonable recommendations and adaptations.

In this paper we will present tagging paradigms like alien tagging, reputation-based tagging, quantitative tagging, anti tagging, tag expiry, contextual tagging, and describe how these can be used to refine our models and to perform even more valuable adaptations or to issue more valuable recommendations.

2 Related Work

Using collaborative metrics to get to know what is of relevance to users or entire user groups has been done before. Especially collaborative ranking, i.e. ranking which takes into consideration entire communitys' interests, has recently become more important. Access patterns are used to assess the importance of single web pages [1]. Improved versions of the original PageRank [6] and HITS [3] algorithms have been developed (cp. FolkRank [2], CollaborativeRank [4]).

Other work focuses on the personalized recommendation of content based on its relatedness to certain tag terms. [7] proposes a modified version of the HITS algorithm to determine experts and high-quality documents related to a given tag.

3 Concepts

3.1 Alien tagging

As said before Web 2.0 communities can be rather heterogeneous. The expertise of users contributing (and consuming) content can vary a lot. What might be obvious for one user might be completely unknown to others. *Alien tagging* allows more experienced users to tag content for less experienced ones. In our prototypical implementation tag widgets allow power users to apply tags to resources on behalf of other users (or even user groups). Next time one of the users for which alien tags have been applied logs-in, he or she is notified about the availability of these and can inspect the underlying resources. The same way we used "normal" tags in our previous work [5] to refine user models that describe users interests and preferences we can use these alien tags, too. In real environments *alien tagging* could be used e.g. by managers pretagging content for their new hires, by team- or technical leads to point their team members to relevant content which they otherwise might have missed. Thus *alien tagging* opens another opportunity to prevent users from missing out content by issuing recommendations provided by "alien" users.

3.2 Reputation-based tagging

In our previous solutions we always assumed that the weight (i.e. the importance) of tags only depends on the frequency of their occurrence. I.e. a tag applied more often with respect to a certain scope was regarded of higher importance than a tag applied less often. In our new prototype we additionally assume that the weight of a tag can depend on the reputation (or expertise) of a user. I.e. that tags applied by more experienced users have higher weights, and thus higher influence on what content the community is presented (or recommended) with, than tags from less experienced users. This way we can point users to more relevant content as we assume experts to know better what the community should focus on. E.g., in development team we assume the tagging behavior of the team- or technical lead of higher importance. With *reputation-based tagging* we also ensure that "incorrect or less suited" tags perceive lower weights (influence). E.g., a newbie might apply a more "incorrect/less suited" tag as he just misunderstands (due to his insufficient knowledge) what he is looking at. The way we determine users' expertise has already been described in [5].

3.3 Quantitative tagging

Previously we also assumed that tags can only have "positive character". I.e. that we assumed that a resource can be tagged with a term to describe that the resource has something to do with this term, but also assumed that a resource cannot be tagged with a term to describe that the resource has nothing to do with it. In addition to that aspect we did not provide means for single users to express that a certain tag is of less relevancy for them. *Quantitative tagging* provides a solution to both problems: in our prototypical implementation a plus- and a minus sign is presented besides each tag being displayed. In addition, when applying a tag, a not-sign is presented. Clicking the not-sign when applying a tag allows users to express that a resource has nothing to do with the term applied, a helpful feature for more fine-granular categorization of resources: e.g., users could tag some resources with the term *Web 2.0* and a few of them with "not" *scientific*. This helps users to quickly find all Web 2.0 related resources and to quickly distinguish between the scientific and non scientific ones among them. Clicking the plus- and minus-signs when working with tags allows single users to express that they are less interested in a tag (or a certain tag associated to a certain resource) or can additionally express that a tag is of less relevancy for the entire community. Thus, these mechanisms allow for further refinement of our user models.

Anti tagging *Anti tagging* describes an enhancement to *quantitative tagging* (cp. 3.3). Here we automatically increase or decrease tags' relevancy for the entire community by analyzing tags semantics (cp. [5]). One option we have evaluated is to take into consideration antonyms. E.g., when a resource is tagged with "good" and "bad" we regard it as not tagged at all with either of these two terms as they annihilate each other. Antonyms can e.g. be found using the antonym

thesaurus [1]. As *anti tagging* is not trivial to be realized as most examples are much more complicated and less obvious than the one just provided we have not yet incorporated it in our prototype.

3.4 Tag expiry

In our previous work we also assumed that tags can be applied once and stay alive until they are manually deleted again. This let to tag-space littering as most users never deleted tags anymore even if they became obsolete. The fact that tags do not remain valid forever occurs in Portals that provide dynamic content very often. This resulted in having a lot of tags assigned to resources that did not describe the resource adequately nor express the resources relevancy to the community appropriately anymore. In our prototype *tag expiry* allows users to specify a chronological validity for tags when assigning them to a resource. Taggers can give tags a start date, an end date or a time frame in between they live. We also allow tags that are assigned a "lifetime" to become more (or less) important as time passes by. E.g. if there is a page in the Portal system providing information about the Olympic Games 2012, this page might become more and more interesting to users as we get nearer to the year 2012 and less interesting after 2012. Thus users can specify that the tag should not be available before 2011, vanish after 2013 and become more important from 2011 till 2012 and less important from 2012 till 2013. Thus, *tag expiry* is yet another mechanism to help the community to focus on what is currently really relevant. Moreover, *tag expiry* allows us to neglect "invalid" tags from being considered when doing content adaptation or recommendation.

3.5 Tagging tags and meta-tagging

Previously we have also worked on solutions to solve major problems of tagging systems: most of these problems discussed dealt with synonyms (multiple tags having the same meaning) and polysemies (a single tag having different meanings). Current tag engines often try to overcome these issues by applying stemming and normalization algorithms which most often only solve problems resulting from morphological variations. Semantical variations can most often not be detected to be a synonym e.g. In our latest prototype we allow the community to resolve the resulting tag-space littering. In our tag-clouds we allow users to drag and drop tags on each other to consolidate them. In addition to that we allow users to create meta-tags (or meta-tag bags as we call them) under which other tags can be organized. Users could create private meta-tag bags only they can see or community meta-tag bags all users part of the community can see. That way users could e.g. create a meta-tag bag "sports" drag all sports related tags into that bag; users could also create a meta-tag bag "favorite-stuff" and just drag what he/she likes most into it.

[1] http://www.synonym.com/synonyms/

3.6 Other concepts

We are also allowing for tag sharing among subcommunities. Most current tagging systems allow to either create public or private tags but do not allow for a granularity in between. Our prototype allows to share tags with a dedicated set of other users. We also allow for contextual tagging where we can associate tags a certain context (for our context modeling approaches refer to [5]) to prevent irrelevant tags (irrelevant in a certain context) to appear. The latter helps focusing on currently relevant content again.

4 Conclusion and Future Work

In this paper we have presented tagging paradigms which we are using to refine our user- and context modeling approaches presented in our previous work [5] in order to perform content adaptation and recommendation. The concepts described have already been prototypically implemented and can be presented at the workshop. We have not yet performed in-depth evaluation on these early ideas described in this short paper but are looking forward to discuss them and receive initial feedback. Of course, especially the usefulness of each single concept has still to be evaluated.

For the future we plan to merge our Web 2.0 collaborative tagging approaches with Semantic Web ideas heading towards the Web 3.0.

References

1. J. Caverlee, Ling Liu, and W.B. Rouse. Link-based ranking of the web with source-centric collaboration. *colcom*, 0:11, 2006.
2. Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Folkrank : A ranking algorithm for folksonomies. In *LWA*, pages 111–114, 2006.
3. Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In *SODA*, pages 668–677, 1998.
4. Amir Michail. CollaborativeRank: Motivating People to Give Helpful and Timely Ranking Suggestions. http://collabrank.web.cse.unsw.edu.au/collabrank.pdf, April 2005. (work in progress).
5. Andreas Nauerz, Stefan Pietschmann, and Rene Pietzsch. Social recommendation and adaptation in web portals. In *Proceedings of the International Workshop on Adaptation for the Social Web*, Hannover, Germany, 2008.
6. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford Digital Library Technologies Project, 1998.
7. Harris Wu, Mohammad Zubair, and Kurt Maly. Harvesting social knowledge from folksonomies. In *Proc. of the 17th Conf. on Hypertext and hypermedia*, pages 111–114, New York, NY, USA, 2006. ACM Press.