A Tag-Based Blog Recommendation System for Smart-M3 Multi-Blogging

Diana Zaiceva, Dmitry Korzun
Petrozavodsk State University
Petrozavodsk, Russia
{zaiceva, dkorzun}@cs.karelia.ru

Abstract

Blogosphere is a massive and heterogeneous media-space that recently lacks of mechanisms for "smart" information discovery. The latter becomes even more challenging when mobile multi-blogging appears, and any user can access many blogs and blog-services in parallel using a multitude of available mobile user devices. SmartScribo is a reference application for mobile multi-blogging based on the Smart-M3 information sharing platform. In this state-of-the-progress paper we consider the problem of personalized blog recommendation focusing on basic recommendation system design for SmartScribo. In accordance with the ubiquitous computing vision, details of the search mechanism are made invisible to users; they are fed only with recommendations for "worth-to-read" blogs. For each user and tag the recommendation system computes a personalized tag rank that reflect recent activity of the user for this tag. Then a small set of blogs with high-ranked tags can be retrieved from the blogosphere and provided to the user. The core of our design is a tag ranking model to support dynamic personalized rank computation in Smart-M3 smart spaces. Its implementation and experimental validation are our further research directions.

Index Terms: Smart Spaces, Blogosphere, Smart-M3, Multi-blogging, Recommendation system.

I. Introduction

Recently the size of the blogosphere has been exhibiting the exponential increase. According to [1], WordPress continued in 2010 its growth to 27 millions publishers and 13.9 million blogs hosted. Another fact from [1] is that BlogPulse has 152,948,304 new blogs created in 2010. The hugeness, heterogeneity, and rapid evolution result in difficulties for many bloggers for searching personally-interesting topic-based discussions in the blogosphere.

Mobile blogging is also exposes the significant growth. According to [2], about 25% of all bloggers are already engaged in mobile blogging. Relatively low computational power and local memory of mobile devices, their small screens, and other specifics lead to additional problems in searching the blogosphere. Infrastructural support is needed to which low-performance user devices can delegate complex processing. Examples of such infrastructural designs are Cobra [3] and blog search engines like Technorati [4].

Multi-blogging [5] is a consequence of the service convergence trend of smart computing environments [6]. It leads to additional challenges for searching the blogosphere. A user can access many blogs in parallel; she or he would like to operate with the blogosphere as a whole, independently on which blog-services appropriate blogs are hosted. Corresponding infrastructural support has to have an efficient and interoperable information sharing mechanism that connects multiple clients with their multiple services.

The smart spaces paradigm aims at this kind of information sharing and interoperability [7]. Smart-M3 [8] is a concrete open-source platform for smart spaces. SmartScribo [5], [9] is a Smart-M3 application that aims at providing advanced access to the blogosphere using

the smart spaces paradigm. SmartScribo users can interact with multiple blogs at many blog services using their (mobile) devices as clients. Blog-related data and knowledge about the person and her/his current interests are stored in blogger's personal smart space, which is shared (partially) with other user's devices and applications as well as with other users.

In this paper we show our initial attempt to consider the problem of searching personally-interesting topic-based discussions in the blogosphere. We suggest a design solution of a personalized blog recommendation system to be implemented in SmartScribo application. The problem of personalized recommendations for social networks is well elaborated, e.g., see surveys [10], [11], [12]. In the case of blogging, many techniques have been proposed in the last decade, e.g., see [13], [3], [14]. Nevertheless, to the best of our knowledge no blog recommendation system is available for Smart-M3 platform. The paper makes a step towards this direction considering some basic architectural and design peculiarities appropriate for Smart-M3 based application.

Since our focus is on high-level design solutions we limit ourselves with a simple tag ranking model. The latter aims at identifying those blog tags (keywords) that are of the most user interest. Accordingly, proactive blog retrieval from multiple blog-services is possible, and the user is provided with personal recommendations of blogs with high-ranked tags. In our design, the recommendation system computes tag ranks adaptively with the dynamics of user's current interests. The system design is intentionally set up to exploit properties of distributed multi-agent architecture of SmartScribo and its smart spaces infrastructural support.

The paper is organized as follows. Section II introduces the SmartScribo application within which we plan to implement our blog recommendation system. Section III describes the recommendation scenario from the user perspective. Section IV describes tag ranking model, which is a core of our recoimendation system and allows adaptive computations based on SmartScribo architecture and its blogosphere smart space. Section V introduces the related work for the problem of blog search and recommendation. Section VI summarizes the paper.

II. SmartScribo application

SmartScribo [5], [9] is a Smart-M3 application for mobile multi-blogging. It connects bloggers with their blogs hosted at multiple blog-services. We use term "user" for a blogger who uses SmartScribo. She/he can perform the same blogging functions as a non-SmartScribo user and additionally benefits from smart features that SmartScribo provides to its users, e.g., because of cooperative blog sharing and dissemination.

SmartScribo constructs a blogosphere smart space, which acts as a mediator infrastructure for bloggers. User and blog data are kept dynamically and semantically related in the blogosphere smart space. It reflects only a part of the blogosphere that is actually used by the users. The blogosphere smart space is a composition of user personal spaces. Every smart personal consists of three subspaces: person (user profile), context, and blog data. An OWL ontology specification is used to define the structure of a personal smart space [9].

The distributed architecture consists of three types of agents (knowledge processors or KPs in Smart-M3 terms): clients, blog processors, and mediators. Clients run on user devices (mobile or stationary), operate with blogs locally, and send notifications to the blogosphere smart space for operations at blog-services. Blog processors track notifications from clients relaying particular blog-services. Mediators extend blogging with smart features.

In this paper, we design the recommendation system that can be integrated into SmartScribo based on the above architectural properties. In particular, we assign a dedicated mediator to

track the user activity in the user personal space, to compute tag ranks, and to query blog search engines for appropriate blogs.

We use the context subspace to track "statuses" of blog data. Status reflects such user activity as sending, reading, or deleting blog messages. The person subspace makes this context person-aware, i.e., it can be related to a given user. Relations between personal spaces support collective activity of users, and status can reflect the affection of a user group (e.g., all user's friends) to a given user.

SmartScribo provides a notification mechanism for coordination between KPs. Notifications initiate KP to execute some actions or to inform about execution result. A KP subscribes to notification data in the smart space. Whenever such data have been changed the KP is notified. This notification mechanism supports proactive service discovery, which is important for our recommendation system. Client can receive notifications about blog recommendations without explicit activation from its side.

III. Recommendation scenario

In this section we consider a scenario of proactive post recommendation. A post is an entry point of a blog discussion, hence the scenario aims at discovering intelligently entry points of blog discussions that are potentially interesting for a given blogger. With this scenario, recommended blog discussions become available in the SmartScribo blogosphere smart space and then they are provided proactively to the user.

A known way to describe post content semantics is tag inclusion into posts. A tag t is a keyword that a blogger augments to her/his post, e.g., $t =$ 'football' or $t =$ 'soccer'. Tags can be used to decide either a given post fits the user thematic interests. If a post p includes t then we denote $t \in p$.

Another useful knowledge source for post recommendation is the user activity context. We assume that if a blogger frequently reads posts with a certain tag then any post with this tag seems interesting for the blogger. Note that SmartScribo is oriented particularly to context-aware blogging scenarios and context can be ontologically represented in blogger personal smart spaces [5], [9].

Our recommendation scenario is based on tag ranking. The model of ranking will be introduced in the next section. Now let us consider the basic steps of the scenario and their possible implementation scheme.

1. *Activity accounting.* When a user is reading a post p or sending it to a blog-service, this action is accounted in the personal smart space (by SmartScribo clients). Let T_{pu} be the latest access time of p by user u. This parameter is stored as a post attribute, which is easily implemented in the SmartScribo ontological data model by adding a data property to the class 'post'. The attribute can be generally treated a part of *the post status* in the blogosphere smart space. Note that there can be different strategies for selecting which posts are subject to the status accounting. In the simplest case, all posts are accounted for any user.

2. *Post status tracks.* A dedicated blog mediator has subscribed to post statuses in the personal smart space of each user. Whenever the mediator is receiving a notification about a post status change, the former determines the following parameters (using additional queries to the smart space if needed): the affected post p the user u who has changed the status, the access time T_{pu}, and all n_p tags that p includes. Other knowledge can be determined as well; it depends on what input data the ranking algorithm requires.

3. *Tag index and ranks.* The mediator maintains a special blog tag index, where for every tag t all posts with t and all users that own a post with t are kept as references. Also, the

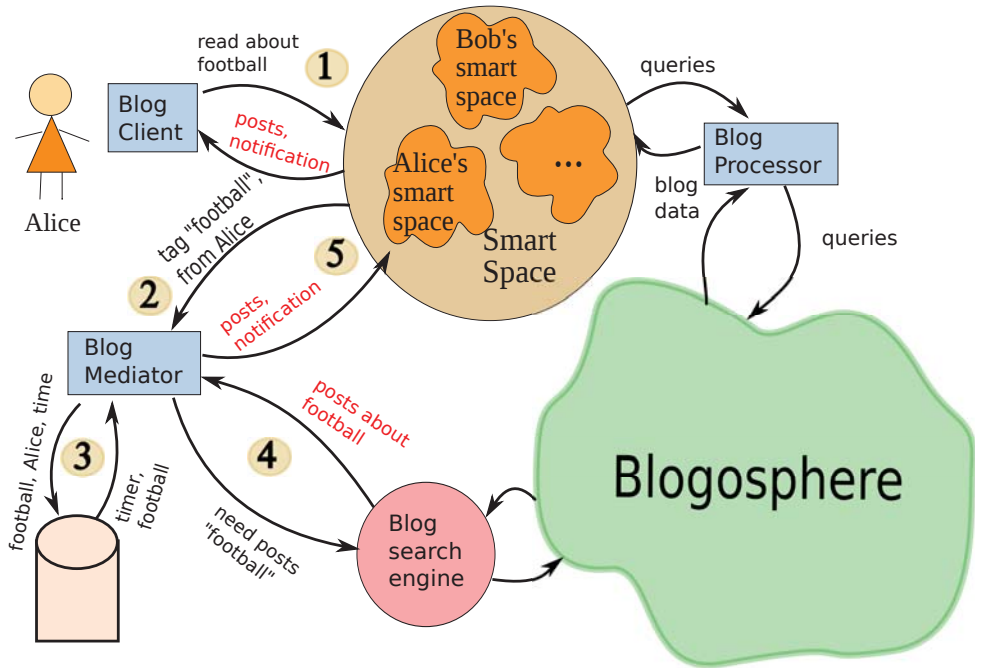

Fig. 1. Example of the post recommendation scenario. Step 1: Alice regularly reads posts with tag $t =$ 'football'. Steps 2&3: Her blog mediator M_{Alice} tracks Alice's activity updating $R_{t,\text{Alice}}$ in the tag index. Step 4. When M_{Alice} detects a burst in activity for t then M_{Alice} queries the blog search engine for posts with t. Step 5: M_{Alice} publishes incoming posts, making them available to Alice.

mediator computes and stores the tag rank R_{tu} for every t and user u, where R_{tu} is non-negative real value. Higher values are for those tags that are more interesting from u's point of view. Our tag ranking model and algorithm will be introduced in Section IV.

4. *Blog discovery and retrival.* The mediator has a threshold rank value. Whenever a rank has exceeded the threshold the blog discovery is activated. The mediator queries a blog search engine (specified in advance) using this tag as a search key. A particular candidate for this engine is Technorati [4]. The mediator publishes the posts it receives in u's personal smart space. The latter function can also be delegated to SmartScribo blog processors using the notification mechanism. There can be other strategies for the activation of blog discovery. For instance, by the tag timer expiration, where the timer keeps the time elapsed from the previous post update for the given tag.

5. *Blog recommendation.* SmartScribo client receives notifications about new posts in the blogosphere smart space and shows a banner with a list of recommended posts. The user selects to either accept or refuse. Indeed, there is an option in the client to switch the recommendation feature off as well as additional control, e.g., changing the recommendation threshold or selecting another mediator.

A simple example of the above steps is shown in Figure 1. SmartScribo notification mechanism [9], [15] provides synchronization of blog clients and blog mediators. Our scenario does not require sharing her/his interests as explicit knowledge in the smart space. Only information about tags and tag-related user activity is utilized. The scenario has certain time latency because of the sequential control and data flows between the participating entities.

IV. TAG RANKING MODEL

Tag rank R_{tu} characterizes the importance of blogs with tag t to user u. In our recommendation scenario from Section III), Step 4 requires recomputation of R_{tu} whenever u's activity is affecting t. In this section we discuss a simple tag ranking model, which can be used the tag rank recomputation.

Let a blog mediator M be notified about a status update of post p. Then M indentifies the user u who changed the status and analyzes all tags $t \in p$. If t is new then M adds t to the tag index with rank $R_{tu} = 1/n_p$, where $n_p \geq 1$ is the sum number of tags in p. Otherwise, M updates the rank from R_{tu}^{old} to

$$R_{tu} = (1 - \alpha)R_{tu}^{\text{old}} + \frac{\alpha}{\delta_t n_p}, \tag{1}$$

where $\delta_t \geq 1$ is the time elapsed from the latest access to a post with tag t and $0 < \alpha \leq 1$ is the weight for averaging old and new information. Intuitively, the larger δ_t the lower appearance of t in all posts the bloggers read, and t becomes of less importance to u. The larger n_p the more blurred (less focused thematically) the post p is, and t again tends to be less important to u.

The straightforward way to calculate δ_t for (1) is

$$\delta_t = \delta_t(u) = T_0 - \max_{t \in q, q \neq p} T_{qu}, \tag{2}$$

where T_0 is the current time. That is, if u has read posts with t very recently then u likely accepts other blogs with t.

Instead of u's individual activity in (2), we can capture collective activity for blogs with t. Let C be a group of users, e.g., friends of u. Assume also $u \in C$. If blogs with t are actively used by $v \in C$ then t becomes important to u, and (2) is generalized to

$$\delta_t = \delta_t(C) = T_0 - \max \left\{ \max_{t \in q, q \neq p} T_{qu}, \max_{v \in C \setminus \{u\}} \max_{t \in q} T_{qv} \right\}. \tag{3}$$

Another functional can be used instead of $\max$ in (2) and (3). For instance, M can use moving average of T_{qv} to aggregate short-term history of v's activity. Instead of the group maximum, the group average can be used to estimate the collective activity. Another possible modification is using δ_t not as a time estimate but as a frequency estimate, i.e., let δ_t be the access rate for blogs with t.

If M maintains a full tag index (see step 3 in Section III) then M can access locally the knowledge about all q with $t \in q$ and about all their users v. Therefore, recomputation (1) can be performed without expensive queries to the smart space.

V. RELATED WORK

The problem of personalized recommendations in social networks has at least 15-year history of research and development. An introduction to the area can be started with surveys [10], [11], [12]. Clearly, some generic methods can be also used in the Web, Semantic Web, and in particular for blog recommendations. Our design, however, is primarily oriented to integration into SmartScribo, a Smart-M3 application. Due to this primary goal, we keep the underlying ranking model very simple. The core scenario and architecture will remain the same even if the model is further evolved or replaced with another one.

Recently many techniques and designs have been proposed for blog recommendations, e.g., see [13], [3], [14]. They have certain architectural difference with our Smart-M3 based approach.

The blog distillation problem is about identification of other key bloggers with similar interests to their own. The problem was considered in [13] and they showed that expert search techniques can be effectively adapted for identifying key bloggers. In this case, experts have to be employed, and recommendations cannot be done in an automatic mode.

There are blog recommendation systems that are based on RSS. An example is Cobra (Content-Based RSS Aggregator) [3], a distributed system of crawlers, filters, and reflectors that crawls, filters, and aggregates RSS feeds. It delivers to each user a personalized feed based on user interests. Interactions between elements of this large system follow a specific protocol, which essentially differs from the interoperable Smart-M3communication model.

Content-based recommendation with a stochastic graph method was proposed in [14]. Given a topic, it finds a small subset of high-ranked blogs. Blog ranks are calculated recursively in terms of the ranks of their neighbors in a lexical similarity graph. In this graph, nodes represent blogs (or posts in the simplest case) and edges link lexically similar blogs. This approach requires analysis of a large graph to produce reasonable recommendations. In the case of multi-blogging, when parts of this graph are distributed among many blog-services, this approach is difficult for implementation.

VI. Conclusion

This paper proposed a blog recommendation system design based on personalized tag ranking. We considered the basic scenario for searching and recommending blogs appropriate for a given user. We primarily orient to integration into SmartScribo, an application for multi-blogging in smart spaces. It follows Smart-M3 principles for coupling multiple users and heterogeneous sources of information.

At the current phase, evaluation of this design can use simple blog recommendation techniques. We introduced the tag ranking model assuming that worth-reading blogs have posts with tags of user's most interest. The system identifies the most important tags in the blogosphere and the identification is personalized to a user. We dynamically compute rank for every tag and user based on observable activity for the tag in the blogosphere. This design allows distributed computation of tag ranks. The ranking model can be further evolved or replaced with another one.

References

[1] A. Morris, "A look into the 2010 state of the blogosphere: Report by Technorati," Jan. 2011. [Online]. Available: http://www.ignitesocialmedia.com/blogging/2010-state-of-the-blogosphere-report-by-technorati/

[2] J. Sobel, "State of the blogosphere 2010: Introduction," Technorati Blog, Nov. 2010. [Online]. Available: http://technorati.com/blogging/article/state-of-the-blogosphere-2010-introduction/

[3] I. Rose, R. Murty, P. Pietzuch, J. Ledlie, M. Roussopoulos, and M. Welsh, "Cobra: Content-based filtering and aggregation of blogs and RSS feeds," in *Proc. 4th USENIX Symp. on Networked Systems Design & Implementation (NSDI)*, Apr. 2007, pp. 29–42. [Online]. Available: http://www.usenix.org/events/nsdi07/tech/rose.html

[4] "Technorati blog search engine and directory," Sep. 2011. [Online]. Available: http://technorati.com/

[5] D. Zaiceva, I. Galov, A. Sannikov, and D. Korzun, "Mobile multi-blogging in Smart-M3: Architecture and scenarios," in *Proc. 8th Conf. of Open Innovations Framework Program FRUCT*, Nov. 2010, pp. 225–233.

[6] S. Balandin and H. Waris, "Key properties in the development of smart spaces," in *Proc. 5th Int'l Conf. Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments (UAHCI '09)*. Springer-Verlag, 2009, pp. 3–12.

[7] I. Oliver, "Information spaces as a basis for personalising the semantic web," in *Proc. 11th Int'l Conf. Enterprise Information Systems (ICEIS 2009)*, May 2009, pp. 179–184.

[8] J. Honkola, H. Laine, R. Brown, and O. Tyrkkö, "Smart-M3 information sharing platform," in *Proc. IEEE Symp. Computers and Communications*, ser. ISCC '10. IEEE Computer Society, Jun. 2010, pp. 1041–1046.
[9] D. Zaiceva, I. Galov, and D. Korzun, "A blogging application for smart spaces," in *Proc. 9th Conf. of Open Innovations Framework Program FRUCT and 1st Regional MeeGo Summit Russia–Finland*, Apr. 2011, pp. 154–163.
[10] A. Abraham, A.-E. Hassanien, and V. Snsel, *Computational Social Network Analysis: Trends, Tools and Research Advances*, 1st ed. Springer Publishing Company, Incorporated, 2010.
[11] A. Micarelli, F. Gasparetti, F. Sciarrone, and S. Gauch, "The adaptive web," ser. LNCS 4321, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Springer-Verlag, 2007, ch. Personalized search on the World Wide Web, pp. 195–230.
[12] M. Gao, K. Liu, and Z. Wu, "Personalisation in web computing and informatics: Theories, techniques, applications, and future research," *Information Systems Frontiers*, vol. 12, pp. 607–629, Nov. 2010.
[13] C. Macdonald and I. Ounis, "Key blog distillation: ranking aggregates," in *Proc. 17th ACM Conf. Information and knowledge management (CIKM '08)*. ACM, 2008, pp. 1043–1052.
[14] A. Hassan, D. Radev, J. Cho, and A. Joshi, "Content based recommendation and summarization in the blogosphere," in *Proc. 3rd Int'l AAAI Conf. Weblogs and Social Media*. AAAI, 2009, pp. 34–41.
[15] D. Korzun, I. Galov, A. Kashevnik, K. Krinkin, and Y. Korolev, "Integration of Smart-M3 applications: Blogging in smart conference," in *Proc. 11th Int'l Conf. Next Generation Wired/Wireless Networking (NEW2AN'11) and 4th Conf. Smart Spaces (ruSMART'11)*. Springer-Verlag, Aug. 2011, pp. 51–62.