Supporting Personalized User Concept Spaces and Recommendations for a Publication Sharing System

Antonina Dattolo, Felice Ferrara, and Carlo Tasso

University of Udine, Via delle Scienze 206, 33100 Udine, Italy
{antonina.dattolo,felice.ferrara,carlo.tasso}@dimi.uniud.it

Abstract. Current publication sharing systems weakly support creation and personalization of customized user concept spaces. Focusing the attention on the user, SharingPapers, the adaptive publication sharing system proposed in this paper, allows users to organize documents in flexible and dynamic concept spaces; to merge their concept map with a social network connecting people involved in the domain of interest; to support knowledge expansion generating adaptive recommendations. SharingPapers presents a multi-agent architecture and proposes a new way of representing user profiles, their evolution and views of them.

1 Introduction

Over the last decade, the Web has undergone great changes; there is a growing evidence of two parallel worlds, the traditional world constituted by expert and selected contributors and the new Web 2.0-based world, in which each user may become author, tag and share documents with a world wide community.

In this new context, an interesting example is provided by the publication sharing systems [1,2]; unfortunately, these systems weakly support creation and personalization of customized user concept spaces [3], representing them in a static and flat way. This problem has been partly analyzed in Bibsonomy [1], that allows users to organize the tags into hierarchies by exploiting an $if...then$ relation; this approach enhances the manual tagging activity, but it does not offer either support for organizing knowledge or for personalized recommendations.

On the other hand, collaborative [4], content [5] and hybrid [6] recommendation frameworks improve searches over the available information bases, but few works (such as [7,8,9]) use the tags for recommending new resources: in [7], the authors use a extension of the PageRank algorithm for ranking resources, tags and users in a folksonomy; in particular, in [8], the authors use hierarchical clustering of tags for personalizing navigational recommendations; in [9], the authors measure the users' similarity considering their past tag activity and inferring tags' relationships based on their association to content.

Nevertheless, such recommendation systems consider only the tags and not the goals and the context of the user's tagging activity.

G.-J. Houben et al. (Eds.): UMAP 2009, LNCS 5535, pp. 325–330, 2009.

In this paper, we present SharingPapers, an adaptive publication sharing system, that allows users: to organize documents in flexible and dynamic concept spaces, using innovative and dynamic data structures (the Nelson's zz-structures [10]); to merge their concept map with a social network connecting people involved in the domain of interest; to support knowledge expansion generating adaptive recommendations. These recommendations are generated analyzing the user's concept space, and evaluating the similarities among them in order to reveal the similarity among goals and perspectives of each user. The paper is organized as follows: in Section 2 we describe the architecture of SharingPapers; then we deepen the discussion about the organization of user concept spaces in Section 3, and we propose a simple schema of recommendations in Section 4. Finally, Section 5 concludes the paper.

2 SharingPapers

SharingPapers presents an agent-based architecture shown in Figure 1.

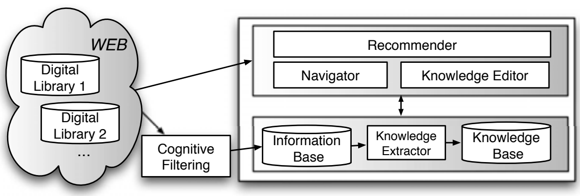

Fig. 1. System architecture

The main modules are:
- the *Cognitive Filtering* module uses the IFT algorithm [11] and specialized agent classes for browsing and accessing a set of external sources (Web sites and digital libraries), looking for relevant documents. The filtering operation is performed according to a set of defined information needs and populates the *Information Base*.
- The *Knowledge Extractor* module is specialized in extracting, from documents present in Information Base, attributes (such as the title of a paper, its authors, its year of publication) and relations (such as the network constituted by co-authors, or by people having a same affiliation, etc.), in order to populate the *Knowledge Base* (see Section 3);
- The *Navigator* module provides views on the Knowledge Base, enabling users to navigate among documents and social networks. Examples of views have been proposed in [12].
- The *Knowledge Editor* module implements the features users can invoke in order to manually modify and re-arrange their personal space, defined as concept space (see definition in Section 3); more specifically, each agent keeps track of the interaction of each user and translates the actions performed by himself into a set of operations on his/her concept space: users can create new entities, add

them to their concept spaces, or connect them with existing entities.
- The *Recommender* module suggests tags, recommends to visit parts of concept spaces (belonging to other users) and calculates personalized rankings on papers.

3 Organizing the Knowledge Base

In our system, the users are represented by their *concept space*: it contains a collection of *papers* and a *social network*.

Papers are connected in an innovative structure by links (indicating, for example, common keywords or tags), while the social network is constituted by users sharing interests and/or contents. A user concept space presents a dynamic structure, evolving in accordance to user behavior (new searches, adding-deleting new contents or tags, etc.).

The **concept space** (Map) related to the user u is formally defined by $M_u = (S_u, En_u, Re_u, Ac_u)$ where: S_u represents its *topological structure*; $En_u = \{\eta_{1_u}, \eta_{2_u}, \ldots\}$ defines its local *environment*; $Re_u = \{\rho_{1_u}, \rho_{2_u}, \ldots\}$ is the finite set of incoming *requests*; $Ac_u = \{\alpha_{1_u}, \alpha_{2_u}, \ldots\}$ is the discrete, finite set of possible *actions*.

In particular, $S_u = (MG_u, T_u, t)$ is a zz-structure, an *edge-colored multigraph* where $MG_u = (V_u, E_u, f)^1$ is a multigraph, in which the set of vertices $V_u = \{P_u, U_u\}$. P_u is the collection of papers of the user u, U_u the set of users connected to u; T_u is a set of colors (T refers to Tag), and $t : E_u \to T_u$ is an assignment of colors (tags) to edges of the multigraph; $\forall x \in V_u,\ \forall k = 1, 2, ..., |T_u|,\ deg^k(x) = 0, 1, 2^2$. Interested readers will find a deeper discussion about zz-structures in [10], [3], and [12]. In Figure 2 (left) is shown a graphical example of a generic M_u.

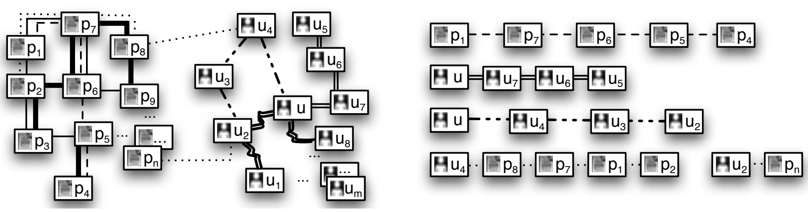

Fig. 2. An example of user concept space (left) and four dimensions of it (right)

$P_u = \{p_1, \ldots, p_n\}$ contains papers of interest for u, while $U_u = \{u, u_1, \ldots, u_m\}$ contains his/her social network; 7 different colors-tags (identified with different types of line style - normal, thick, dashed, double, etc.) are associated to the edges. Each tag identifies a link among vertices; for example, the tag (dashed line) connecting p_1, p_7, p_6, p_5, p_4 represents papers sharing a same tag or topic;

[1] Multigraph definition: $MG_u = (V_u, E_u, f)$ is a multigraph composed of a set of *vertices* V_u, a set of *edges* E_u and a surjective function $f : E_u \to \{\{v, v'\} \mid v, v' \in V_u, v \neq v'\}$.

[2] $deg^k(x)$ denotes the degree (that is, the number of edges incident to x) of color t_k.

the tag (double line) connecting u, u_7, u_6, u_5 indicates co-authors of one or more papers; the tag (long dashed line) connecting users u, u_4, u_3, u_2 groups members of the same research group; the tag (dotted line) connecting users and papers in u_4, p_8, p_7, p_1, p_2 and u_2, p_n identifies the author and a set of his/her papers.

For each color t_k, we may isolate a specific sub-graph of M_u, constituted by the set of vertices V_u and edges $E_u^k \in E_u$, containing edges of the unique color t_k. Each sub-graph of M_u is called *dimension* of color t_k and is denoted by D_u^k. Formally, a dimension $D_u^k = (V_u, E_u^k, f_u, \{t_k\}, t_u)$, with $k = 1, ..., |T_u|$, is a graph such that (1) $E_u^k \neq \emptyset$; (2) $\forall x \in V_u$, $deg_u^k(x) = 0, 1, 2$.

Using dimensions, the topological structure of M_u can be seen as $S_u = \bigcup_{k=1}^{|T_u|} D_u^k$. In this way, a dimension is defined in terms of one or more connected components, that it, some paths and a set (eventually empty) of isolated cells. For example, four paths present in M_u are shown in Figure 2 (right).

When the user enters in the system for first time, his/her concept space is automatically initialized by a set of dimensions. Papers that the user wrote, cited or tagged are imported in specific dimensions, as well as the papers presented in the events (conferences, journal, workshop) that (s)he attended. Similarly, co-authors and other people involved in the user research activity are also imported in the social network considering common publications, events and organizations. As second step, users can invoke the Knowledge Editor in order to manually modify and re-arrange their concept spaces. In this way, users can create new entities, add them to their concept spaces or connect them with existing entities. In its entirety, the concept space represents the user and model him/her; the interaction with the system is stored in it, generating new dimensions or updating the existing ones. Each dimension groups the resources labelled by the same tag and specifies a user interest, while sets of dimensions are used to identify his/her goals and perspectives. Specialized classes of agents manage the user model and calculate personalized recommendations, as described in the next Section.

4 Recommendations in SharingPapers

An important feature of the zz-structures is the intrinsic simplicity to contextualize information and to retrieve all documents and info related to a given resource, starting from the resource itself. On this feature is based our collaborative approach for recommendations: starting from the set of tags (that is, dimensions), that identify the current user's interests, we apply a four steps process: (1) expanding the set of tags for similarity; (2) comparing the collections of documents, associated to the set of tags; (3) ordering similar collections, assigning them a score of similarity; (4) ordering similar papers, assigning them a score of similarity. Each step enables the system to provide intermediate specific types of recommendation: (1) new tags for selected resources; (2) new similar users; (3) new collections of resources; (4) new specific resources.

In order to simplify our discussion, we identify with t_i a generic topic (tag or set of tags), and with D_u^i the dimension related to the user u, containing only the papers tagged also with t_i. Here, we propose the application of the recommendation mechanism to a specific user dimension D_u^k.

(1) Expanding the set of similar tags. In order to obtain a high recall, we are interested to find tags similar to the starting tag t_k; for this reason, we apply a non-adaptive reasoning for stating tag similarity considering the frequency of association to a certain paper.
Let $w^k(p)$ be the number of times that t_k has been associated to the paper p:

$$w^k(p) = \sum_{u' \in U} w^k_{u'}(p) \text{ where } w^k_{u'}(p) = \begin{cases} 1 \text{ if } deg^k_{u'}(p) \neq 0 \\ 0 \text{ otherwise} \end{cases}$$

$w^k(p)$ is expressed in terms of the number of times that t_k has been associated to the paper p from each generic user u' (that is, $w^k_{u'}(p)$); in particular, $deg^k_{u'}(p) \neq 0$ indicates that the paper p has been tagged with t_k in the concept space of user u'.

Now, we consider a set P of papers, and the vector $\bar{w}^i = (w^i(p_1), \ldots, w^i(p_N))$ if $N = |P|$, specified for the generic tag t_i.

In order to measure the similarity between a chosen tag t_k, and another generic t_j, we apply the cosine similarity on related vectors $\bar{w}^k$ and $\bar{w}^j$.

$$tag\_sim(t_k, t_j) = cos(\bar{w}^k, \bar{w}^j) = \frac{\bar{w}^k \cdot \bar{w}^j}{\|\bar{w}^k\| * \|\bar{w}^j\|}$$

This measure allows us to assign a score of similarity to each $t_j \in T$ in respect to t_k. So, we consider top scored tags, T^k, as the most similar tags to the input t_k.
(2) Comparing user dimensions. As second step we compare the dimensions labelled by tags in T^k, evaluating the number of resources that they share; in fact, as stated from traditional collaborative techniques, if two users share a lot of resources (in our system, if their concept spaces contain a common set of resources), there is a greater probability that they have a common information need. The Jaccard similarity coefficient is applied as user similarity metric, $\forall t_j \in T^k, \forall u' \in U$:

$$user\_sim(D^k_u, D^j_{u'}) = \frac{\left| V^k_u \cap V^j_{u'} \right|}{\left| V^k_u \cup V^j_{u'} \right|}$$

This metric compares the dimension of interest for u (that is, D^k_u) with the dimensions of other users and allows us to assign them a score of similarity.
(3) Ordering dimensions. For obtaining an order, which considers both tag and user similarities, we define, $\forall t_j \in T$, $\forall u' \in U$, the following metric:

$$score^{t_k}_u(t_j, u') = tag\_sim(t_k, t_j) * (user\_sim(D^k_u, D^j_{u'}) + 1)$$

This value can be used for suggesting, to the user u, personalized navigation paths on dimensions defined from other users.

(4) Ordering papers. Finally, we associate a score to each paper present in the chosen dimensions:

$$score_u^{t_k}(p) = \sum_{\forall t_j : deg_{u'}^{j}(p) \neq 0 \quad \forall u' \in U} score_u^{t_k}(t_j, u')$$

Top scored resources are suggested.

5 Conclusion

Web 2.0 users share a huge size of user generated content and assign them tags for simplify new searches, but current systems do not provide users with tools for organizing own concept spaces, allowing only a flat organization of them. This paper proposed a concept model focused on a dynamic and flexible organization of user concept spaces, and an adaptive and customized recommendation mechanism. Implementation is currently ongoing and experimental evaluation is planned for the next future.

References

1. Bibsonomy, `http://www.bibsonomy.org/`
2. Citeseer, `http://citeseer.ist.psu.edu/`
3. Dattolo, A., Luccio, F.: A new concept map model for e-learning environments. In: Cordeiro, J., et al. (eds.) WEBIST 2008. LNBIP, vol. 18, pp. 404–417. Springer, Heidelberg (2009)
4. Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
5. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)
6. Burke, R.: Hybrid web recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 377–408. Springer, Heidelberg (2007)
7. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking (2006)
8. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in collaborative tagging systems using hierarchical clustering. In: Proc. of the 2nd Int. Conf. on Recommender Systems, Lausanne, Switzerland (2008)
9. Zanardi, V., Capra, L.: Social ranking: Finding relevant content in web 2.0. In: Proc. of the 2nd ACM Int. Conf. on Recommender Systems, Lausanne, Switzerland (2008)
10. Nelson, T.H.: A cosmology for a different computer universe. Jodi 5, 298 (2004)
11. Minio, M., Tasso, C.: User modeling for information filtering on internet services: Exploiting an extended version of the umt shell. In: UM for Information Filtering on the WWW, 5th UM Int. Conf. (1996)
12. Dattolo, A., Luccio, F.: Visualizing personalized views in virtual museum tours. In: Proc. of the Int. Conf. on Human System Interaction, Krakow, Poland, pp. 339–346 (2008)