Integrating Social Network Data for Empowering Collaborative Systems

Xuan Truong Vu, Marie-Hélène Abel, and Pierre Morizet-Mahoudeaux

Abstract. Over the past years, online social networks with websites such as Facebook, Twitter or LinkedIn, have become a very important part of our everyday life. These websites are increasingly used for creating, publishing and sharing information by users. This creates a huge amount of information a part of which may match the interests of a given group. However the distributed and protected nature of these information make it difficult for retrieving. In this paper, we present a user-centered approach for aggregating social data of members of a group to promote the collaboration and the sharing of knowledge inside collaborative systems. The members will be able to delegate the proposed system to aggregate their different social profiles and to make available the relevant part of information to other members of the group.

1 Introduction

Online social networks, or more commonly known as social networks, have become very popular with websites such as Facebook, Twitter and LinkedIn. These websites provide constantly open and evolving socio-technical ecosystems which promote the interactions and the inter-relationships between distributed users - both individuals and organisations, and facilitate the creation and the sharing of cross-domain information and knowledge.

A huge amount of information is therefore available and increasingly grows every day. A part of these information may match the scope of interest of a given group which could be a professional group, a community of interest or any other community. It is interesting to gather and share such information within the group.

Xuan Truong Vu · Marie-Hélène Abel · Pierre Morizet-Mahoudeaux
UMR CNRS 7253 Heudiasyc, Université de Technologie de Compiègne, France
e-mail: {xuan.vu,marie-helene.abel,
pierre.morizet-mahoudeaux}@utc.fr

V.-N. Huynh et al. (eds.), *Knowledge and Systems Engineering, Volume 2*,
Advances in Intelligent Systems and Computing 245,
DOI: 10.1007/978-3-319-02821-7_11,

Commercial solutions such as GNIP[1] allow an access to real-time social media data streams from dozens of social websites via one single API. However, they are addressed to the enterprises for divers business processes (i.e. Business Intelligence, Community Management, etc.) not for the collaboration and the sharing of knowledge.

Today social networks had not been initially designed to be interoperable together. This creates disconnected websites and subsequently isolated data silos. Moreover, most of social networks require the provision of authentications to access to their user data. Such disconnected and protected nature of social networks make it difficult and time-consuming if one want to manually browse one by one in order to retrieve relevant information.

We present, in this paper, the idea of using collaborative peer-sourcing to integrate data available on social networks into collaborative systems with the objective to enhance the sharing of knowledge. Information to gather is not from all over the social networks but had been published by the members of the group. Moreover, it is intended that only information relevant to his/her respective owner' private preferences and the group' common interests will be capitalized and available to other members.

The paper is organized as follows. In the next section, we present some of the main beneficial features of such a social and collaborative approach. Then we introduce an extensible system architecture designed for aggregating the users' social data and for filtering collected data. An illustrative case is also provided in the fourth section to show the applicability of our proposed system. In the fifth section, we discuss about some related works and our originality. Finally, we conclude and present our future works.

2 Motivation

Users publish and exchange different types of information on social networks such as user profile information ranging from demographic information to personal interests, relationships between users, *user-created contents* including photos, videos, statuses, bookmarks, blogs, etc. [8]. All of them form so-called *social data*. These social data are therefore very rich and frequently updated.

It is interesting to consider social data as a very useful source of cross-domain information. A part of them could be relevant to a group which is driven by some common interests. If each of its members agrees to share their social data with other members, then more knowledge can be reachable and even new knowledge may emerge. Let consider the following possibilities :

1. *Members' additional interests/expertise* : when joining a group, the person only declares a part of his/her profile including his/her interests and/or expertise. However, these information are evolving and changing over time. Social data can be an alternative source for updating and enriching one's profile by allowing

[1] `http://gnip.com/`

to uncover his/her last interests and/or expertise [12, 6]. Therefore, the group will learn more about each of its members.
2. *New web resources* : social networks are intensively used for publishing and spreading news and Web resources. For example, a significant part of tweets, *i.e.* short messages published by Twitter users, can be considered as *information sharing* [11]. Most of them contain *URLs* referring to web pages, thus allowing to discover new resources matching the interests of the group.
3. *Emerging topics* : by watching recently captured members' additional interests and new Web resources, emerging topics could be identified.
4. *Possible sub-groups* : Each member can be connected to some other members on one or several social networks. These relationships and their interaction degree [16] will provide extra indicators beside the similarity indicators [7] for efficiently locating some possible sub-groups.
5. *Extended membership* : some external users, not actually belong to the group, might be considered as extended memberships when they are in the lists of contacts of several members of the group. These people could be invited to join the group or given a certain truth if other members would like to reach them for information.

These identified benefits, among others, provide the motivation for the members of a given group to collaboratively share their social data to other members. However, it is obviously impossible to ask to each of them to duplicate information considered pertinent that his/her has already published in other social networks. Thus, to help users avoid making extra manual efforts, an automated or semi-automated process is needed for aggregating their social data and subsequently for filtering the relevant part of gathered information.

As people are the key components of our approach, we must keep in mind that one of the main tasks is to provide them with the entire control over the information that they will possibly share with others. Therefore, we have adopted a user-centered approach which allows each member to delegate the system by suitable permissions to aggregate his/her different social profiles. He/she is also able to set a variety of personal private settings which will be taken in consideration during the information filtering process.

3 The Proposed System

In this section, we describe a general and extensive system architecture for aggregating the social data of the members of a given group and for filtering out the relevant part of information with respect to the personal private setting of each member as well as to the common interests of the group. Our proposed system is made up of three main modules (Figure 1) : (i) a user data integration module, (ii) an information filtering module, and (iii) a collaborative platform. The first module replies on different social networks to gather members' social data and then output preprocessed data that the second module should filter out by the means of a number

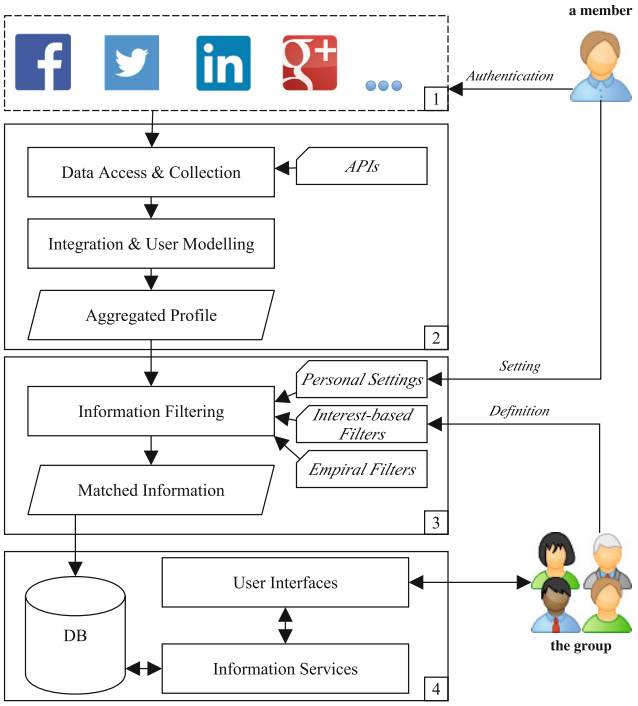

Fig. 1 System architecture for integrating users' aggregated profiles into collaborative systems : (1) social networks (2) user-centered social network aggregation (3) information filtering (4) collaborative knowledge management system

of techniques. Finally, the collaborative platform provides the group with a digital support so that they can access to and share capitalized knowledge.

3.1 Social Networks

There are recently a lot of social websites[2]. Amongst others, Facebook[3], Twitter[4] and LinkedIn[5] are the three most popular websites in terms of the number of users, the volume of daily generated data and the traffic.

Each of them have been trying to provide users with different services and user experiences. A public social networking service like Facebook allows anyone aged 13 and over to register. They come and use Facebook as a communication platform

[2] http://en.wikipedia.org/wiki/List_of_social_networking_websites
[3] https://www.facebook.com
[4] https://www.twitter.com
[5] https://www.linkedin.com

to connect with others and to join common-interest user groups [4]. LinkedIn is also a social networking website but since it is mainly devoted to people in professional occupations. Unlike Facebook and LinkedIn, Twitter is rather a microblogging service that enables users to send messages, known as "tweets", about any topic within the 140-character limit and to follow others for receiving their tweets. Twitter has therefore become a new and powerful medium of information sharing [9].

Social networks all allow users to set up visible profiles and to link to other individuals' profiles. User profile is a unique page where one can type oneself and display an articulated list of friends. This page might also include frames, where different kinds of information may appear such as user-created contents (i.e. posts, statuses, tags, messages, etc.).

By investigating the coverage of user profiles handled by these three services and those from Google+[6], OpenSocial[7] [17], we have identified the most frequent user data and categorized into six dimensions as follows :

- *Personal Characteristics* includes the user's basic information such as name, city, gender, and so forth;
- *Friends* includes the user' social connections;
- *Interests* contains user interests, preferences, and expertise;
- *Groups* contains the user's memberships:
- *Studies* and *Works* describe the user's schools and the user's workplaces respectively.
- *User-created contents* includes contents, social activities produced by the user.

These dimensions also form our general social user model.

3.2 User Data Integration

The *user data integration* is the process dealing fist with accessing and collecting each member's social data then with merging various data and modelling them into a unified profile.

3.2.1 Data Access and Collection

Most of actual social networks make it possible for their users to grant selected third-party applications an access to user data via their own application programming interface (API). With respect to this policy, the system always asks users for permission to access their profiles on subscribed social networks.

Since each social network support its proprietary API (e.g. Graph API for Facebook, REST API 1.1 for Twitter), different *aggregators* are then needed. Each of them is dedicated to a specific social network and recover information corresponding to the previous user dimensions if available.

[6] `https://plus.google.com/`
[7] `http://opensocial.org/`

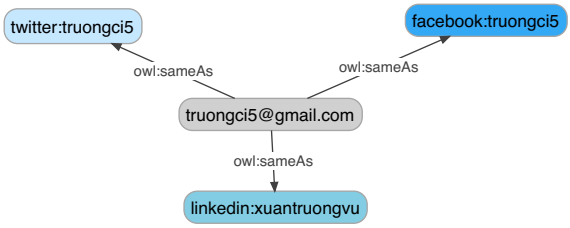

Fig. 2 User aggregated profile representation

Moreover, the aggregators are programmed to regularly (i.e. daily) crawl members' profiles so that their last published information on social networks will be early taken into account for filtering and possibly sharing with others.

3.2.2 Integration and User Modelling

Once members' social data are gathered, they should be merged together. For that purpose, a target common representation is needed. For this purpose, we have used the Friend Of A Friend (FOAF) vocabulary[8].

To map each gather information to a specific FOAF concept we simply base on a hand-crafted set of rules [17]. For example, *friends* will be represented as *foaf:knows* while *interests* can be described by *foaf:topic_interest*.

Each member is identified by an identifier, for example a unique email. His various social accounts will be represented as separated entities and linked to the member by the *owl:sameAs* concept (Figure 2). Each entity therefore has its own attributes and corresponding values. Such a representation allows to preserve in an explicit way the provenance of data (i.e. source) which is an important feature in our system cause it eases users' control over their data and the information sharing.

To improve the members' plain-text profiles, we have followed the *semantic web* approach introduced in [12, 2] matching the concepts contained in profiles to DBpedia[9] resources. Such approach allows to semantically enrich user information and to possibly refer to other linked resources. For example, the resource found at `http://dbpedia.org/resource/Compigne` stands for the Compi'egne city and is linked to the `http://dbpedia.org/resource/Picardy` resource which is the region where the city is located.

DBpedia Spotlight[10] and DBpedia keyword search API[11] are thus used for extracting concepts including named entities from texts end finding related DBpedia resources.

[8] `http://xmlns.com/foaf/spec/`
[9] `http://dbpedia.org`
[10] `https://github.com/dbpedia-spotlight/dbpedia-spotlight`
[11] `https://github.com/dbpedia/lookup`

Regarding user interests, the process is less straightforward. We should extract the user' interests either from a list of things for which users have explicitly claimed their interest or from user-created contents, in particular those containing *URLs*. In the second case, the referred web-pages will be explored to extract the titles and the keyword tags.

3.3 Information Filtering

The *Information Filtering* step is very important. Because, it should decide which information will be ignored or kept according to the member' private settings and the group's common interests. To this end, we have developed different techniques including manual as well as automated methods following :

- *User Private Settings* : the users will be free to set and modify the following settings :
  - The social account(s) the information of which could be aggregated and filtered,
  - The dimension(s) the information of which could be shared,
  - An extra verification might be needed before sharing. Either users can verify which information is shareable one by one, or all matched information will be available to other members.
- *Hashtag Method* : a hashtag is a word or a phrase prefixed with the symbol #, for example, #UTC could stand for the University of Technology of Compiegne. Hashtags have become very popular and efficient means for grouping and retrieving messages related to a given topic on social websites. The users of our system will be encouraged to use their commonly defined hashtags across social networks. Gathered contents including such hashtags will be considered pertinent for the group and directly accessible.
- *Keyword-based Method* : keywords and their synonyms can be also used for matching information. There is a need however for an analysis before constructing the keyword list due to a lot of ambiguities and noises.
- *Ontology-based Method* : in comparison to keywords, ontology gives more powerful performance for matching information. Firstly, there would be much less ambiguities. Secondly, it would not be necessary to list all concepts, named entities in particular, which belong to certain categories. Let's consider an example from DBpedia, in which the *Social Networking Services* category[12] is the subject (i.e. *dcterms:subject*) of a lot of networking services such as *dbpedia:Facebook*, *dbpedia:Twitter*, *dbpedia:Myspace*, *dbpedia:Instagram*, *dbpedia:FOAF_(software)*, etc. In such a case, only the category will be needed in order to match information related to one of its members.
- *Empirical Methods*, additional empiric methods can be also used to reduce the number of information extracted from user-created contents. They actually tend

[12] `dbpedia.org/resource/Category:Social_networking_services`

to filter out personal messages which are self-describing or addressed to a particular person. For this purpose, they rely on some simple detection patterns such as containing emoticons (e.g. “tired and upset :(”) or including other usernames (e.g. “take a look at these photos http://bit.ly/Ywg7p6 @truongci5”).

3.4 Collaborative Platform

The collaborative platform could be any collaborative system available on the market. It should provide among others, two essential functions such as the storage of data and the user interfaces for collaborating (e.g. defining interest-based filters) and accessing to capitalized knowledge.

Since, we have oriented to Semantic Web technologies in our approach, we prefer collaborative systems which reply on or support Semantic Web too. One of the identified candidates is the Memorae[13] platform which has been developed at the Université de Technologies de Compiègne. This platform is an ontology-based collaborative environment easing organizational learning and knowledge capitalization and is recently improved to integrate and index resources from social networks[5].

4 An Illustrative Scenario

In this section, we present a use case example of our system previously proposed. The example is taken for a group composed of several professors and Ph.D. students from the University of Technology of Compiègne.

Being recently interested by the topics like *knowledge management* and *digital ecosystem*, they have been collaborating within the Memorae platform with the objective to share interesting resources related to the two topics. Thought, a certain amount of knowledge have been therefore capitalized thanks to the manual input of each member, they would like to enrich and expand further this common knowledge base.

On the other hand, they are all using online social networks but for divers purposes. They did not necessarily join to the same types of social networks neither put the same efforts for publishing and sharing information. For example, the professors use essentially professional networks such as LinkedIn while the students are much more active on large-scale social networking services such as Facebook and Twitter.

Each of them have been publishing various information and content on his/her profiles handled by different social networks. There is a part of these information matching the two interested topics such as specialised web resources, related scientific events or researchers in the field. Unfortunately, they are not all reachable by all members of the group. Thus, the group wants to integrate these social data into the shared knowledge base.

To this end, the members of the group use our previously proposed aggregation and filtering extension which is supposed to be already operational on the platform

[13] `http://www.hds.utc.fr/memorae/`

Memorae. A hashtag #KE for *knowledge ecosystem* and several keywords such as *knowledge engineering*, *digital ecosystem*, *knowledge ecosystem* and *collaborative platform* have been then defined by the group for filtering relevant information. We suppose moreover that in *private settings*, they all agree to share the entire aggregated data including interests, contacts, published contents, etc.

Now, let consider three specific cases Pierre, Etienne and Xuan. Pierre is a professor while two others are both students. Each of them grants specific permissions to the system for accessing their social profiles and for collecting data. For example, Pierre has given access to his LinkedIn profile and Etienne has authorized access to his Facebook and LinkedIn profiles while Xuan has given access to his LinkedIn and Twitter profiles. During a couple of weeks, they continue normally using their subscribed social networks when the collaborative platform starts to detect some relevant information from the collected data. Here follows several examples.

Xuan has seen a tweet from one of his Twitter contacts concerning *the fifth international conference on knowledge and systems engineering* (KSE2013), then he re-posts this tweet by adding the #KE hashtag. By containing the predefined hashtag, this message will be straightforward selected to share.

Etienne has found out an interesting document about *knowledge ecosystem* available at[14]. He shares this link on his Facebook profile. The system first detects the url and then investigates the referred webpage. Since the page contains *Knowledge ecosystem - Management - Part 1* as its title, it will be retrieved.

The professor Pierre has published on LinkedIn his latest research paper. Since its title includes *digital ecosystem*, this publication will also be considered relevant to the group. Besides, Pierre has some LinkedIn contacts who have been self-qualified as expert in *knowledge engineering*. Such information allows other members to follow these people on social networks so that more interesting resources can be discovered. They could be moreover invited to join the group.

Thus, the shared knowledge base of the group is increased.

5 Discussion

Social networks, by providing rich user data and social interactions, have recently received a wide interest in many areas of Computer Science literature. In particular, many authors have focused on the identification of trust groups [3], *recommender systems* running on multiple social networks [10] and the detection of real-world event using crowd-sourcing [13].

There are some works related to the integration of user profiles in social networks, especially for constructing users' profiles of interests[1, 12], preferences[14] or expertise [16]. The users' output profiles are generally used in the context of recommender systems or social search engines.

The originality of our work is a new use of users' aggregated and enriched profiles in a collaborative way. The members of a given group, supported by our system,

[14] `http://www.slideshare.net/Presentationsat24point0/knowledge-ecosystem-powerpoint-slide`

agree to put together their own unified profiles. Once aggregated and shared, more information is accessible and new knowledge can even emerge. This therefore allows to feed and to enhance the collective intelligence within the same group of users around mutual interests.

In other words, our work consists in contribution to a collaborative knowledge ecosystem by the use of other well-known digital ecosystems such as social networks. To our knowledge, this is the first time that such an social and collaborative approach has been proposed.

Another interesting work that we could take into consideration has been introduced in [15]. The authors have defined a set of requirements, sketched a security model and presented a framework of cryptographic protocols for securing friendship requests and user-generated content within groups formed outside of the social network, around secret, sensitive or private topics.

6 Conclusion and Future Work

In this paper, we have presented a new approach for using social network data to empower collaborative systems. We have introduced some interesting benefits that can motivate each member of a given group to collaborate in sharing their social data. A user-centered approach is recommended for enabling the access to the members' cross-social-network data and for allowing them to have a better control on their information to share with other member of the group. Therefore, we have described a general and extensive system architecture for aggregating and integrating member' social profiles into collaborative systems. A variety of methods for matching and filtering information have been also developed.

On-going work will mainly focuses on the implementation of the social extension on a concrete collaborative system, namely Memorae. It will then be possible to investigate the willingness of users, the potential of social data and to subsequently to evaluate actual benefits of our proposed approach for real groups of end-users whose shared divers interests.

Acknowledgments. Part of this work has been developed in cooperation with the 50A Company[15] who is funding this work.

References

1. Abel, F., Gao, Q., Houben, G.-J., Tao, K.: Analyzing Temporal Dynamics in Twitter Profiles for Personalized Recommendations in the Social Web. In: Proceedings of ACM WebSci 2011, 3rd International Conference on Web Science, Koblenz, Germany. ACM (June 2011)
2. Abel, F., Henze, N., Herder, E., Krause, D.: Linkage, aggregation, alignment and enrichment of public user profiles with mypes. In: Proceedings of the 6th International Conference on Semantic Systems, I-SEMANTICS 2010, pp. 11:1–11:8. ACM, New York (2010)

[15] `http://www.50a.fr`

3. Caverlee, J., Liu, L., Webb, S.: The socialtrust framework for trusted social information management: Architecture and algorithms. Inf. Sci. 180(1), 95–112 (2010)
4. Chiu, P.-Y., Cheung, C.M.K., Lee, M.K.O.: Online social networks: Why do we use facebook? In: Lytras, M.D., Carroll, J.M., Damiani, E., Tennyson, R.D., Avison, D., Vossen, G., De Pablos, P.O. (eds.) WSKS 2008. CCIS, vol. 19, pp. 67–74. Springer, Heidelberg (2008)
5. Deparis, E., Abel, M.-H., Mattioli, J.: Modeling a social collaborative platform with standard ontologies. In: 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, pp. 167–173 (2011)
6. Gao, Q., Abel, F., Houben, G.-J.: Genius: generic user modeling library for the social semantic web. In: Pan, J.Z., Chen, H., Kim, H.-G., Li, J., Wu, Z., Horrocks, I., Mizoguchi, R., Wu, Z. (eds.) JIST 2011. LNCS, vol. 7185, pp. 160–175. Springer, Heidelberg (2012)
7. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 431–440. ACM, New York (2010)
8. Kim, W., Jeong, O.-R., Lee, S.-W.: On social web sites. Inf. Syst. 35(2), 215–236 (2010)
9. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 591–600. ACM, New York (2010)
10. De Meo, P., Nocera, A., Terracina, G., Ursino, D.: Recommendation of similar users, resources and social networks in a social internetworking scenario. Information Sciences 181(7), 1285–1305 (2011)
11. Naaman, M., Boase, J., Lai, C.-H.: Is it really about me?: message content in social awareness streams. In: Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, CSCW 2010, pp. 189–192. ACM, New York (2010)
12. Orlandi, F., Breslin, J., Passant, A.: Aggregated, interoperable and multi-domain user profiles for the social web. In: Proceedings of the 8th International Conference on Semantic Systems, I-SEMANTICS 2012, pp. 41–48. ACM, New York (2012)
13. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 851–860. ACM, New York (2010)
14. Shapira, B., Rokach, L., Freilikhman, S.: Facebook single and cross domain data for recommendation systems. User Modeling and User-Adapted Interaction 23(2-3), 211–247 (2013)
15. Sorniotti, A., Molva, R.: Secret interest groups (sigs) in social networks with an implementation on facebook. In: Proceedings of the 2010 ACM Symposium on Applied Computing, SAC 2010, pp. 621–628. ACM, New York (2010)
16. Vu, T., Baid, A.: Ask, don't search: A social help engine for online social network mobile users. In: 2012 35th IEEE Sarnoff Symposium (SARNOFF), pp. 1–5 (2012)
17. Vu, X.T., Morizet-Mahoudeaux, P., Abel, M.-H.: User-centered social network profiles integration. In: Proceedings of the 9th International Conference on Web Information Systems and Technologies, Aachen, Germany (2013)