Computers in Human Behavior xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

A user-centered and group-based approach for social data filtering and sharing

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

UMR CNRS 7253 Heudiasyc, Unviersity of Technology of Compiegne, France

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Social networking sites
Social Data Aggregation
Information filtering
Information sharing
Content management
Collaboration

ABSTRACT

Social networking sites (SNSs) like Facebook, Google+, Twitter, LinkedIn have become a very important part of our daily life. People are connected to multiple SNSs for networking, communicating, collaborating, sharing and seeking for information. Although, the diversity of current SNSs increases and enriches our online experience, they cause some problems. One of the major issues is that users are often overwhelmed by the huge number of social data. It is even worse as these social data are scattered across disconnected SNSs. To address such problems, we propose a user-centered and group-based approach for social data filtering and sharing. First, it allows users to aggregate their social data from different SNSs and to extract relevant contents. Users explicitly define their interests via specific queries, using information filtering techniques, the system will retrieve new corresponding contents. Second, it is expected to extend its first user-centered purpose by allowing group-based information sharing and management. Users can share some part of their own social data with and collectively define the information organization within their respective groups. To describe further and illustrate our proposed approach, a system architecture and a prototype are also presented in this paper. A primary test was carried out and showed encouraging results confirming the added values of our approach.

1. Introduction

With social media sites, social networking sites (SNSs) have made up today social Web sites (Kim, Jeong, & Lee, 2010) and become a very important part of our daily life. People spend more time in SNSs than ever for three main uses: (i) gathering, sharing information and contents, (ii) keeping in touch with family, friends, colleagues, and (iii) finding, making new friends (Bonds-Raacke & Raacke, 2010).

There are a large number of SNSs available (Solis, 2013). Facebook, Google+, Twitter and LinkedIn are some of the most successful examples. By attracting millions of active users around the world, they occupy a central place in the social media landscape (Cavazza, 2014). Each of them provides users with its different and unique features: Facebook allows third party applications to build on its application programming interfaces (APIs), Google+ embeds other Google's services, Twitter offers real-time micro-blogging, and LinkedIn focuses on professional networking. Thus, it is very common that one user is simultaneously connected to several SNSs.

Despite such growing popularity and a wide range of provided services, current SNSs raise some issues associated with users' social lifestyle and interaction such as privacy, identity theft, addiction, and spread of bad information (Kwak & Lee, 2011). While these issues are very important in SNSs, they are however not in the scope of our work. We are more concerned with issues related to the usage experience of SNSs, two of which have been particularly the focus of our work: (i) *Information Overload*, and (ii) *"Walled garden" Problem*.

Information Overload: Users are increasingly facing with information overload in SNSs (Borgs, Chayes, Karrer, & Meeder, 2010). They are often overwhelmed by the huge number of coming information. For example, a normal user has, on average 338 friends on Facebook (Smith, 2014) and 208 followers on Twitter (Smith, 2014). As such, he/she receives per day hundreds of diverse *social data* which can be posts, status, updates, photos, videos, links, tags, check-ins, etc., via these social connections. This is much beyond what users could process manually. Moreover, most recent SNSs put all social data generated by a user's social connections into a single stream (e.g. Facebook New Feed, Twitter User Timeline) and sort them in a chronological order. New social data are constantly appearing within and flooding these streams. As a result, many important and interesting pieces of information remain unnoticed by the user, whereas lots of irrelevant and not worth reading contents keep showing up.

"Walled Garden" Problem: Current SNSs, namely Facebook, Google+, Twitter, Linkedin, all operate as "walled gardens", where user data and generated contents are exclusive to the SNSs Chisari

* Corresponding author.
E-mail address: xuan.vu@hds.utc.fr (X.T. Vu).

http://dx.doi.org/10.1016/j.chb.2014.11.079
0747-5632/

(2009). Most importantly, they provide very little interaction with each other. Users are always required to create a new profile when joining a SNS. Then, they have to manage their social data as well as to keep track of all recent updates across SNSs. On the one hand, it worsens the problem of information overload as social data are scattered over different SNSs. On the other hand, the users of a given SNS are not able to interact and share interesting information or useful contents with the users of another SNS.

To address both the problems, we propose a user-centered and group-based approach for social data filtering and sharing. First, this approach enables the aggregation of social data from different SNSs and the extraction of relevant contents. The users' social data are constantly retrieved and gathered from different SNSs by means of dedicated programs previously granted by the users. Every new social data is then processed, enriched and indexed before being filtered and selected according to the users' interests, which have been explicitly defined beforehand by the users by using specific queries. The users can therefore quickly access and view the contents matching a particular topic by selecting it.

Beyond such a user-centered objective, the approach is expected to also allow group-based information sharing and management. To be more practical, the users can join one or several groups driven by common interests or topics where they can share some part of their social data. Conversely, the users can reach more appropriate contents. In particular, each member of a given group is encouraged to contribute their own intelligence to collectively define and improve the group's evolving areas of interest.

Besides these two principal contributions, we also present, in this paper, an extensible system architecture for implementing our approach. The system is composed of a number of specific modules, which are easy to improve and extend.

The rest of the paper is organized as follows: In the next section, we present a summary of related work on Social Network Aggregation, information filtering, and content curation and sharing. Then, we introduce our user-centered and group-based approach for social data filtering and sharing. After, we describe the important modules for the implementation of the system. In Section 5, we present our web-based prototype and discuss some encouraging results of the first test using this prototype. Finally, we summarize the contributions of our work.

2. Related work

Our work attempts to implement a new emergent paradigm called *Social Internetworking System* (SIS), where a SNS can be seen as a part of a more complex system comprising many users, social networks and resources (Meo, Nocera, Terracina, Ursino, & De Meo, 2011). SIS has raised more and more interests from researchers for enabling strategic applications whose main strength is the integration of different communities that nevertheless preserves their diversity and autonomy (Buccafurri, Lax, Nocera, & Ursino). While the Social Network Analysis can play a very important role in the SIS scenario (Fan & Gordon, 2014), we are more interested in information management approaches that can fit both personalized and collective uses. The approach presented in this paper involves three different fields such as *Social Data Aggregation*, *information filtering*, and *content curation and sharing*. We review below some representative works related to these fields. Then, we discuss about their respective limitations, and outline the added values of our proposed approach.

2.1. Social Network Aggregation

Social Network Aggregation is the process of collecting, aggregating and organizing data spread across multiple SNSs. Such process could be brand-oriented as well as user-oriented. The former allows brands to track and capture as much as possible public messages and comments concerning their reputation (Fire, Puzis, & Elovici, 2013; Gao, Wang, Luan, & Chua, 2014), while the latter attempts to organize a user's social networking experience as a whole.

Here, we solely discuss about the user-oriented aggregation, the first challenge of which is to identify unique users across SNSs. *People search engines* such as Peekyou,[1] Pipl[2] allow to search the different social profiles of a user based on public personal attributes (e.g. name, username, email or location). However, a user may set different values to these attributes, even leave them undefined that makes user identification incomplete. Google proposed an alternative, named *Social Graph API*,[3] which crawls users' personal web pages (e.g. blogs) and extracts links referring to their social profiles. Unfortunately, it is no longer available. In this work, we do not deal with this non-trivial problem as we directly ask users for indicating their different social identifiers.

Another important requirement for integrating social data from different SNSs is to define a common model for the representation of social data. A number of light-weight ontologies have therefore been developed. Some of them such as Friend Of A Friend (FOAF),[4] Semantically-Interlinked Online Community (SIOC),[5] Weighted Interests (WI),[6] Open Provenance Model (OPM)[7] and Activity Streams[8] have already been widely adopted. The authors of Kapanipathi and Orlandi (2011) and Orlandi, Breslin, and Passant (2012) showed that these vocabularies could be combined and used as a domain ontology for integrating social data from many SNSs. Furthermore, they proposed to use linked open data (e.g. DBpedia,[9] OpenCalais[10]) in order to semantically enrich aggregated data. Although, it could be very interesting to follow such a Semantic Web principle, we describe, in the next section, another generic model which fits better our previously cited objectives.

Commercial solutions such as FriendFeed,[11] Hootsuite,[12] Flock[13] attempt to implement Social Network Aggregation, and are so called *Social Network Aggregator* (SNA). They allow to consolidate at a single point the various social activities in such a way that the user is not required to log in each SNS and perform same social activity (Virmani, Pillai, & Juneja, 2014). The user performs a given social activity within a SNA and the information is synchronized to all of the social networks that the user specifies. Each sSNA provides the users with specialized features to integrate the SNSs but none of them tries to integrate the information available within SNSs. Compared to them, our proposed solution allows the users to not only collect data from different SNSs, but also to extract useful information.

2.2. Information filtering

Information filtering deals with the delivery of information that the user is likely to find interesting or useful. An information filtering system assists users by filtering the data source and delivers relevant information to them (Ghorab, Zhou, O'Connor, & Wade). Information filtering is very close with *Information Retrieval* and

[1] http://www.peekyou.com/.
[2] https://pipl.com/.
[3] https://developers.google.com/social-graph/.
[4] http://www.foaf-project.org/.
[5] http://www.sioc-project.org/.
[6] http://smiy.sourceforge.net/wi/spec/weightedinterests.html.
[7] http://openprovenance.org/.
[8] http://activitystrea.ms/.
[9] http://dbpedia.org/.
[10] http://www.opencalais.com/.
[11] http://friendfeed.com/.
[12] https://hootsuite.com/.
[13] http://sourceforge.net/projects/flock/.

Please cite this article in press as: Vu, X. T., et al. A user-centered and group-based approach for social data filtering and sharing. *Computers in Human Behavior* (2014), http://dx.doi.org/10.1016/j.chb.2014.11.079