Information Systems 56 (2016) 1–18

Contents lists available at ScienceDirect

Information Systems

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

Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms

Mohamed Reda Bouadjenek [a,b,*], Hakim Hacid [c], Mokrane Bouzeghoub [d]

[a] INRIA & LIRMM, University of Montpellier, France
[b] Department of Computing and Information Systems, University of Melbourne, Parkville 3010, Victoria, Australia
[c] Zayed University, United Arab Emirates
[d] PRiSM Laboratory, University of Versailles-Saint-Quentin-en-Yvelines (UVSQ), France

ARTICLE INFO

Article history:
Received 18 October 2014
Accepted 9 July 2015
Available online 28 August 2015

Keywords:
Information Retrieval
Social networks
Social Information Retrieval
Social search
Social recommendation

ABSTRACT

There is currently a number of research work performed in the area of bridging the gap between Information Retrieval (IR) and Online Social Networks (OSN). This is mainly done by enhancing the IR process with information coming from social networks, a process called Social Information Retrieval (SIR). The main question one might ask is *What would be the benefits of using social information (no matter whether it is content or structure) into the information retrieval process and how is this currently done?*

With the growing number of efforts towards the combination of IR and social networks, it is necessary to build a clearer picture of the domain and synthesize the efforts in a structured and meaningful way. This paper reviews different efforts in this domain. It intends to provide a clear understanding of the issues as well as a clear structure of the contributions. More precisely, we propose (i) to review some of the most important contributions in this domain to understand the principles of SIR, (ii) a taxonomy to categorize these contributions, and finally, (iii) an analysis of some of these contributions and tools with respect to several criteria, which we believe are crucial to design an effective SIR approach. This paper is expected to serve researchers and practitioners as a reference to help them structuring the domain, position themselves and, ultimately, help them to propose new contributions or improve existing ones.

1. Introduction

With the emergence of the social Web, the Web has evolved from a static Web, where users were only able to consume information, to a Web where users are also able to produce information. This evolution is commonly known as Social Web or Web 2.0. Thus, the Web 2.0 has introduced a new freedom for the user in his relation with the Web by facilitating his interactions with other users who have similar tastes or share similar resources. Social platforms and networks (such as *MySpace*,[1] *Facebook*,[2] and *LinkedIn*[3]), collaborative tagging sites (like *delicious*,[4] *CiteULike*,[5] and *Flickr*[6]), and microblogging sites (like *Twitter*[7] and *Yammer*[8]) are

* Corresponding author. Tel.: +33 6 43 83 76 91.
E-mail addresses: rbouadjenek@gmail.com (M.R. Bouadjenek), hakim.hacid@zu.ac.ae (H. Hacid), mokrane.bouzeghoub@prism.uvsq.fr (M. Bouzeghoub).

http://dx.doi.org/10.1016/j.is.2015.07.008
0306-4379/

[1] https://myspace.com/
[2] https://www.facebook.com
[3] https://www.linkedin.com/
[4] https://delicious.com/
[5] http://www.citeulike.org/
[6] https://www.flickr.com/
[7] https://twitter.com

certainly the most adopted technologies in this new era. These platforms are commonly used as ameans to communicate with other users, exchange messages, share resources (photos and videos), comment news, create and update profiles, interact and play online games, etc. In addition to dedicated social platforms, traditional content providers sites like newspapers, tend to be more social since they provide to users means for sharing, commenting, constructing, and linking documents together [1,2], e.g., through the buttons *"share on social networks"*. This has been also facilitated by initiatives like OpenID[9] and OpenSocial.[10]

These collaborative tasks that make users more active in generating content are among the most important factors for the increasingly growing quantity of available data. In such a context, a crucial problem is to enable users to find relevant information with respect to their interests and needs. This task is commonly referred to as Information Retrieval (IR). IR is performed every day in an obvious way over the Web [3], typically using a search engine. However, classic models of IR do not consider the social dimension of the Web. They model web pages[11] as a mixture of static homogeneous terms generated by the same creators, i.e., the authors of the web pages. Then, the ranking algorithms are often based on (i) a query and document text similarity (e.g., the cosine measure and the Okapi BM25 [4]), and (ii) the existing hypertext links that connect these web pages (e.g., PageRank [5], and HITS [6]).

Therefore, classic models of IR and even the IR paradigm have to be adapted to the socialization of the Web in order to fully leverage the social context that surrounds web pages and users. Indeed, exploiting social information has a number of advantages (for IR in particular): first, feedback information in social networks is provided directly by the user, so users' interests accurate information can be harvested as people actively express their opinions on social platforms. Second, a huge amount of social information is published and available with the agreement of the publishers. Exploiting this information should not violate user privacy, in particular social tagging information, which does not contain sensitive information about users. Finally, social resources are often accessible, as most of social networks provide APIs to access their data (even if often, a monetized contract must be established before any large scale use).

There is currently a number of research works undertaken to improve the IR process with information coming from social networks. This is commonly known as "Social Information Retrieval" (SIR). This paper intends to provide a clear understanding of the various efforts performed in the domain of SIR, and in this perspective, we propose

1. An objective review of some of the most representative research contributions and existing tools in this domain to understand the principles of SIR as they are currently formulated.
2. A taxonomy to categorize these contributions in order to structure this wide domain.
3. An analysis of some of these contributions and tools with respect to several criteria considered as crucial to design an effective SIR approach or to appreciate one.

This study is useful for researchers, engineers, and practitioners in this domain to help in understanding trends, challenges and expectations of a SIR approach. The rest of this paper is organized as follows: in Section 2, we discuss the main concepts used throughout this paper. In Section 3, we introduce our taxonomy of SIR approaches and tools, while describing each of these categories in Sections 4–6. In Section 7, we elaborate on an analysis of some of these contributions and tools based on several dimensions. Finally, we give some future directions in Section 8 and we conclude in Section 9.

2. Background

Information Retrieval is the process of recovering stored information from large datasets to satisfy users' information needs. Salton [7] and Baeza et al. [3] defined IR as follows:

Definition 1 (*Information Retrieval*). Information Retrieval (IR) is the science that deals with the representation, storage, organization of, and access to information items in order to satisfy the user requirements concerning this information.

An IR system is evaluated in its accuracy and ability to retrieve high quality information/documents, which maximize users' satisfaction, i.e. the more the answers correspond to the users' expectations, the better the system is. Information Retrieval is a very well established domain. Several articles and books are devoted to IR, e.g., [3,7]. Since our focus is on the social part of IR, we do not go into more detail in the description of the IR domain.

This section introduces and defines the basic concepts used throughout this paper. In Section 2.1 we define what is a social network, and discuss the main models of social relationships.[12] Then, in Section 2.2, we introduce the notion of Social Information Retrieval that bridges social networks and information retrieval.

2.1. Social networks and social network analysis

Nowadays, social networks are at the heart of the Web 2.0. A social network is defined as follows:

Definition 2 (*Online social network*). An online social network is the social structure, which emerges from human interactions through a networked application.

[8] https://www.yammer.com/

[9] http://www.openid.net/

[10] http://www.opensocial.org/

[11] In this paper, we also refer to web pages as documents.

[12] Note that we do not intend to provide a complete review of the SNA domain, but rather to provide the main underlying principles, which could be leveraged in IR systems and which are helpful to understand the analysis provided in this paper.