Available online at www.sciencedirect.com
SCIENCE DIRECT®
Web Semantics: Science, Services and Agents on the World Wide Web 3 (2005) 211–223
JOURNAL OF Web Semantics
www.elsevier.com/locate/websem

Flink: Semantic Web technology for the extraction and analysis of social networks

Peter Mika*

Department of Computer Science, Vrije Universiteit Amsterdam (VUA), De Boelelaan 1081, 1081HV Amsterdam, The Netherlands

Received 24 May 2005; accepted 25 May 2005

Abstract

We present the Flink system for the extraction, aggregation and visualization of online social networks. Flink employs semantic technology for reasoning with personal information extracted from a number of electronic information sources including web pages, emails, publication archives and FOAF profiles. The acquired knowledge is used for the purposes of social network analysis and for generating a web-based presentation of the community. We demonstrate our novel method to social science based on electronic data using the example of the Semantic Web research community.

Keywords: Semantic Web; Social networks; Ontology extraction; Social ontology

1. Introduction

The possibility to publish and gather personal information (such as the interests, works and opinions of our friends and colleagues) has been a major factor in the success of the web from the beginning. Remarkably, it was only in the year 2003 that the web has become an active space of socialization for the majority of users. That year has seen the rapid emergence of a new breed of web sites, collectively referred to as social networking services (SNS). The first-mover Friendster[1] attracted over 5 million registered users in the span of a few months [13], which was followed by Google and Microsoft starting or announcing similar services.

Although these sites feature much of the same content that appear on personal web pages, they provide a central point of access and bring structure in the process of personal information sharing and online socialization. Following registration, these sites allow users to post a profile with basic information, to invite others to register and to link to the profiles of their friends. The system also makes it possible to visualize and browse the resulting network in order to discover friends in common, friends thought to be lost or potential new friendships based on shared interests. (Thematic sites cater to more specific goals, such as establishing a business contact or finding a romantic relationship.)

* Tel.: +31 20 5987753; fax: +31 20 5987653.
E-mail address: pmika@cs.vu.nl
URL: http://www.cs.vu.nl/pmika
[1] http://www.friendster.com

1570-8268/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.websem.2005.05.006

The latest breed of social networking services combine social networks with the sharing of content such as bookmarks, documents, photos, reviews. The idea of network-based knowledge sharing is based on the sociological theory that social interaction creates similarity and vice versa, interaction creates similarity (friends are likely to have acquired or develop similar interests). Lately, the notion of ratings and social networks-based trust are also investigated as a filtering mechanism in loosely controlled environments.

Despite their early popularity, users have later discovered a number of drawbacks to centralized social networking services. First, the information is under the control of the database owner who has an interest in keeping the information bound to the site. The profiles stored in these systems cannot be exported in machine processable formats, and therefore the data cannot be transferred from one system to the next. (As a result, the data needs to be maintained separately at different services.) Second, centralized systems do not allow users to control the information they provide on their own terms. Although Friendster follow-ups offer several levels of sharing (e.g. public information versus only for friends), users often still find out the hard way that their information was used in ways that were not intended.

These problems have been addressed with the use of Semantic Web technology. The friend-of-a-friend (FOAF) project[2] is a first attempt at a formal, machine processable representation of user profiles and friendship networks. Unlike with Friendster and similar sites, FOAF profiles are created and controlled by the individual user and shared in a distributed fashion.[3] Much like the way web pages are linked to each other by anchors, these profiles link to the profiles of friends by using the *rdfs:seeAlso* relation, creating the so-called FOAF-web.

The alert reader may note that for the purposes described above, namely providing a structured representation of user profiles, the use of XML technologies would have sufficed. In fact, the real value of FOAF is that it represents an agreement on key terms and that it is described in a semantic format (namely, OWL full). These properties make FOAF the ideal basis for the semantic integration of personal information extracted from heterogeneous knowledge sources.[4]

Flink, our system to be introduced is the first to our knowledge that exploits FOAF for the purposes of *social intelligence*. By social intelligence, we mean the semantics-based integration and analysis of social knowledge extracted from electronic sources under diverse ownership or control. In our case, these sources are largely the natural byproducts of the daily work of a community: HTML pages on the web about people and events, emails and publications. From these sources, Flink extracts knowledge about the social networks of the community and consolidates what is learned using a common semantic representation, namely the FOAF ontology.

The raison d'etre of Flink can be summarized in three points. First, Flink is a demonstration of the latest Semantic Web technology (and as such a recipient of the Semantic Web Challenge Award of 2004). In this respect, Flink is interesting to all those who are planning to develop systems using Semantic Web technology for similar or different purposes. Second, Flink is intended as a portal for anyone who is interested to learn about the work of the Semantic Web community, as represented by the profiles, emails, publications and statistics. Hopefully Flink will also contribute to bootstrapping the nascent FOAF-web by allowing the export of the knowledge in FOAF format. This can be taken by the researchers as a starting point in setting up their own profiles, thereby contributing to the portal as well. Lastly, but perhaps most importantly, the data collected by Flink is used for the purposes of social network analysis, in particular learning about the nature of power and innovativeness in scientific communities.

In this paper the focus is on the first two aspects of Flink. We begin with the introduction of the system from a user perspective in Section 2. In Section 3, we describe the architecture of Flink in detail and discuss the lessons that have been learned while developing its components. We briefly introduce the idea of network analysis using Flink in Section 4. Related and future work are discussed in the last two sections of this paper.

[2] http://www.foaf-project.org.

[3] FOAF profiles are typically posted on the personal website of the user and linked from the user's homepage with the HTML LINK tag.

[4] While FOAF carries a necessary level of commitment, the maintainers of ontology are also careful not to overly restrict the interpretation of the ontology in order to keep its wide appeal to different communities and usage scenarios.