Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 111 (2017) 8–16

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Conference on Advances in Information Technology, IAIT2016, 19-22 December 2016, Macau, China

Increasing the expression power of persons' profiles in semantic social networks

Juha Puustjärvi[a,]*, Leena Puustjärvi[b]

[a]*Department of Computer Science, University of Helsinki, P.O. Box 68, Helsinki, Finland*
[b]*The Pharmacy of Kaivopuisto, Neitsytpolku 10, Helsinki, Finland*

Abstract

The amount of social information published in the Web has dramatically increased in the past years. FOAF, Dublin Core, and vCard are examples of popular vocabularies that are used to present computer understandable profiles of persons in social media. These profiles describe facts about a person such as his or her interests, and thereby we can easily find persons having the same interests. However, in reality we are rather interested in about the closeness of persons' interests than whether they have precisely the same interests. Such aspects are easily understandable for humans but they are not machine understandable without additional semantics. Therefore we have extended the vocabularies of social media by domain specific taxonomies. In addition, to increase the expression power of persons' profiles we have also attached weights to persons' interests. We have also consider a variety of ways for expressing and computing the closeness of weighted profiles.

Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology

Keywords: Semantic social network; Semantic web; FOAF; Ontologies; OWL; RDF ;

1. Introduction

A social network is a collection of individuals linked together by a set of relations[1]. A semantic social network is the result of the application of Semantic Web technologies to social networks[2]. For example, several vocabularies

* Corresponding author. Tel.: 358452756720.
E-mail address: juha.puustjarvi@cs.helsinki.fi

1877-0509
Peer-review under responsibility of the organizing committee of the 8th International Conference on Advances in Information Technology
10.1016/j.procs.2017.06.003

such as FOAF (friend of a friend)[3], Dublin Core [4], and vCard[5], are used in social networks. These vocabularies can be used in describing facts about a person such as his or her work place, job title and interests.

FOAF is a machine-readable vocabulary describing persons, their activities and their relations to other people and objects. Anybody can use it to describe him- or herself. It is expressed using the Resource Description Framework (RDF)[6] and the Web Ontology Language (OWL) [7]. Computers may use these FOAF profiles to find, for example, all people interested in tennis, or to find out all people both you and a friend of yours know.

Technically, FOAF vocabulary (ontology)[8] is comprised of things and links. Links are called properties and the types of things are called classes. Hence, FOAF vocabulary is comprised of terms, each of which is either a class or a property.

Class foaf:Person is one of the most used class of the FOAF vocabulary. A set of properties, such as gender, birthday, interest and topic_interest, can be attached to the instances of this class. Hence, we might claim that a person, say John, has an interest in RDF by saying that "John stand in an interest relationship to RDF". Likewise we might claim that another person, say Mary, is interested in Semantic Web.

John might be interested in whether he and other persons, such as Mary, do have interests in common? Individuals familiarized on these topics, such as John and Mary, understand that they do have some interests in common as John is interested in RDF and Mary is interested in Semantic Web. However, computer cannot make such a conclusion without machine interpretable knowledge about the dependency of the terms RDF and Semantic Web. Hence, machine understandable data is of prime importance as the amount of data in social media is increasing all the time.

Using the FOAF property topic_interest John might also state that he has an interest in classic blues and punk. However, he cannot state that he is extremely interested in classic blues and has minor interest in punk, as such expressions would require some kind of weighting of interests. Yet, ignoring weighting considerably restricts the expression power of persons' profiles.

In this paper, we will restrict ourselves on the above mentioned inconveniences:

- First, we deal with taxonomy-based semantic similarity, i.e., how the semantics of taxonomies can be exploited in computing the closeness of persons' interests. In particular, we adapt the methods of genealogy for computing the closeness of persons' profiles.
- Second, we consider how weighting can be used in making profiles more specific, and how the closeness of such profiles can be computed. First we extend the taxonomy-based semantic similarity by weighting, meaning that a person may attach weights low, medium or high to his or her interests. Then, we consider the usability of the Vector Model[9] for computing the closeness of persons' interests. The key idea behind this model is that persons' interests as well as queries can be presented as vectors. Hence, we can process a query by computing the distance of the query vector and the profile vectors.

The rest of the paper is organized as follows. In Section 2, we first consider the structure and semantics of taxonomies, and then we present how taxonomies can be used in extending the expression power of the FOAF vocabulary. In Section 3, we give an example of a FOAF file, and illustrate how FOAF files are queried by SPARQL[10]. Taxonomy-based semantic similarity is the topic of Section 4. Then, in Section 5, we deal with profile weighting in the context of taxonomy-based similarity, and in the context of the Vector Model. We also shortly consider the matching algorithms that are suitable for the Vector Model. Finally, Section 6 concludes the paper by discussing our future research.

2. Taxonomies

2.1. The structure and semantics of taxonomies

One of the main challenges faced by information retrieval systems is to efficiently manage the large number of documents they hold. Such systems make it easier to give users access to relevant resources that satisfy their information needs. Information retrieval systems are based on an information retrieval model. They determine the ways the data of the resources are presented, and the way the information needs (queries) are presented.