Towards Trust in Web Content Using Semantic Web Technologies

Qi Gao

TU Delft, Web Information Systems, EWI/ST/WIS,
P.O. Box 5031, 2600 GA Delft, The Netherlands
q.gao@tudelft.nl

Abstract. Since the amount of user-generated content has been sharply increasing in recent years, mainly due to Web 2.0 technology and effects of social networking, it is necessary to build mechanisms to assess the reliability of the content. On the web this notion of trust is a key ingredient for an effective manipulation of knowledge on a (world-wide) web scale. The web of trust has thus become an important research area both for web science and semantic web. In the PhD research we have laid out for us, we focus on the notion of trust and methods for representing and computing trust of users in the web content. This paper outlines the vision at the start of the PhD period on the research problem and the semantic web-based approach to solve that problem.

Keywords: trust, web science, semantic web.

1 Problem Statement and Motivation

The volume of web content is sharply increasing in the recent years mainly because of Web 2.0 technology and applications for social networking, e.g. discussion boards in online communities, social network profiles, videos, photos, reviews and ratings of products, hotels, restaurants.

With the enormous and still growing amount of content on the Web, it becomes more and more necessary for users to be able to assess the trustworthiness of content on the Web. For example, it will help for information retrieval systems or recommender systems to find the most relevant and reliable content that matches the user's preference. Building a trust mechanism for the Web is also helpful to make the interaction between users and content more reliable and controllable.

Dealing with trust on the Web is a difficult and complex task, since it involves a wide range of factors such as user history, user preferences, demographical information and the context in which the trust relationship is built. These factors have very different characteristics and it is a challenge to measure these factors. So, it is necessary to create an approach for modeling and computing the trustworthiness of content on the Web for individual user.

Ideally, there would be a trust model that tells for each user and each content element how trustworthy that content element is for that user. A giant barrier for obtaining such a trust model is that if we would have obtained accurately the trustworthiness

for a given set of users and content elements, it is very hard to predict the trustworthiness for new content.

Instead of focusing on the content, many approaches therefore choose to build the notion of trust and trust metrics within networks of people, agents or peers that have trust relationships between them [3]. Such trust networks are often obtained by taking explicit trust relationships between pairs of people firstly and then propagating the trust through the network. In a social environment, such a trust network can easily be acquired from social networks of friends or connected people. The trustworthiness of content that is unknown for a given user is then predicted in an indirect way by utilizing the trust relationships for that user to other people.

A major shortcoming with these approaches is that they cannot adequately deal with the semantics of the content and so we observe that the current approaches are limited to work only in a specific context or domain. For example, the trust relationship predicted using a friends network acquired from a website for books does not work effectively in the domain of movies, even for the same people.

Since it is now feasible to know more of the content's semantics, exploiting explicit and implicit representations with semantic web technology and using access to linked open data, we see the opportunity to build a better global trust metric. In our approach, each user maintains the trustworthiness for a set of content elements. Instead of propagating this through the network of users, we directly predict the trustworthiness of new, unknown content by analyzing the semantic relationship between these two pieces of content. So, by using a more explicit representation of what we know about the concept and their relationships, we make it possible for the users to get a better grip on their trust in content and not only in the people behind the content.

This PhD work will propose a model for trust of users in Web content in order to make the user-oriented exploitation of data on the Web more reliable and controllable. In the model we will represent trust with semantic standards and utilize the analysis of semantic relations between content elements for the trust computation and inference.

2   State of the Art

In this section, we briefly discuss the state-of-the-art of approaches that are related to our work. The Web of trust [4] has become an important area both for academia and industry. Many trust metrics have been developed. Each approach stresses various characteristics of trust. Artz and Gil [1] demonstrated a comprehensive overview of existing trust metrics in computer science and semantic web. Due to the lack of space we focus on two tasks that are most relevant to our problem and motivation.

2.1   Representing Trust

One category of approaches to representing trust on the Web is based on the concept of a Web of Trust that is a network of people, agents or peers with trust relationships. Golbeck [3] represented trust within web-based social networks by defining the functional properties of trust. In Ray's paper [12], trust is specified as a relationship between a truster and a trustee at a particular time instance and for a particular context. Heath and Motta [8] provided a way to collect reviews and ratings data from distributed sources and generated a metric to represent trust relationships, which is used as

an input to a Web-based system with the purpose of information seeking and recommendation. These approaches focus on the trust among users on the Web.

In contrast to these approaches, some work has been done related to trust in the content itself. Gil et al. [2] used the term of "content trust" to acquire trust in the content provided by a web resource. Rich factors are included in their model. However, their work did not provide a mechanism to deal with the semantic of the content. Groth et al. [5] proposed a content-based trust assessment in electronic contracts based on similarity of content in different contracts. Hartig et al. [7] presented a trust model for RDF-based content for representing the trustworthiness of the data published on the Web. But these approaches do not provide a sufficient mechanism to represent the trustworthiness of certain content for individual user.

2.2 Computing Trust

In computer science, Marsh [10] was the first one to analyze trust as a computational concept. Guha et al. [6] proposed a method to propagate both trust and distrust through social networks. In Ziegler's work [13], a trust algorithm called Appleseed has been developed. It employs a spreading activation model that is originally from psychology, to produce a trust ranking of individuals in the network. TidalTrust [3] is another algorithm that performs a breadth-first search, using trust values within the social network. In a trust network, it calculates trust values for a sink node with a weighted average of the source's neighbors' ratings of the sink. Massa [11] proposed a trust propagation metric of searching for trustworthy users over the trust network to alleviate the cold start problem in a collaborative filtering (CF) algorithm.

3 Approach

The previous section addresses the lack of a uniform model for representation and computation of trust. We propose an approach that concentrates on the following research issues:

- How to represent a trust relationship between a user and web content?
- How to start to assess the trust value in that trust relationship?
- How to further compute trust by investigating the semantic relation between different pieces of content?

We discuss these issues now in more detail. First of all, our approach requires a representation model of trust that defines and represents a trust relationship between a user and content on the Web. There are three critical elements to be investigated in our representation model: the trust value that is a rating to certain content from a certain user, the semantics of the content, and the context. The trust model will be formalized using RDF-based vocabularies.

Based on this approach of trust representation, in order to boot the computation process of trust, we first need to acquire initial trust values. Basically there are two solutions: providing the trust values: manually by users and generating them automatically by utilizing existing data. For the first one, it is very difficult to choose a set of Web content as a sample for rating when our task is not domain specific. Another

disadvantage is that sometimes it is not practical to push the users to give these trust ratings. Since there are already many data sources that include the user's ratings and reviews on the Web content and more and more web content is enriched with semantic data, we adopt the latter one as our solution to start assessing the trust. Here we propose two kinds of automatic trust assessment methods: user-based rating methods that integrate various data of ratings and reviews; and semantic-based methods that consider meta-information such as provenance to assign the initial trust values.

The third task is the further computation and inference of trust, of which the main goal is to predict a trust value of web content that the user never rated or obtained before. Comparing the propagation method we mentioned in section 2, we reach this goal by analyzing the semantics of the relationship between two pieces of content, of which one already has a trust value and the other one not yet. Given the semantically annotated and linked data, it is possible to reveal the semantic relationship between the content. To illustrate, we give a simple example. Using "RelFinder"[9], a semantic relation discovery tool for DBpedia, we can find the relationship between two entities in DBpedia: Johannes Vermeer[1] and Night Watch[2]. By observing the graph of paths that represent the relationship, we found six different paths, each of which connects these two entities by chaining several other entities from the knowledge base (DBpedia). The discovered entities located in each path can demonstrate the specific semantics of the relationship that connects these two requested entities. For example, they can be geographically connected because of the entities "Netherlands" and "Amsterdam" in one path or connected by artistic characteristics because of the entity "Baroque painting" in another path. After matching these semantics and context information such as user preferences to analyze which path is more relevant to the trust computation in the specific context, a new trust value can be inferred.

4   Methodology

Our methodology for solving the above questions has three main phases. In the first phase, we will build the representation model for trust, which will take the context into account. Based on the model, a trust vocabulary will be developed to enable the description and access of trust information. In the second phase, based on the representation model and the approach we discussed in the previous section, we will process the computation of trust. For the first implementation, we limit the web content to structured data, e.g. DBpedia or DBLP, or some web pages, e.g. some movie list page, which can be easily transferred to structured data. We will develop a browser plug-in to recommend a trust value when the user browses the web content. The third phase will include the evaluation of our approach. Because of the nature of trust, we will conduct several user studies to evaluate our representation and computation methods. Furthermore, due to the semantic web technologies that we are using, we will consider uncertainty as a key factor in the evaluation.

---

[1] http://dbpedia.org/ page/Johannes_Vermeer
[2] http://dbpedia.org/page/ Night_Watch_%28painting%29

5 Conclusion

We identified the lack of a suitable model for trust between user and content on the Web. This PhD symposium paper sketched our approach that addresses these two main problems: we provide a mechanism for an individual user to represent his/her trust in certain web content and to assess trust values by collecting data from distributed sources. We also propose to compute the trust by utilizing the semantics of the relationship between the various content elements. With our work we hope to make the user-centric data exploitation on the web more reliable, efficient and controllable than it is today. For the first phase of the PhD work, according to the approach and methodology we proposed, the next step is to formalize trust in our scenario. We will build a representation model and an RDF-based vocabulary based on which the further actions can be pursued.

References

1. Artz, D., Gil, Y.: A Survey of Trust in Computer Science and the Semantic Web. J. Web Semantic 5(2), 58–71 (2007)
2. Gil, Y., Artz, D.: Towards Content Trust of Web Resources. In: WWW 2004, pp. 565–574. ACM, New York (2004)
3. Golbeck, J.A.: Computing and Applying Trust in Web-based Social Networks. PhD thesis, University of Maryland (2005)
4. Golbeck, J.A., Parsia, B., Hendler, J.: Trust Networks on the Semantic Web. In: Proceedings of Cooperative Intelligent (2003)
5. Groth, P., Miles, S., Modgil, S., et al.: Determining Trustworthiness of New Electronic Contracts. In: Engineering Societies in the Agents World X, pp. 132–147 (2009)
6. Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of Trust and Distrust. In: WWW 2004, pp. 403–412. ACM, New York (2004)
7. Hartig, O.: Querying Trust in RDF Data with tSparql. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvönen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E. (eds.) ESWC 2009. LNCS, vol. 5554, pp. 5–20. Springer, Heidelberg (2009)
8. Heath, T., Motta, E.: Revyu: Linking Reviews and Ratings into the Web of Data. J. Web Semantic 6(4), 266–273 (2008)
9. Heim, P., Hellmann, S., Lehmann, J., Lohmann, S., Stegemann, T.: RelFinder: Revealing Relationships in RDF Knowledge Bases. In: SAMT 2009, pp. 182–187 (2009)
10. Marsh, S.P.: Formalising Trust as a Computational Concept. PhD thesis (1994)
11. Massa, P.: Trust Metrics in Recommender Systems. In: Golbeck, J.A. (ed.) Computing with Social Trust, pp. 259–285. Springer, London (2009)
12. Ray, I., Ray, I., Chakraborty, S.: An Interoperable Context Sensitive Model of Trust. J. Intell. Inf. Syst. 32(1), 75–104 (2009)
13. Ziegler, C.N., Lausen, G.: Propagation Models for Trust and Distrust in Social Networks. Information Systems Frontiers 7, 337–358 (2005)