Trustworthiness of Data on the Web

Olaf Hartig
Humboldt-Universität zu Berlin
Department of Computer Science
hartig@informatik.hu-berlin.de

Abstract: We aim for an evolution of the Web of data to a Web of trusted data. In our work we develop a trust model for RDF data that is more suitable for the Web of data than the existing coarse-grained approaches that consider trustworthiness on the level of data sources. To enable a trust infrastructure for the Web of data we develop concepts for automatic trust assessment that are based on provenance information and on the opinion of other information consumers. Furthermore, we provide trust-aware data access methods and concepts to implement trust-aware systems.

1 Introduction

Today, a large amount of RDF data is published on the Web; large datasets are interlinked; new applications emerge that utilize this data in novel and innovative ways. However, the openness of the Web and the ease to combine RDF data from different sources creates new challenges. Unreliable data could dominate the result of queries, taint inferred data, affect local knowledge bases, or may have negative or misleading impact on software agents. Hence, questions of reliability and trustworthiness must be addressed. While several approaches consider trustworthiness of information sources, little has been done considering the data itself. Simply adopting the trustworthiness of a source for its data does not consider cases where statements have multiple sources, where information providers (re)publish data aggregated from the original sources, or where inference engines discover implicit facts from statements of different sources. Hence, source-level approaches are too coarse-grained and, thus, insufficient for the Web of data. What is missing is a uniform approach to represent and to assess the trustworthiness of the data itself and standardized mechanisms to access those assessments. Our research work addresses these open issues; we aim to provide concerted and all-embracing solutions that enable a trust infrastructure for the Web of data, solutions that empower systems based on data from the Web to consider the trustworthiness of the involved data in their processing and decisions.

This paper is organized as follows. First, we review related work in Section 2. Section 3 presents our main contributions. In Section 4 we give an overview of our approaches and our current work. Finally, Section 5 concludes this paper with a summary.

2 Related Work

Formalizing trust and trust in the Web is a topic of research since several years. In computer science, Marsh [Mar94] was the first to analyze trust as a computational concept; in the context of software agents he proposed a model with trust values in the interval [-1,1].

Since then, various different trust models have been developed; each approach stresses different characteristics of trust. Artz and Gil [AG07] provide a comprehensive overview of existing trust models. The most common approach to address trustworthiness in the Web are trust infrastructures that are based on a Web of Trust [GPH03]; a Web of Trust is a network of people, agents or peers with trust relationships and a metric to calculate a trust value for each member. For instance, Golbeck et al. [GPH03] propose a trust metric upon a FOAF-based trust model. These Web of Trust approaches consider the trustworthiness of members of the Web.

In contrast to the above mentioned approaches, we focus on the trustworthiness of the data published on the Web, instead of the publishers. Gil and Artz [GA07] call this *content trust* which "is a trust judgment on a particular piece of information in a given context." As the units of content that are being judged Gil and Artz identify Web resources. We doubt this level of granularity is suitable for RDF data where the smallest piece of information is a fact represented by an RDF statement. Hence, our approach is based on the trustworthiness of single statements. To the best of our knowledge, only Mazzieri [Maz04] and Richardson et al. [RAD03] propose trust models that represent content trust on a statement-level. Mazzieri introduces fuzzy RDF; a *membership value* associated with each statement represents the likelyhood the statement belongs to the RDF graph. By equating those membership values with trustworthiness of statements Mazzieri inappropriately mixes two different concepts; trustworthiness is not the same as a fuzzy notion of truth nor is trustworthiness of RDF statements tied to a specific RDF graph. Richardson et al. represent a user's personal belief in a statement by a value in the interval [0,1]. Besides the vague explanation that a "high value means [...] the statement is accurate, credible, and/or relevant" the approach lacks a more formal definition of those values. Thus, what is missing in all cases is a well-founded definition of the meaning of trustworthiness of RDF data.

Several authors propose the basic idea of considering the trustworthiness of data in processing tasks and for decisions. Systems such as TRELLIS [GR02], FilmTrust [GH06], and IWTrust [ZdM05] implement this idea. The TriQL.P approach [BO04] is the most relevant to our work because it explicitly targets RDF data from the Web. The TriQL.P system permits filtering of data that has been aggregated from the Web. Filtering is based on trust policies; these policies are constraints that are enforced during query evaluation and that restrict the resultset of queries. Furthermore, the system explaines why data should be trusted, more precisely, why results passed the filters. The TriQL.P approach does not use trust values for data. However, missing trust values prevent comparisons of the trustworthiness of different pieces of data; moreover, without explicit ratings it is impossible to compare the opinions of multiple information consumers regarding the trustworthiness of the same data. Additionally, TriQL.P does not consider the trustworthiness of inferred statements.

3 Contributions

With our work we address the lack of a uniform approach to represent and to assess the trustworthiness of RDF data on the Web: we develop a trust model and trust assessment methods. Our trust model defines trust values that represent the trustworthiness of RDF

data on a statement-level. To determine those trust values we develop automatic trust assessment methods that are based on provenance information and on the opinion of other information consumers. Furthermore, we provide standardized mechanisms to access and to use those trust values. To gain wide-spread use, our solutions will integrate seamlessly in existing technologies and follow common practices on the Web. In brief, our main contributions are the following:

- an adequate trust model for RDF data
- methods to determine the trustworthiness of RDF data
- access methods for trust values
- techniques to manage trust values in an efficient manner

4 First Achievements and Intended Tasks

In the following we describe our approaches for a Web of trusted data. This section is organized according to our main objectives as introduced in the previous section.

A Trust Model for RDF Data Our fundamental understanding of the *trustworthiness of RDF statements* is the subjective belief or disbelief in the truth of the statements. To enable machine-based processing we introduce a quantifiable measure; we represent the trustworthiness of RDF statements by a *trust value* which is either unknown or a value in the interval [-1,1]. We define the meaning of these values by a specification of the interval boundaries: a trust value of 1 represents absolute belief in the statements; -1 represents absolute disbelief; 0 represents the lack of belief/disbelief. Unknown trust values represent unknown trustworthiness in cases where the trust management system has no information. For a more precise understanding of the exact meaning of a trust value, especially for the intermediate values, we currently study uncertainty and ignorance that causes different degrees of belief [Hal03]. However, to determine the trust values our model defines a *trust function*. A trust function assigns every statement a subjective trust value that represents the trustworthiness of the statement specific to an information consumer. The model does not prescribe a specific implementation of determining the trust values. Instead, we allow each system to provide its own trust function. Nonetheless, we develop concepts for trust assessment methods as we describe below. Additionally, our model defines a *trust aggregation function* to determine the trust value for a set of related RDF statements. As with trust functions, we do not prescribe a specific implementation. However, the minimum and the median are reasonable trust aggregation functions. The minimum is a cautious choice; it assumes the trustworthiness of a set of statements is only as trustworthy as the least trusted statement. The median is a more optimistic choice. To explicitly assert the trustworthiness of RDF data we defined a simple vocabulary (cf. http://trdf.sourceforge.net/trustvocab).

Methods for Trust Assessment To implement trust functions that determine trust values we investigate three distinguishable trust assessment strategies: *user-based rating methods* prompt human information consumers for trust judgements; *provenance-based methods* consider meta-information about the provenance of the data; *opinion-based methods* aggregate trust ratings from other information consumers as well as the reliability of those consumers. Provenance- and opinion-based methods are automatic approaches; these are

more promising with respect to practicability. Hence, our research focuses on them. For our provenance-based approach our aim is to develop a provenance model for RDF data. This model must represent all kind of provenance information that may affect trust decisions. So far, we identified the following relevant information: publisher of the dataset, creation method and creation time of the dataset, and publisher and publication time of possible original sources. Based on our analysis we derive requirements for provenance descriptions. We study existing practices and possibilities to describe provenance of RDF data on the Web such as named graphs [CBHS05] and semantic sitemaps [CSD+08]; we will identify missing options and we will propose complementing extensions. Besides explicit assertions inference engines allow to discover implicit facts. We will develop concepts to derive the trustworthiness of inferred data from the trustworthiness of the premise set.

Access Methods for Trust Ratings Users as well as software agents have to be able to utilize the trust values and base their decisions upon them. To enable access and use of the trust values we develop concepts to extend access methods for RDF data accordingly. For instance, we extended the RDF query language SPARQL [PS08]. Our trust-aware extension *tSPARQL* (cf. http://trdf.sourceforge.net/tsparql) adds two new clauses, namely the `TRUST AS` clause and the `ENSURE TRUST` clause. The first clause allows access to the trust value associated to the triples that match a specific graph pattern; the second clause expresses a trust requirement. We implemented a prototypical extension to the ARQ (cf. http://jena.sourceforge.net/ARQ) query engine to evaluate our approach. We measured the time to complete the ARQ test suite for the ARQ query engine and for our extension that processes generated trust values. These preliminary tests indicate that processing of trust values has no significant impact on the query execution performance. Besides SPARQL, other methods to access RDF data exist; we will investigate them in order to develop concepts similar to our SPARQL extension.

Efficient Management of Trust Ratings Systems that consider the trustworthiness of RDF data must represent and process trust values in an efficient manner. For instance, we investigate caching strategies for trust values because trust assessment may become costly. To study these implementational issues we developed a prototypical navigation system (cf. http://trdf.sourceforge.net/navi); in order to determine relevant target locations with user-specified properties this system does not use a static local database as usual, instead, our system aggregates relevant RDF data from the Web. It is this kind of systems that must consider the trustworthiness of RDF data and that will benefit from our contributions. As another testbed we are adding trustworthiness-based decisions to a semantic tagging system. Based on our experiences with these sample applications we will provide suitable concepts to implement trust-aware systems.

5 Conclusions

We identified the lack of suitable concepts to deal with the trustworthiness of data on the Web. Our work addresses this problem: we provide a trust model for RDF data, trust assessment methods, methods to utilize trust assessments, and implementation strategies for trust-aware systems. This paper sketched our current achievements and approaches. With our work we hope to make the Web of data more reliable than it is today.

References

[AG07] Donovan Artz and Yolanda Gil. A Survey of Trust in Computer Science and the Semantic Web. *Journal of Web Semantics*, 5(2), 2007.

[BO04] Christian Bizer and Radoslaw Oldakowski. Using Context- and Content-Based Trust Policies on the Semantic Web. Poster at WWW, 2004.

[CBHS05] Jeremy J. Carroll, Christian Bizer, Pat Hayes, and Patrick Stickler. Named Graphs, Provenance and Trust. In *Proc. of WWW*, 2005.

[CSD+08] Richard Cyganiak, Holger Stenzhorn, Renaud Delbru, Stefan Decker, and Giovanni Tummarello. Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web. In *Proc. of ESWC*, 2008.

[GA07] Yolanda Gil and Donovan Artz. Towards Content Trust of Web Resources. *Journal of Web Semantics*, 5(4), 2007.

[GH06] Jennifer Golbeck and James Hendler. FilmTrust: Movie Recommendations using Trust in Web-based Social Networks. In *Proc. of CCNC*, 2006.

[GPH03] Jennifer Golbeck, Bijan Parsia, and James A. Hendler. Trust Networks on the Semantic Web. In *Proc. of the 7th Int. Workshop on Cooperative Information Agents (CIA)*, 2003.

[GR02] Yolanda Gil and Varun Ratnakar. Trusting Information Sources One Citizen at a Time. In *Proc. of ISWC*, 2002.

[Hal03] Joseph Y. Halpern. *Reasoning about Uncertainty*. MIT Press, Cambridge, MA, USA, 2003.

[Mar94] Stephen P. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, Department of Mathematics and Computer Science, 1994.

[Maz04] Mauro Mazzieri. A Fuzzy RDF Semantics to Represent Trust Metadata. In *Proc. of SWAP*, 2004.

[PS08] Eric Prud'hommeaux and Andy Seaborne. SPARQL Query Language for RDF. W3C Recommendation, January 2008.

[RAD03] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust Management for the Semantic Web. In *Proc. of ISWC*, 2003.

[ZdM05] Ilya Zaihrayeu, Paulo Pinheiro da Silva, and Deborah L. McGuinness. IWTrust: Improving User Trust in Answers from the Web. In *Proc. of iTrust*, 2005.