Representing, Proving and Sharing Trustworthiness of Web Resources Using *Veracity*

Grégoire Burel, Amparo E. Cano, Matthew Rowe, Alfonso Sosa

Oak Group, Department of Computer Science,
University of Sheffield,
Sheffield, United Kingdom
{g.burel,e.cano,m.rowe,a.sosa}@dcs.shef.ac.uk

Abstract. The World Wide Web has evolved into a distributed network of web applications facilitating the publication of information on a large scale. Judging whether such information can be trusted is a difficult task for humans, often leading to blind trust. In this paper we present a model and the corresponding *veracity* ontology which allows trust to be placed in web content by web agents. Our approach differs from current work by allowing the trustworthiness of web content to be securely distributed across arbitrary domains and asserted through the provision of machine-readable proofs (i.e. by citing another piece of information, or stating the credentials of the user/agent). We provide a detailed scenario as motivation for our work and demonstrate how the ontology can be used.

1 Introduction

The World Wide Web (WWW) facilitates the contribution of ideas spanning a large information network. The visibility of this information introduces the problem of *what* to trust and *whom* to trust. Currently it is up to web users to make a conscious decision whether to believe what they are reading or not. Trust needs to be derived using automatic means despite relying on error prone user generated content. For example, the *Wikipedia biography controversy*[1] highlighted the issue of information trustworthiness for human and software agents that might exist independently of the reliability of an information source [1]. As a consequence, it is necessary to not only trust information provenance but also information content in order to confirm the quality of a piece of data. This paper addresses trust in a piece of information published on the WWW through the use of Semantic Web (SW) technologies which we define as a *proposition*. A *proposition* can be any piece of web content which is identified by a URI - akin to a resource or statement via reification in SW terms. We present a lightweight decentralised trustworthiness model describing the need to assert proofs in a trust decision and an ontology named *Veracity*.

[1] The Wikipedia biography controversy, http://en.wikipedia.org/wiki/Wikipedia_biography_controversy.

2 Trustworthiness Evaluation of Online Resources

The definition of trustworthiness is not absolute but highly contextual since it encapsulates social and personal concepts such as reputability, popularity, reliability and likelihood [2]. We define a proposition as trustworthy if the carried information is reliable, commonly accepted as true given pre-existing trusted knowledge and stable over time. Basing trustworthiness only on social or personal concepts may lead to mistakes, as there is no explicit rationale behind this type of trust endorsement. As a consequence, it is important to base trustworthiness on rational and socially independent variables.

2.1 Trustworthiness Evaluation Scenario

Consider Alice, a journalist, preparing an article about Einstein. In order to write her article, she needs to find as much reliable information as possible. During her search on Internet she may find different information about the famous scientist written by different known and unknown authors. Imagine that Alice finds the following proposition on some obscure website: “Einstein worked at the University of Berlin and was a physicist”. For evaluating the trustworthiness of the proposition Alice can 1) Use her personal knowledge about Einstein; 2) Identify the author of the proposition as a “trusted authority” or as a reliable person given the current context; 3) Identify if the author has the knowledge required for asserting the information about Einstein; 4) Search for external trustworthy information that asserts a similar statement; 5) Ask a domain expert for estimating the trustworthiness of the proposition.

In the first case, Alice just needs to compare if her beliefs match the proposition. This case is unlikely since she does not actually know anything about the scientist yet. The second case implies that the editor of the information is reliable in general or in the terms of Alice. The third technique requires Alice to look for information about the author of the proposition that confirms that he knows directly or not about Einstein (e.g. the author may be a “physicist”). In the fourth case, she needs to look for external information or a reference that confirms the considered proposition. Finally, the last case requires Alice to find somebody that knows directly about the topic.

3 Related Work

Artz and Gil [3] distinguish between different methods that can be applied for deriving trust on the WWW and SW such as policy management, provenance analysis and content trust. We now review approaches in each of these areas.

Trust Policies: Information integrity and identification ensure that the parties involved in a trust situation cannot be altered separately. Generally, two approaches are taken for ensuring this integrity: the first involves the utilisation of a centralised server designed for managing a trust assertion [4–6];

the second involves distributed mechanisms for managing integrity such as digital signatures[2]. Compared to distributed systems, centralised systems are rather limited since a relation of trust must exist between a trust server and a relying party to work properly. These systems are also often unable to cope with proposition changes or require cache systems [7].

Information Provenance: Information provenance relies on the assumption that trust in a proposition can be estimated using the trust owned by an agent and network analysis. Hartig and Zhao [8] and Golbeck [9] explain how trustworthiness of data is based on assessing its provenance (author, timeliness, etc). Heath et al [6] identify factors influencing the trust between social entities given a particular proposition, Ziegler et al [5] and Carroll et al [10] present different trust metrics that can be used in social networks and similar work in [6] calculates trust ratings from a semantic social network. Several ontologies exist for expressing trust in a given proposition such as the Web Of Trust (WOT)[2] ontology - expressing author identity and trust for a given RDF resource - and the Trust Ontology[3] - for modelling the relation of trust between agents according to a given topic.

Content Trust: According to Gil and Artz [2], evaluating trustworthiness based on information provenance is limited due to the utilisation of indirect and non-contextual knowledge, instead the authors propose evaluating the trustworthiness of information based on its content and context rather than its author. Similarly the TRELLIS [7] application constructs relations between propositions, thereby providing trust in content without provenance information and the Proof Markup Language [11] (PML) ontology is designed to represent metadata about propositions - in the context of question answering. Both TRELLIS and PML do not provide proposition signatures and versioning thus confining its applicability to centralised data.

Current research focuses on the information provenance [4–6], content trust [7], policies and metrics [4, 5] independently. Metrics can be applied independently of a model meaning that trust metrics may be applied using a similar model. Unfortunately, there is no existing model encompassing the representation of provenance, content trust and supporting the reliability of these statements independently of server side assumptions (centralised server/authority).

4 Requirements

Our scenario (section 2) outlines the need for a model that supports each of the techniques that Alice can apply for evaluating the trustworthiness of a proposition. Referring back to the scenario; in the first case Alice merely relies on her knowledge of the given proposition, which therefore does not necessitate trustworthiness. As a consequence this step is out of scope with modeling trustworthiness, thus to support the other cases the following needs must be fulfilled:

[2] Web Of Trust (WOT), http://xmlns.com/wot/0.1/.

[3] The Trust Ontology, http://trust.mindswap.org/ont/trust.owl

Identify a proposition: To establish the veracity in a piece of information the proposition must be identifiable.

Describe the trustworthiness of a proposition: Once a trust decision is made, a formal description of that decision must be given.

Identify an agent: It is necessary to identify accurately the people or agents involved in the evaluation of the trust of a proposition. Particularly, the person asserting trust on the proposition must be identified. Moreover the identity of the agent must be protected so that it cannot be reused.

Provide agent credentials: Should an agent state their position on the veracity of a proposition, it is essential that the agent provides proof of their background knowledge when making such a statement about the proposition.

Provide supporting information: Should an agent find a piece of information external to a given proposition that supports or refutes the proposition's veracity, then the agent should be able to cite this piece of information.

Security of assertions: In order to provide a secure environment for the trustworthiness assertion, a mechanism should exist to verify that an agent really asserted a particular trustworthiness value on a particular proposition.

Reliable assertions: A modification in a proposition should invalidate all the previous trustworthiness assertions related to this piece of information.

No predefined trust assumptions: The veracity of a proposition should not require any a priori trust relations. Particularly, it should not assume the existence of a central authority for verifying the validity of a trust assertion.

5 Towards a Shared Model of Rational Trustworthiness

Previous ontologies are either incomplete or rely on a controlled network or do not define clearly the mechanism behind an assertion of trust over a proposition. In the context of our scenario, it is required to provide a distributed model that enable the assertion of explanatory trust.

5.1 Knowledge Factors in Trust

According to our definition, a proposition is trustworthy if it is admitted to be true given some *proven* background knowledge. Our definition relies on the use of rational and explicit contextual information for the assertion of trust over a proposition. However, current models [7, 5, 10, 6, 8] have no formal descritpion that ensures that a trustworthiness assertion has been performed according to our definition. Gil and Artz define *Entity Trust* and *Content Trust* models for asserting the trustworthiness of a proposition. However, despite adding a stronger context to their content trust model, Gil and Artz's definition incorporates fuzzy parameters that remain hard to evaluate automatically and impartially due to their social entailment. As a result, a strict model that relies on automatically processable and rational variables is requiered. Such type of representation needs to symbolise the factual knowledge that a *third-party* uses for making a trust decision. In this paper, we refer to *Social Trustworthiness* as the model of trust based on fuzzy factors while we use the term of *Rational Trustworthiness* for a trust assertion based on verified and valid information.

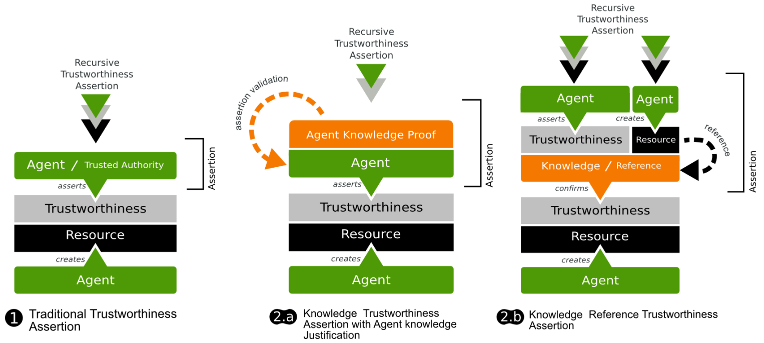
Recursive Trustworthiness Assertion
Agent / Trusted Authority
asserts
Trustworthiness
Resource
creates
Agent
Assertion
1 Traditional Trustworthiness Assertion
Recursive Trustworthiness Assertion
assertion validation
Agent Knowledge Proof
Agent
asserts
Trustworthiness
Resource
creates
Agent
Assertion
2.a Knowledge Trustworthiness Assertion with Agent knowledge Justification
Recursive Trustworthiness Assertion
Agent
Agent
asserts
creates
Trustworthiness
Resource
reference
Knowledge / Reference
confirms
Trustworthiness
Resource
creates
Agent
Assertion
2.b Knowledge Reference Trustworthiness Assertion

Fig. 1. Social Trustworthiness (1) and Rational Trustworthiness (2) — 2.a) Justifying Social Trustworthiness; 2.b) Knowledge Reference.

5.2 Social Trustworthines vs Rational Trustworthines

Even if trustworthiness is asserted through social means (Fig. 1.1), its reliability can be verified through rationality. The difference between social and rational trust is that the former relies exclusively on an unconditional trust in a person's judgment, while the latter uses supporting trust statements that can be used for evaluating the quality of a trust judgment of a person in a particular concept based on known facts. However, Social trust can still become rational by adding judgment justification statements that prove the personal knowledge of the user. This type of rational trust is summarised in Fig. 1.2. When asserting trustworthiness on a piece of information (resource), a user can justify his decision by proving that he has knowledge about the information he wants to prove as trustworthy or not (for example, by showing that his judgement about the proposition 'Einstein was a physicist' is valid because 'As Einstein, he is a physicist'). Rational trust can also be asserted by directly referring to supporting information rather than referring to a social assertion. In this context, the trustworthiness of a statement is not endorsed by a person given some knowledge but endorsed directly by another source of information or reference through a user assertion (Fig. 1.3). In this context, the trustworthiness relation becomes dereferenced to the cited knowledge.

5.3 Sharing Trustworthiness

Representing trustworthiness on the WWW demands some attention to the effect of the distribution of information in an open network. Without particular measures, it becomes evident that it is easy to either falsify the content of a proposition, its authorship or the agent behind a trust assertion. It is important to not depend on a closed or controlled network for maximising the spreading of trustworthiness information across the network. Because, the necessity of secure trustworthiness assertions is a technical issue our approach for dealing with these problems is discussed in the implementation described in the following section.

6 The Veracity Ontology

The Veracity Ontology[4] (VO) is designed for representing our trustworthiness model described in the previous section. The ontology is organised through three levels of trustworthiness: 1) Social; 2) Knowledge; 3) Knowledge Reference.

6.1 Imported Ontologies

The VO reuses the FOAF ontology for modeling the agents asserting the trustworthiness information on a resource while the WOT ontology is imported for managing `foaf:Agent`[5] signatures. FOAF fits perfectly the requirement of representing the social components of our ontology. The WOT ontology supports the insertion of a public key into a FOAF profile. This assertion enables the creation of a web of trust through the use of digital signatures. Digital signature ensures that: 1) The provenance of a web resource cannot be falsified easily; 2) A web resource cannot be modified without revoking the provenance of the information. This ontology is useful in a distributed environment where trustworthiness can be asserted anywhere. The WOT ontology ensures that these assertions cannot be falsified thus providing a solid base for valid assertions.

6.2 Core Components

The VO relies on the concepts of agent, proposition and trustworthiness. An agent models the agents asserting a trustworthiness value to a specific entity. It represents the social component of our ontology. A proposition is a model of a web resource (or a semantic statement) on which trustworthiness information can be inserted. Trustworthiness defines if an information is trustworthy or not.

foaf:Agent An agent is modeled using the FOAF ontology (Fig. 2). We decided not to directly reuse the `foaf:Document` class since it does not map to an arbitrary `rdfs:Resource`. However, the resources of an entity are modeled using similar concepts in order to be aligned easily to DC Terms[6] and, depending on the context, to `foaf:Document` if necessary.

Proposition: A `Proposition` borrows concepts from FOAF and DC without referring directly to them. However, it is possible to align the properties of a `Proposition` to DC properties depending of the final application. The `Proposition` properties are inspired by several properties from the FOAF ontology, such as `foaf:sha1` for the `has_checksum` property and properties from the DC ontology such as `dc:creator` for `has_maker`. The checksum enables the identification of a specific version of a `Proposition` and the creator property asserts which `foaf:Agent` created the resource. The `topic` property is similar to `foaf:topic`; it can be used for deriving the creator's knowledge and the trustworthiness of an assertion (Fig. 2).

[4] The Veracity Ontology, http://purl.org/net/veracity/ns#.

[5] The classes and properties are written using the `Typewriter font` and prefixed with the corresponding ontology. If no prefix is provided, the Veracity ontology is used.

[6] DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms/

Trustworthiness: For modeling the trustworthiness of a `Proposition` we must assert a value which denotes an entity as trustworthy or not. Our model (Fig. 2) identifies two trust levels represented as a property of `TrustWorthiness`: A resource may be trusted or not. Therefore we define the `TrustWorthiness` class which has the `trusted` property. The range of this property depicts the instance of the class as being either trustworthy or not. `TrustWorthiness` is associated with a `Proposition` using the `has_trustworthiness` relation. Therefore the `Proposition` now has an associated trust value.

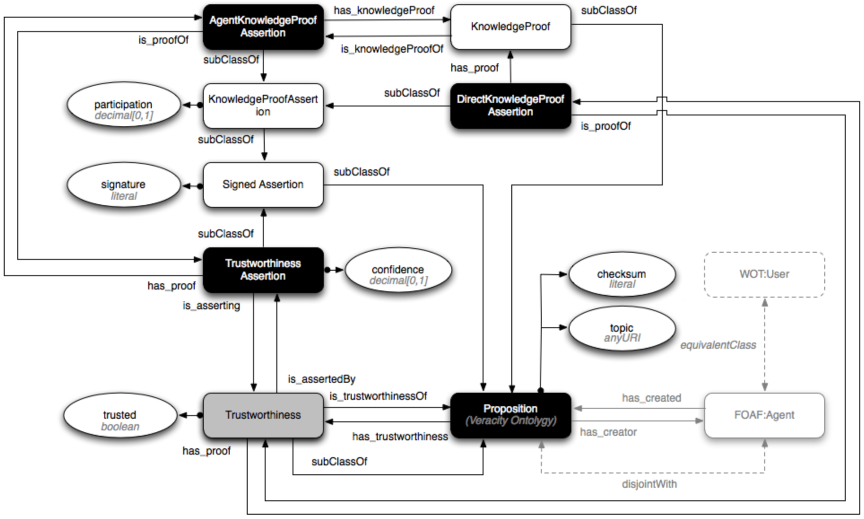

Fig. 2. The Veracity Ontology

6.3 Trustworthiness Assertion Components

A `TrustWorthiness` value cannot be really used before being endorsed by another entity that vouches for a particular `TrustWorthiness` value (trustworthiness assertion). The VO defines three different types of assertions that can be combined in a recursive fashion in order to model deep assertions and support the application of complex trust analysis metrics. In order to ensure that a resource is not changed after being endorsed, each type assertion is defined as a subclass of `SignedAssertion`. As a consequence, an assertion inherits the `checksum` property of a `Proposition`. A `TrustWorthiness` value also inherits from `Proposition` for the same reasons. As said previously, the model needs to prevent an agent from using the identity of another one for making fake assertions. A way to guarantee that an assertion is correct is to use Digital Signatures (DS). In the VO, DS are asserted using the WOT ontology. By merging the `wot:User` class from WOT with the `foaf:Agent` class from FOAF we can associate a public key with any foaf:Agent for matching the agent with a unique public key. As a consequence, each trust assertion can be associated with a unique DS that can be matched with a specific Agent.

TrustworthinessAssertion: This class manages a type of assertion that only relies on social information. This assertion implements the model described in the Fig. 1. As a consequence, for being used by a trust metric, the entity asserting a TrustworthinessAssertion must be identified as a trusted authority during a trustworthiness evaluation. Because the VO uses FOAF for modeling entities and any type of assertion is a subclass of SignedAssertion, a TrustworthinessAssertion is a relation between a foaf:Agent DS (using the signature property) and a TrustWorthiness (using has_trustworthiness and is_assertedBy properties). For allowing a more precise description of trustworthiness, the confidence property inserts a confidence level of the endorsement of TrustWorthiness by a foaf:Agent. The confidence models the accuracy of a foaf:Agent judgement for asserting the trustworthiness of a Proposition. For instance, by supporting a distrust on a resource (trusted set to *false*) with a confidence of *1.0* means that an foaf:Agent is sure that the information is not valid.

AgentKnowledgeProofAssertion: This trustworthiness assertion goes on the top of a TrustworthinessAssertion. An AgentKnowledgeProofAssertion assertion implements the first level of Rational Trustworthiness described in the Fig. 1.2.a. This assertion enables an entity to justify the personal knowledge used by a particular foaf:Agent. Typically, a foaf:Agent that performs a TrustworthinessAssertion may justify his decision by using a proof of his knowledge through the use of the has_proof property of a TrustworthinessAssertion. Similarly to the confidence property used in a TrustworthinessAssertion, an agent may use the participation property for defining how much of the personal knowledge participates in his decision. The linked information may have different formats. However, the relation between the user, the knowledge and the endorsed proposition should be semantically defined for enabling automatic trustworthiness validation.

DirectKnowledgeProofAssertion: This assertion refers to the Rational Trust described in the Fig. 1.2.b. A DirectKnowledgeProofAssertion enables an agent to cite a resource that tends to validate or invalidate a Proposition. Contrary to a AgentKnowledgeProofAssertion, this type of proof does not imply that the agent citing the resource has knowledge concerning the cited resource. The agent makes an explicit relation between two documents and presents how a cited document participates in a Trustworthiness relation. So, this relation enables the propagation of trust from a cited resource to a Proposition. An agent may also use the confidence property for specifying how strong is the relation between the two resources.

6.4 Fulfillment of the Requirements

The VO fulfils the previous model requirements. The proposed ontology enables the assertion of a trustworthiness value in a secure way thanks to the

WOT ontology and the `trusted` property on usual web resources or semantic web resources. By reusing the FOAF and WOT ontologies, agents can be uniquely identified. The need for a contextual knowledge and knowledge references is satisfied respectively by an `AgentKnowledgeProofAssertion` and a `DirectKnowledgeProofAssertion`. Because each assertion can be performed by anybody, the ontology satisfies the requirement of third-party assertions. Finally, the use of digital signatures enforces trusted assertions. The Table 1 summarises the differences between the VO and the existing ontologies.

	Prop. Ident.	Desc. of Trust.	Agent Ident.	Agent Cred.	Info. Supp.	Secure Assert.	Reliable Assert.	No Prev. Trust
Trust Ont.	○	○	●	●	–	–	–	–
WOT/Konfidi	○	○	●	●	–	●	○	●
TRELLIS [7]	●	●	●	●	●	●	○	–
PML [11]	●	●	●	○	●	●	–	–
Veracity	●	●	●	●	●	●	●	●

Abbreviations: ● = Yes. ○ = Limited/Implicit. − = No.

Table 1. Veracity Compared to the Existing Models

7 Trustworthiness Evaluation using *Veracity*

Considering the scenario from section 2, Alice might find that the University of Berlin trust assertion on the proposition “Einstein was a physicist” is enough to confirm the veracity of the proposition. The identification of the university as a “trusted authority” or an agent, can be asserted through the use of a `TrustworthinessAssertion`.

Alice can also find a similar trust assertion on another website. However, the author of the assertion, Bohr, is unknown to Alice. As expressed in the scenario, Bohr can prove that he is a physicist using a `AgentKnowledgeProofAssertion` in order to be trusted by Alice in the context of the proposition.

In the fourth case of the scenario, Alice needs to find external resources that provide a similar proposition. The VO can be used by a third-party agent for asserting a reference to external information. For instance, one could refer to the DBpedia page of Einstein since it states that “Einstein was a physicist”. As a consequence, if DBpedia is considered by Alice as a “trusted authority”, the initial proposition becomes trustworthy automatically. This assertion can be supported directly by using a `DirectKnowledgeProofAssertion`:

```
@prefix vo: <http://purl.org/veracity/ns#> .
<http://example.com/Albert_Einstein#physics> a vo:Proposition .
<http://example.com/Albert_Einstein#physics> vo:has_trustworthiness _:bnode1 .
_:bnode1 a vo:Trustworthiness .
_:bnode1 vo:trusted "true"^^xsd.boolean.
_:bnode1 vo:has_proof _:bnode2 .
_:bnode2 a vo:DirectKnowledgeProofAssertion .
_:bnode2 vo:has_proof <http://dbpedia.org/resource/Albert_Einstein> .
```

The last case is solved indirectly: Alice cannot seek expert advice but she has access to third-party assertions since each assertion can be done by any agent.

8 Conclusions

We have presented an approach for modeling the veracity of information on the web. Our model enables the verification of a proposition at the information level in a reliable and distributed fashion rather than relying only on its provenance. To prove the veracity of a piece of information it employs: 1) A Trustworthiness assertion, which links a statement to a trusted authority or agent; 2) An Agent Knowledge Proof Assertion, which proves that an agent is credited to label a statement as being trusted or not; and 3) A Direct Knowledge Proof Assertion, which proves that a statement is trusted by providing a reference to another statement.

Acknowledgements The research leading to these results has received funding from the EU project WeKnowIt (ICT-215453).

References

1. Chesney, T.: An empirical examination of wikipedias credibility. First Monday **11**(11) (2006)
2. Gil, Y., Artz, D.: Towards content trust of web resources. Web Semantics: Science, Services and Agents on the World Wide Web **5**(4) (December 2007) 227–239
3. Artz, D., Gil, Y.: A survey of trust in computer science and the semantic web. Web Semantics **5**(2) (2007) 58–71
4. Golbeck, J., Hendler, J.: Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. Engineering Knowledge in the Age of the SemanticWeb (2004) 116–131
5. Ziegler, C., Lausen, G.: Spreading activation models for trust propagation. In: Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04), IEEE Computer Society (2004) 83–97
6. Heath, T., Motta, E., Petre, M.: Computing word-of-mouth trust relationships in social networks from semantic web and web 2.0 data sources. In: Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web. (2007)
7. Gil, Y., Ratnakar, V.: Trusting information sources one citizen at a time. In: The Semantic Web ISWC 2002. (2002) 162–176
8. Hartig, O., Zhao, J.: Using Web Data Provenance for Quality Assessment. SWPM (2009)
9. Golbeck, J., Hendler, J.: Inferring binary trust relationships in web-based social networks. ACM Trans. Internet Technol. **6**(4) (2006) 497–529
10. Carroll, J.J., Bizer, C., Hayes, P., Stickler, P.: Named graphs, provenance and trust. In: Proceedings of the 14th international conference on World Wide Web, Chiba, Japan, ACM (2005) 613–622
11. McGuinness, D.L., Ding, L., da Silva, P.P., Chang, C.: Pml 2: A modular explanation interlingua. In: Proceedings of AAAI. Volume 7. (2007)