DC Proposal: Evaluating Trustworthiness of Web Content using Semantic Web Technologies

Jarutas Pattanaphanchai
Advisors: Professor Dame Wendy Hall and Dr. Kieron O'Hara

Electronics and Computer Science, Faculty of Physical and Applied Science,
University of Southampton,
Southampton, SO17 1BJ, United Kingdom
{jp11g09, kmo, wh}@ecs.soton.ac.uk

Abstract. Trust plays an important part in people's decision processes for using information. This is especially true on the Web, which has less quality control for publishing information. Untrustworthy data may lead users to make wrong decisions or result in the misunderstanding of concepts. Therefore, it is important for users to have a mechanism for assessing the trustworthiness of the information they consume. Prior research focuses on policy-based and reputation-based trust. It does not take the information itself into account. In this PhD research, we focus on evaluating the trustworthiness of Web content based on available and inferred metadata that can be obtained using Semantic Web technologies. This paper discusses the vision of our PhD work and presents an approach to solve that problem.

1 Introduction

Trust plays an important role in the process of consuming data in many different circumstances such as communication between humans or data exchange between a human and a computer. Untrustworthy data may lead to wrong decisions or may make users misunderstand the concept or story, especially on the Web which has an abundance of information, but there is a lack of any control over the quality of its publications. With the incredible increase in the amount of Web content, it is becoming more necessary for users to be able to evaluate the trustworthiness of the information they use in order to judge whether to trust and use it or not. Therefore, having a trust mechanism for users to assess whether information is trustworthy is important and useful for the proper consumption of such information. However, trust is a subjective issue and dealing with it is a complicated task because it depends on the context in which the information is being considered. For example, people would trust a doctor to provide complete, accurate, and correct information about their health, but not about their finances.

The Semantic Web [1] is a technology that has been designed to make computers more intelligent by allowing them to understand the semantics of information and process it properly. RDF (Resource Description Framework) is used

to describe things in a way that computers can understand and also to portray the relationships between those things. Therefore, RDF can be used to represent metadata about Web resources. It also is used by machines to process this metadata in order to provide trust information to users. With this feature, less effort is needed on the side of humans than if they were to manually assess the trustworthiness of the information themselves [12, 4]. RDF provides an opportunity to produce an effective trust model which uses metadata that is available in the Semantic Web for evaluating the trustworthiness or credibility of Web content. It also helps the Semantic Web at its current stage to be more successful because end users can have greater confidence in it.

Recent work on trust has focused on authentication-, reputation-, and policy-based trust [6, 8, 2], but it does not consider the content itself. The concept of content trust was first proposed by Gil and Artz [4]. They defined content trust as "*a trust judgement on a particular piece of information or some specific content provided by an entity in a given context*". They also described the factors which influence a user's decision on whether to trust content on the Web. Moreover, they introduced the content trust model, which solves the problem of assessing the reliability of Web resources, by inferring the trustworthiness of the content of these Web resources [4]. Their work proposed the preliminary concept of content trust which can be explored more to produce a reasonable model. Similarly, our work considers the trustworthiness of the information on the Web based on the sources of that information. Credibility is another concept which is similar to evaluating the trustworthiness of the information on the Web. It focuses on studying and analysing factors that influence a user's decision on whether to trust the information on the Web. Several works have studied and proposed criteria for use in evaluating the credibility of Web sites or Web information. For example, the authority of the source that creates the information, the accuracy of the information, the appearance/presentation (such as the user interface, graphic design, and navigation) and the speed of loading the document [3, 11, 15]. However, each unique set of criteria presented in the different pieces of research has its limitations (e.g. it is hard to collect the information based on that criterion directly from the Web or it only slightly reflects the credibility of the information content itself). Therefore, we have to select the criteria that can be used in practice and that have a significant impact on the evaluation of the trustworthiness of Web content.

In this paper, we propose a model to evaluate the trustworthiness of information on the Web. This model uses Semantic Web technologies to gather metadata, which is collected based on our credibility evaluation criteria. The main contribution of our work is integrating metadata to build a data model that can be used to evaluate the trustworthiness of the information on the Web. We present these integrated metadata in an easily understandable form to the users who will, in turn, use this information to support their decisions of whether or not to trust the information provided on this Web. The rest of this paper is structured as follows. We review related work in Section 2. In Section 3, we introduce our proposed trust model, describe the concept of the model, and discuss

the criteria that we use in our work. In Section 4, we propose our methodology to address the problem of evaluating the trustworthiness of content and present future work. Finally, in Section 5, we conclude our work.

2 Related Work

In the early stages of trust research, researchers focused on policy-based and reputation-based methods. The policy-based method assesses trust based on a set of rules. Bizer and Cyganiak [2] proposed a framework called *WIQA (Information Quality Assessment Framework)*, which is a set of software components that can be employed by applications in order to filter information based on quality policies. The reputation-based methods estimate trustworthiness by using other users' opinions or recommendations. Several trust metrics have been studied, and algorithms to compute trust across trust networks or social networks which have been presented [7, 9, 14]. The work presented so far focuses on evaluating the trustworthiness of the entities, which are judged based on their identities and their behaviours; for example using digital signatures or rating from recommendation system. However, this work did not take the information provided by such entities into account. More recent research has proposed evaluating trust based on content. Content trust is a concept which judges the trustworthiness of data based on features of that content or information resources. Some research uses RDF or annotations to present information about the source and content of desired information, which can then be used to determine that information's trustworthiness [5]. Other approaches discuss the factors which influence users' decisions on whether to trust the content and use these factors as criteria to evaluate the trustworthiness or credibility of information [4, 13, 3, 15, 11]. Instead of considering only the content, the entity that publishes that content should also be considered– we should assess whether or not authors who provide the information can be considered trustworthy. Therefore, we should consider both dimensions (entity and content), and this will help produce a more reasonable approach to evaluating the trustworthiness of information, and to provide support for making decisions.

3 The proposed model

3.1 Basic Concept

The Semantic Web is an extension of the existing Web, designed with the goal of letting computers deal with data rather than just documents. It describes facts about things and how they are related using RDF (Resource Description Framework) in the form of subject-predicate-object expressions called triples. RDF allows both structured and semi-structured data on the Web to be combined, exposed and shared across different sources or applications. In addition, it allows both users and software to follow links to discover more information related to these data [10]. Accordingly, the Semantic Web provides a way to

gather metadata that is useful for evaluating the trustworthiness of information and also provides an opportunity to adopt trust into the Semantic Web itself.

In our work, we define an entity as a source which provides or publishes information. Therefore, we will evaluate an entity based on its credentials or its identity. In addition, we also analyse the available metadata to estimate the trustworthiness of the information on the Web. Therefore, in this paper, we consider the trustworthiness of the information on the Web as an evaluation of the metadata based on a set of trustworthiness criteria. Our proposed model deals with integrated metadata which is provided alongside the content of the Web (explicit metadata) and metadata inferred from a Semantic Web data resource (implicit metadata). The following sections describe the criteria that we use in our model and the architecture of the model.

3.2 Evaluation Criteria

Previous research proposed several factors that affect users' decisions on trusting content provided by an information resource on the Web [4, 13, 3, 15, 11]. These factors range from the source of that information to the information which is provided. An example of these factors are the author or the organization which publishes the information, the graphic design of the Web page, bias, and likelihood (the probability of the content being correct). We investigated the criteria from those pieces of research. We found that some of these criteria can be adopted for implementation in practice, such as the currency criterion that can be assessed based on the time stamp from the system and the authority of the information. However, some of these proposed criteria required data that is difficult to gather or do not have significant impact. For example, the time a document needs to be loaded may indicate the performance of the system and may influence trust but it does not reflect the information on the Web itself and thus it has less impact on the trustworthiness of information.

Figure 1 shows the criteria for evaluating the credibility or the trustworthiness of the information on the Web which is defined in studies that were discussed above. It shows that a number of characteristics, such as authority (source), accuracy, currency, and relevance, appear three times across the four studies which studied different domains and participants. This indicates that these criteria are the common criteria which can be used across the different circumstances to evaluate the trustworthiness of the information on the Web and therefore, we also use them in our approach. In more detail, we consider the following factors:

In conclusion, we selected the following four factors to use in our model since they have been common factors in several domains and can be adopted for automatic analysis:

- *Authority*: The reputation of the source that produced the content. We consider this criterion on two levels, the institutional and the individual level.
- *Currency*: Whether the content of the document is up-to-date or is regularly updated.
- *Accuracy*: Whether the information in a document is accurately expressed i.e. it is grammatically correct and lacks spelling mistakes.

Gil and Aetz (2007)	Wathen and Burkell (2002)	Fogg et.al. (2000)	Rieh and Belkin (1998)
• Topic • Context and criticality • Popularity • Authority • Direct experience • Recommendation • Related resources • Provenance • User expertise • Bias • Incentive • Limited resources • Agreement • Specificity • Likelihood • Age • Appearance • Deception • Recency	• Surface credibility - Appearance/ presentation e.g. colour, graphics, font size - Usability/interface design e.g. menus, navigability - Organization of information e.g. layers, choice of detail level • Message credibility - Source e.g. expertise, credential. - Message e.g. relevance, currency, accuracy	• Real world presence e.g. an organization's physical address, a phone number, or photographs of the members. • Error e.g. typographical error, grammatical error • User interface/navigation • advertisements	• Source - Institutional level e.g. URL, name of institution. - Individual level e.g. identification of author/ creator, affiliation, qualification of creator. • Content e.g. theories from education sites, bibliographies, contact information, get information that meet users' need. • Format e.g. graphical images, information structure. • Presentation e.g. writing style, references, the size of document. • Currency • Accuracy • Speed of loading

Fig. 1. Comparison of Credibility/Trustworthiness Influence Factors in Four Studies

– *Relevance*: Whether the content meets the users' needs, which means it is useful for them.

We use these criteria to evaluate the trustworthiness of content. Our assessment approach is inspired by Wang et al. [13]. We use a range of metadata, in order to evaluate the trustworthiness of the information based on each criterion as follows:

– *Authority* is determined by the expertise of the author, the author's credentials and their institutional affiliations (the name of a public organization).
– *Currency* is calculated as the difference between the time stamp of the creation of the document or the time stamp of submission of the document and the current time. In the case that the system provides the time stamp of the last modification of the document, we will consider this metadata in our model.
– *Accuracy* is estimated according to the number of errors that appear in the document. Specifically, we measure the percentage of words which are spelled incorrectly or are part of a grammatically incorrect statement.
– *Relevance* is determined by the proportion of the users' search terms which are present in the key areas namely the title and the abstract of the publication. In addition, we will adopt the ontology concept for finding related terms and then use these terms to match in the key area in the future work. This allows the model to match the relevant data more efficiently than matching only exact keywords in the key areas of the document.

3.3 Architecture

The content trust model consists of three main modules. Specifically, the input, trustworthiness criteria and metacollection, and output modules, as shown in

Figure 2. Firstly, the input module gathers information from the Web page according to the user's keywords. In addition, our model considers the context and time frame as input from the environment to the model because evaluating trustworthiness according to a specific context. Also, the time frame has an affect on the judgement of whether or not to trust information (information that was trusted in the past is not guaranteed to be trusted in the future). Our model takes the time frame into account, since it evaluates the trustworthiness of the information every time the user interacts with the system. This means the system obtains the most recent information at the time at which the evaluation is performed. In other words, the time frame is automatically supplied to system. Then, the metacollection function will extract metadata provided alongside the content and aggregate the relevant information of this metadata from Semantic Web data sources based on the factors in each criterion from the trustworthiness criteria and the context from the input module. The trustworthiness criteria define the metadata that the system should collect for each criterion. Finally, the output module displays the result by presenting those collected metadata in simple sentences. The aim of this module is to provide the needed information to support users' judgement of the trustworthiness of the information on the Web. As trust is a very subjective notion, it is important to present information about trustworthiness in meaningful ways to the user. This is a complex problem which will be an important part of our future work.

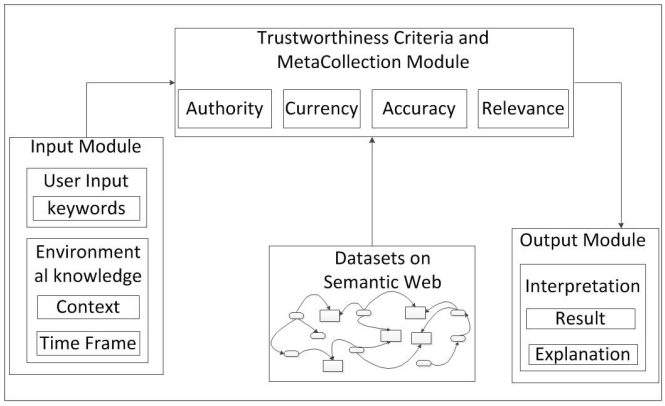

Fig. 2. The Content Trust Model Architecture

4 Methodology

Our methodology for evaluating the trustworthiness/credibility of Web content using Semantic Web technologies has three main phases.

In the first phase, we will build our data model to be used in the system. We will use an RDF graph to represent the data model, called a metadata

graph. This metadata is gathered from information provided alongside the content (explicit metadata) along with information from a Semantic Web data store (implicit metadata). In more detail, the explicit metadata is the data that is published as RDF, alongside the Web page itself. It can be parsed to obtain metadata about the information on the page. In contrast, the implicit metadata is obtained by a query to the Semantic Web data store. We submit a query to the data store to search for more information based on the explicit metadata.

In the second phase, based on the metadata graph, we provide an explanation of this metadata, in order to present it in an easily understandable format. We give the information corresponding to each criterion we presented in Section 3.2. The explanation process will query the data from the metadata graph based on the criteria. Then, the system will interpret these metadata as simple sentences which are then presented to the users as an explanation. As a result, each criterion produces an explanation which is combined with the explanations from other criteria. In the initial implementation, we limit the Web content and Web information to academic publications from the University of Southampton to show that we can collect explicit and implicit data to evaluate the trustworthiness of Web content. We will develop a browser plug-in to present the explanation to users when they browse the Web. We also consider the case in which metadata required for a criterion is not available, in which case we will use other data in the content to help in assessing trustworthiness.

In the third phase, we will evaluate our approach by conducting user based test cases to test our model and conduct a survey to evaluate our system. We will use the publications of the Web Science Conference to be a test case because it is a new conference which has only been established in recent years. Therefore, there might be less information for evaluating the trustworthiness of the publications since we have little background knowledge of the authors who are new researchers in this research area. In addition, the Web Science Conference is a conference for a new, challenging research area which integrates several domains. For this reason, it covers a wide area of research which provides a large range of information. We will evaluate our model by using the questionnaire methodology from the expertise and the general user. For the expert evaluation, we will set up the experiment by choosing the information resources to be evaluated. We will ask the experts to evaluate the trustworthiness of the information. Then, we will compare the result from the experts and the result from our system. For a general user, we will build a questionnaire system to ask the user about their satisfaction with the system when taking part in the group of experiments.

5 Conclusion

In this paper, we proposed a trust model to solve the problem of evaluating the trustworthiness of the information on the Web from available metadata using Semantic Web technologies. We proposed a method to collect the explicit and implicit metadata to build a metadata graph and present this metadata, which will enable the users to judge whether this information is trustworthy. The results

that have been obtained so far show that, by using Semantic Web technologies, we can retrieve relevant data and evaluate the trustworthiness of information based on its information credibility. For the first phase of the PhD work, we can build a metadata graph focusing on the authority and currency criteria and present these metadata to users. In the next step, we will work on dealing with the missing metadata and evaluating our approach.

References

1. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific american 284(5), 34–43 (2001)
2. Bizer, C., Cyganiak, R.: Quality-driven information filtering using the WIQA policy framework. Web Semantics: Science, Services and Agents on the World Wide Web 7(1), 1–10 (Jan 2009)
3. Fogg, B.J., Marshall, J., Osipovich, A., Varma, C., Laraki, O., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M.: Elements that Affect Web Credibility: Early Results from a Self-Report Study. In: Human factors in computing systems. pp. 287–288 (2000)
4. Gil, Y., Artz, D.: Towards content trust of web resources. Web Semantics: Science, Services and Agents on the World Wide Web 5(4), 227–239 (Dec 2007)
5. Gil, Y., Ratnakar, V.: Trusting information sources one citizen at a time. The 1st International Semantic Web Conference pp. 162–176 (2002)
6. 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 pp. 116–131 (2004)
7. Golbeck, J., Parsia, B.: Trust network-based filtering of aggregated claims. International Journal of Metadata, Semantics and Ontologies 1(1), 58–65 (2006)
8. Heath, T., Motta, E.: Reviews and ratings on the semantic web. In: The 5th International Semantic Web Conference (ISWC2006). pp. 7–8 (2006)
9. Heath, T., Motta, E., Petre, M.: Computing word-of-mouth trust relationships in social networks from semantic web and web2. 0 data sources. In: Proceedings of the Workshop on Bridging the Gap between Semantic Web and Web. vol. 2, pp. 44–56. Citeseer (2007)
10. RDF Working Group: Resource Description Framework (RDF) (2004), `http://www.w3.org/RDF/`
11. Rieh, S., Belkin, N.: Understanding judgment of information quality and cognitive authority in the WWW. In: Proceedings of the 61st annual meeting of the american society for information science. vol. 35, pp. 279–89. Citeseer (1998)
12. W3C: RDF Primer (2004), `http://www.w3.org/TR/2004/REC-rdf-primer-20040210/`
13. Wang, W., Zeng, G., Tang, D.: Using evidence based content trust model for spam detection. Expert Systems with Applications 37(8), 5599–5606 (Aug 2010)
14. Wang, X., Zhang, F.: A New Trust Model Based on Social Characteristic and Reputation Mechanism for the Semantic Web. First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008) pp. 414–417 (Jan 2008)
15. Wathen, C.N., Burkell, J.: Believe it or not: Factors influencing credibility on the Web. Journal of the American Society for Information Science and Technology 53(2), 134–144 (2002)