Trust Evaluation through User Reputation and Provenance Analysis

Davide Ceolin, Paul Groth, Willem Robert van Hage,
Archana Nottamkandath, and Wan Fokkink
{d.ceolin,p.t.groth,w.r.van.hage,
a.nottamkandath,w.j.fokkink}@vu.nl

VU University, Amsterdam, The Netherlands

Abstract. Trust is a broad concept which, in many systems, is reduced to reputation estimation. However, reputation is just one way of determining trust. The estimation of trust can be tackled from other perspectives as well, including by looking at provenance. In this work, we look at the combination of reputation and provenance to determine trust values. Concretely, the first contribution of this paper is a standard procedure for computing reputation-based trust assessments. The second is a procedure for computing trust values based on provenance information, represented by means of the W3C standard model PROV. Finally, we demonstrate how merging the results of these two procedures can be beneficial for the reliability of the estimated trust value.
We evaluate our procedures and hypothesis by estimating and verifying the trustworthiness of the tags created within the *Waisda?* video tagging game, launched by the Netherlands Institute for Sound and Vision. Within *Waisda?*, tag trustworthiness is estimated on the basis of user consensus. Hence, we first provide a means to represent user consensus in terms of trust values, and then we predict the trustworthiness of tags based on reputation, provenance and a combination of the two. Through a quantitative analysis of the results, we demonstrate that using provenance information is beneficial for the accuracy of trust assessments.

Keywords: Trust, Provenance, Subjective Logic, Machine Learning, Uncertainty Reasoning, Tags

1 Introduction

From deciding the next book to read to selecting the best movie review, we often use the reputation of the author to ascertain the trust in the thing itself. Reputation is an important mechanism in our set of strategies to determine trust. However, we may base our assessment on a variety of other factors as well, including prior performance, a guarantee, or knowledge of how something was produced. Nevertheless, many systems, especially on the Web, choose to reduce trust to reputation estimation and analysis alone. In this work, we take a multi-faceted approach. We look at trust assessment of Web data based on reputation,

provenance (i.e., how data has been produced), and the combination of the two. We use the term "trust" for the trust in information resources and "reputation" for the trust in agents (see the work of Artz and Gil [1] for complete definitions).

We know that over the Web "anyone can say anything about any topic" [24], and this constitutes one of the strengths of the Semantic Web (and of the Web in general), since it brings democracy in it (everybody has the same right to contribute) and does not prevent a priori any possible useful contribution. However, this principle brings along trust concerns, since the variety of the contributors can affect both the quality and the trustworthiness of the data. On the other hand, the fact that the Semantic Web itself offers the means to, and is putting more effort in recording provenance information, is beneficial to solve this issue. Our contribution is therefore important for two reasons: first, we propose procedures for computing trust assessments of (Semantic) Web data, and some of these procedures have provenance information already available over the Web. Second, by showing that trust assessments based on combinations of reputation and provenance are more accurate than those based only on reputation, we show how a solution to trust issues can be found on the Web itself.

We first propose a procedure for computing reputation that uses basic evidential reasoning principles and is implemented by means of subjective logic opinions [13]. Secondly, we propose a procedure for computing trust assessments based on provenance information represented in the W3C PROV model [23]. Here, PROV plays a key role, both because of the availability of provenance data over the Web recorded by using this standard, and because of its role of interchange format: having modeled our procedure on PROV, then any other different input format can be easily treated after having mapped it to PROV. We implement this procedure by discretizing the trust values and applying support vector machine classification. Finally, we combine these two procedures in order to maximize the benefit of both. The procedures are evaluated on data provided by the *Waisda?*[8] tagging game[1], where users challenge each other in tagging videos. If the tags of two or more users regarding the same video are matched within a given time frame, they both get points. User consensus about tags correlates with tag trustworthiness: the more users agree on a given tag, the more likely it is that the tag is correct. We show how it is possible to predict tag consensus based on who created the tag, how it was created and a combination of the two. In particular, we show that a reputation-based prediction is not significantly different from a provenance-based prediction and, by combining the two, we obtain a small but statistically significant improvement in our predictions. We also show that reputation- and provenance-based assessments correlate.

The rest of the paper is organized as follows: Section 2 describes related work, Section 3 describes the dataset used for our evaluations, Section 4, 5, 6 introduce respectively the trust assessment procedures based on reputation, provenance and their combination, including example associated experiments. Section 7 provides final conclusions.

[1] A zip file containing the R and Python procedures used, together with the dataset, is retrievable at `http://d.pr/f/YXoS`

2 Related work

Trust is a widely explored topic within a variety of computer science areas. Here, we focus on those works directly touching upon the intersection of trust, provenance, Semantic Web and Web. We refer the reader to the work of Sabater and Sierra [21], Artz and Gil [1], and Golbeck [10] for comprehensive reviews about trust in respectively artificial intelligence, Semantic Web and Web. The first part of our work focuses on reputation estimation and is inspired by the works collected by Masum and Tovey [15]. Pantola et al. [16] present reputation systems that measure the overall reputation of the authors based respectively on the quality of their contribution and the "seriousness" of their ratings; Javanmardi et al. [12] measure reputation based on user edit patterns and statistics. Their approaches are similar to ours, but these contributions are particularly tailored for wikis. The second part of our work focuses on the usage of provenance information for estimating trust assessments. In their works, Bizer and Cyganiak [2], Hartig and Zhao [11] and Zaihrayeu et al. [27], use provenance and background information expressed as annotated or named graphs [4] to produce trust values. We do not make use of annotated or named graph, but we use provenance graphs as features for classifying the trustworthiness of artifacts. The same difference is valid also with respect to two works of Rajbhandari et al. [20,19], where they quantify the trustworthiness of scientific workflows and they evaluate it by means of probabilistic and fuzzy models. The use of provenance information for computing trust assessments has also been investigated in a previous work of ours [5] where we determined the trustworthiness of event descriptions based on provenance information by applying subjective logic [13] to provenance traces of event descriptions. In the current paper, we still represent trust values by means of subjective opinions, but trust assessments are made by means of support vector machines, eventually combined with reputations, again represented by means of subjective opinions. Finally, the procedure introduced in Section 4 is a generalization of the procedure that we implemented in a precedent work [6], where we evaluated the trustworthiness of tags of the Steve.Museum [22] artifact collection.

3 The *Waisda?* dataset

Waisda? is a video tagging gaming platform launched by the Netherlands Institute for Sound and Vision in collaboration with the public Dutch broadcaster KRO. The game's logic is simple: users watch video and tag the content. Whenever two or more players insert the same tag about the same video in the same time frame (10 sec., relative to the video), they are both rewarded. The number of matches for a tag is used as an estimate of its trustworthiness. When a tag which is not matched by others is not considered to be untrustworthy, because, for instance, it can refer to an element of the video not noticed so far by any user, or it can belong to a niche vocabulary, so it is not necessarily wrong. In the game, when counting matching tags, typos or synonymity are not taken into consideration.

We validate our procedures by using them to estimate the trustworthiness of tag entries produced within the game. Our total corpus contains 37850 tag entries corresponding to 115 tags randomly chosen. These tag entries correspond to about 9% of the total population. We have checked their representativity of the entire dataset. First, we compared the distribution of each relevant feature that we will use in Section 5 in our sample with the distribution of the same feature in the entire dataset. A 95% confidence level Chi-squared test [18] confirmed that the hour of the day and the day of the week distribute similarly in our sample and in the entire dataset. The typing duration distributions, instead, are significantly different according to a 95% confidence level Wilcoxon signed-rank test [26]. However, the mode of the two distributions are the same, and the mean differs only 0.1 sec. which, according to the KLM-GOMS model [3], corresponds, at most, to a keystroke. So we conclude that the used sample is representative of the entire data set. A second analysis showed that, by randomly selecting other sets of 115 tags, the corresponding tag entries are not statistically different from the sample that we used. We used 26495 tag entries (70%) as a training set, and the remaining 11355 (30%) as a test set.

4 Computing user reputation

Reputation is an abstraction of a user identity that quantifies his reliability as artifact author. Here, we use it to estimate the trustworthiness of the artifact.

4.1 Procedure

We present a generic procedure for computing the reputation of a user with respect to a given artifact produced by him or her.

```
proc reputation(user, artifact) ≡
  evidence := evidence_selection(user, artifact)
  weighted_evidence := weigh_evidence(user, artifact, evidence)
  reputation := aggregate_evidence(weighted_evidence)
```

Evidence Selection Reputation is based on historical evidence, hence the first step is to gather all pieces of evidence regarding a given person and select those relevant for trust computation. Typical constraints include temporal (evidence is only considered within a particular time-frame) or semantics based (evidence is only considered when is semantically related to the given artifact). *evidence* is the set of all evidence regarding *user* about *artifact*.

```
proc evidence_selection(user, artifact) ≡
  for i :=1 to length(observations) do
      if observations[i].user = user then evidence.add(observation[i]) fi
```

Evidence Weighing Given the set of evidence considered, we can decide if and how to weigh its elements, that is, whether to count all the pieces of evidence

as equally important, or whether to consider some of them as more relevant. This step might be considered as overlapping with the previous one since they are both about weighing evidence: evidence selection gives a boolean weight, while here a fuzzy or probabilistic weight is given. However, keeping this division produces an efficiency gain, since it allows computation to be performed only on relevant items.

```
proc weigh_evidence(user, artifact, evidence) ≡
  for i := 1 to length(evidence) do
      weighted_evidence.add(weigh(evidence[i], artifact))
```

Aggregate evidence Once the pieces of evidence (or observations) have been selected and weighed, these are aggregated to provide a value for the user reputation that can be used for evaluation. We can apply several different aggregation functions, depending on the domain. Typical functions are: *count*, *sum*, *average*. Subjective logic [13], a probabilistic logic that we use in the application of this procedure, aggregates the observations in subjective opinions about artifacts being trustworthy based on the reputation of their authors are represented as follows: $\omega(b, d, u)$ where

$$b = \frac{p}{p+n+2} \quad d = \frac{n}{p+n+2} \quad u = \frac{2}{p+n+2}$$

where b, d and u indicate respectively how much we believe that the artifact is trustworthy, non-trustworthy, and how uncertain our opinion is. p and n are the amounts of positive and negative evidence respectively. Subjective opinions are equivalent to Beta probability distributions (Fig. 1), which range over the trust levels interval $[0 \dots 1]$ and are shaped by the available evidence.

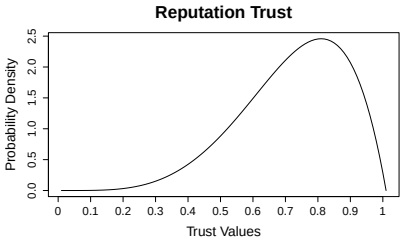

Fig. 1. Example of a Beta probability distribution aggregating 4 positive and 1 negative evidence. The most likely trust value is 0.8 (which is the ratio among the evidence). The variance of the distribution represents the uncertainty about the evaluation.

4.2 Application Evaluation

First, we convert the number of matches that each tag entry has into trust values:

tag selection For each tag inserted by the user, we select all the matching tags belonging to the same video. In other contexts, the number of matching tags can be substituted by the number of "likes", "retweets", etc..

tag entries weighing For each matching entry, we weigh the entry contribution on the time distance between the evaluated entry and the matched entry. The weight is determined from an exponential probability distribution, which is a "memory-less" probability distribution used to describe the time between events. If two entries are close in time, we consider it highly likely that they match. If they match but appear in distant temporal moments, then we presume they refer to different elements of the same video. Instead of choosing a threshold, we give a probabilistic weight to the matching entry. 85% of probability mass is assigned to tags inserted in a 10 sec. range.

tag entries aggregation In this step, we determine the trustworthiness of every tag. We aggregate the weighed evidence in a subjective opinion about the tag trustworthiness. We have at our disposal only positive evidence (the number of matching entries). The more evidence we have at disposal for the same tag entry, the less uncertain our estimate of its trustworthiness will be. Non-matched tag entries have equal probability to be correct or not;

We repeat this for each entry created by the user to compute his reputation.

user tag entries selection Select all the tag entries inserted by *user*.

user tag entries weighing Tag entries are weighed by the corresponding trust value previously computed. If an entry is not matched, it is considered as half positive (trust value 0.5) and half negative (1-0.5 = 0.5) evidence (it has 50% probability to be incorrect), as computed by means of subjective opinions. The other entries are also weighed according to their trust value. So, user reputation can either rise or decrease as we collect evidence.

user tag entries aggregation In turn, to compute the reputation of a user with respect to a given tag, we use all the previously computed evidence to build a subjective opinion about the user. This opinion represents the user reputation and can be summarized even more by the corresponding expected value or trust value (a particular average over the evidence count).

4.3 Results

We implement the abstract procedure for reputation computation and we evaluate its performance by measuring its ability to make use of the available evidence to compute the best possible trust assessment. Our evaluation does not focus on the ability to predict the exact trust value of the artifact by computing the user reputation, because these two values belong to a continuous space, and they are computed on a different basis. What we expect is that these two values hint at trustworthiness in a similar fashion: when a tag is trustworthy, then both trust value and reputation should be higher than a certain threshold and vice-versa. The validation, then, depends upon the choice of the threshold. We run the procedure with different thresholds as presented in Fig.3. Low thresholds correspond to low accuracy in our predictions. However, as the threshold increases, the accuracy of the prediction rises. Moreover, we should consider that: (1) It is preferable to obtain "false negatives" (reject correct tags) rather than

"false positives" (accept wrong tags), so high thresholds are more likely to be chosen (e.g., see [9]), in order to reduce risks; (2) A Wilcoxon signed-rank test at 95% confidence level proved that the reputation-based estimates outperform blind guess estimates (having average probability of accuracy 50%). The average improvement is 8%, the maximum is 49%.

We previously adopted this same procedure to compute the trustworthiness of tags on the Steve.Museum artifacts [6]. Having to adapt it to the *Waisda?* case, we could understand the prominent features of it, hence this helped us in formulating the general procedure above.

5 Computing provenance-based trust

We focus on the "how" part of provenance, i.e., the modality of production of an artifact. (For simplicity, in the rest of the paper, we will use the word "provenance" to refer to the "how" part of it). We learn the relationships between PROV and trust values through machine learning algorithms. This procedure allows to process PROV data and, on the basis of previous trust evaluations, predict the trust level of artifacts. PROV is suitable for modeling the user behavior and provenance information in general.

5.1 Procedure

We present the procedure for computing trust estimates based on provenance.

proc provenance_prediction($artifact\_provenance, artifact$) $\equiv$
  $attribute\_set := attribute\_selection(artifact\_provenance)$
  $attributes := attribute\_extraction(attribute\_set)$
  $trust\_levels\_aggregation$
  $classified\_testset := classify(testset, trainingset)$

attribute_selection Among all the provenance information, the first step of our procedure chooses the most significant ones: agent, processes, temporal annotations and input artifacts can all hint at the trustworthiness of the output artifact. This selection can lead to an optimization of the computation.

attribute_extraction Some attributes need to be manipulated to be used for our classifications, e.g., temporal attributes may be useful for our estimates because one particular date may be particularly prolific for the trustworthiness of artifacts. However, to ease the recognition of patterns within these provenance data, we extract the day of the week or the hour of the day of production, rather than the precise timestamp. In this way we can distinguish, e.g., between day and night hours (when the user might be less reliable). Similarly, we might refer to process types or patterns instead of specific process instances.

trust_level_aggregation To ease the learning process, we aggregate trust levels in n classes. Hence we apply classification algorithms operating on a nominal scale without compromising accuracy.

classification Machine learning algorithms (or any other kind of classification algorithm) can be adopted at this stage. The choice can be constrained either from the data or by other limitations.

5.2 Application evaluation

We apply the procedure to the tag entries from the *Waisda?* game as follows.

attribute selection and extraction The provenance information available in *Waisda?* is represented in Fig. 2, using the W3C PROV ontology. First, for

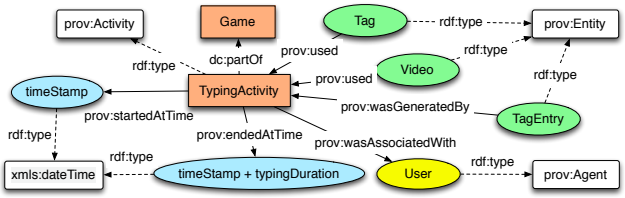

Fig. 2. Graph representation of the provenance information about each tag entry.

each tag entry we extract: *typing duration*, *day of the week*, *hour of the day*, *game_id (to which the tag entry belongs)*, *video_id*. This the "how" provenance information at our disposal. Here we want to determine the trustworthiness of a tag given the modality with which it was produced, rather than the author reputation. Some videos may be easier to annotate than others, or, as we mentioned earlier, user reliability can decrease during the night. For similar reasons we use all the other available features.

trust level classes computation In our procedure we are not interested in predicting the exact trust value of a tag entry. Rather we want to predict the range of trust within which the entry locates. Given the range of trust values $[0 \ldots 1]$, we split it into 20 classes of length 0.5: from $[0 \ldots 0.5]$ to $[9.5 \ldots 10]$. This allows us to increase the accuracy of our classification algorithm without compromising the accuracy of the predicted value or the computation cost. The values in each class were approximated by the middle value of the class itself. For instance, the class $[0.5 \ldots 0.55]$ are approximated as 0.525.

regression/classification algorithm We use a regression algorithm to predict the trustworthiness of the tags. Having at our disposal five different features (in principle, we might have more), and given that we are not interested in predicting the "right" trust value, but the class of trustworthiness, we adopt the "regression-by-discretization" approach [14], that allows us to use Support Vector Machines algorithm (SVM) [7] to classify our data. The training set is composed by 70% of our data, and then we predict the trust level of the test set. We used the SVM version implemented in the e1071 R library [25]. In the future, we will consider alternative learning techniques.

5.3 Results

The accuracy of our predictions depends on the choice of the thresholds. If we look at the ability to predict the right (class of) trust values, then the accuracy is of about 32% (which still is twice as much as the average result that we would have with a blind guess), but it is more relevant to focus on the ability to predict the trustworthiness of tags within some range, rather than the exact trust value. Depending on the choice of the threshold, the accuracy in this case varies in the range of 40% - 90%, as we can see in Fig. 3. For thresholds higher than 0.85 (the most likely choices), the accuracy is at least 70%. We also compared the provenance-based estimates with the reputation-based ones, with a 95% confidence level Wilcoxon signed-rank test that proved that the estimates of the two algorithms is not statistically different. *For the Waisda? case study, reputation- and provenance-based estimates are equivalent: when reputation is not available or it is not possible to compute it, we can substitute it with provenance-based estimates.* This is particularly important, since the ever growing availability of PROV data will increase the ease for computing less uncertain trust values.

If we apply the "regression-by-discretization" approach for making provenance-based assessments, then we approximate our trust values. This is not necessary with the reputation approach. Had we applied the same approximation to the reputations as well, then provenance-based trust would have performed better, as proven with a 95% confidence level Wilcoxon signed-ranked test, because reputation can rely only on evidence regarding the user, while provenance-based models can rely on larger data sets. Anyway, we have no need to discretize the reputation and, in general, we prefer it for its lightweight computational burden.

6 Combining reputation and provenance-based trust

We combine reputation- and provenance-based estimates to improve our predictions. If a certain user has been reliable so far, we can reasonably expect him/her to behave similarly in the near future. So we use reputation and we also constantly update it, to reduce the risk on relying on over-optimistic assumptions (if a user that showed to be reliable once, will maintain his/her status forever). However, reputation has an important limitation. To be reliable, a reputation has to be based on a large amount of evidence, which is not always possible. So, both in case the reputation is uncertain, or in case the user is anonymous, other sources of information should be used in order to correctly predict a trust value. The trust estimate based on provenance information, as described in Section 5, is based on behavioral patterns which have a high probability to be shared among several users. Hence, if a reputation is not reliable enough, we substitute it with the provenance-based prediction.

6.1 Procedure

The algorithm looks like the following:

proc provenance_prediction($user, artifact$) $\equiv$
$q\_ev = evaluate\_user\_evidence(user, artifact)$
if $q\_ev > min\_evidence$ then $predict\_reputation$ else $predict\_provenance$ fi

evaluate_user_evidence This function quantifies the evidence. Some implementation examples: (1) *count*; (2) compute a subjective opinion and check if the uncertainty is low enough. As future work we plan to investigate how to automatically determine q_ev and $evaluate\_user\_evidence$.

6.2 Application evaluation

We adopted the predictions obtained with each of the two previous procedures. The results are combined as follows: if the reputation is based on a minimum number of observations, then we use it, otherwise we substitute it with the prediction based on provenance. We run this procedure with different values for both the threshold and the minimum number of observations per reputation. We instantiate the *evaluate_user_evidence(user,artifact)* function as a *count* function of the evidence of *user* with respect to a given *tag*.

6.3 Results

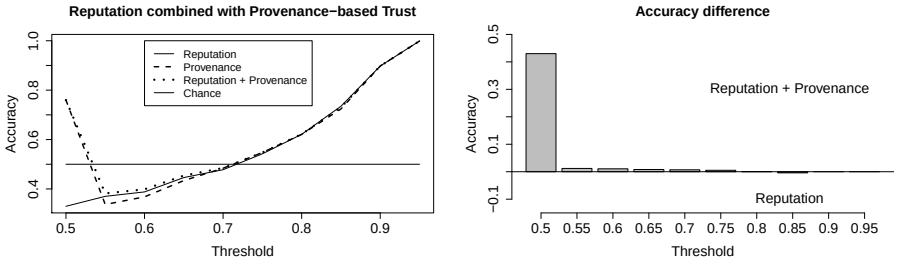

Fig. 3. Absolute and relative (Reputation+Provenance vs. Reputation) accuracy. The gap between the prediction (provenance-based) and the real value of some items explains the shape between 0.5 and 0.55: only very low or high thresholds cover it.

The performance of this algorithm depends both on the choice of the threshold for the decision and on the number of pieces of evidence that make a reputation reliable, so we ran the algorithm with several combinations of these two parameters (Fig. 3). The results converge immediately, after having set the minimum number of observations at two. We compared these results with those obtained before. Two Wilcoxon signed-rank tests (at 90% and 95% confidence level with respect to respectively reputation and provenance-based assessments) showed that *the procedure which combines reputation and provenance evaluations*

in this case performs better than each of them applied alone. The improvement is, on average, about 5%. Despite the fact that most of the improvement regards the lower thresholds, which are less likely to be chosen (as we saw in Section 4), even at 0.85 threshold there is a 0.5% improvement. Moreover, we would like to stress how the combination of the two procedures performs better than (in a few cases, equal to) each of them applied alone, regardless of the threshold chosen.

Combining the two procedures allows us to go beyond the limitation of reputation-based approaches. Substituting estimates based on poorly reliable reputations with provenance-based ones improves our results without significantly increasing our risks, since we have previously proven that the two estimates are (on average) equivalent. Hence, when a user is new in a system (and so his/her history is limited) or anonymous, we can refer to the provenance-based estimate to determine the trustworthiness of his/her work, without running higher risks. This improvement is at least partly due to the existing correlation between the reputation and provenance-based trust assessments. A little positive correlation (0.16) has been proved by a Pearson's correlation test [17] with a confidence level of 99%. Thanks to this, we can safely enough substitute uncertain reputations with the corresponding provenance-based assessments. This explains also the similarity among the results shown in Fig. 3.

7 Conclusion

This paper explores two important components of trust assessments: reputation and provenance information. We propose and evaluate a procedure for computing reputation and one for computing trust assessments based on provenance information represented with the W3C standard PROV. We show that it is important to use reputation estimation for trust assessment, because it is simple, computationally light and accurate. We also show the potential of provenance-based trust assessments: these can be at least as accurate as reputation-based ones and can be used to overcome the limitations of a reputation based approach. In *Waisda?* the combination of the two methods revealed to be more powerful than each of the two alone. In the future we will investigate the possibility of automatically extracting provenance patterns usable for trust assessment, to automate, optimize and adapt the process to other case studies. We will also focus on the use of trust assessments as a basis for information retrieval.

Acknowledgements We thank the Netherlands Institute for Sound and Vision for launching and guiding the *Waisda?* project, and our colleagues Michiel, Riste and Valentina for their support. This research was partially supported by the PrestoPRIME project, in the EC ICT FP7 program, and by the Data2Semantics and SEALINC Media projects in the Dutch national program COMMIT.

References

1. D. Artz and Y. Gil. A survey of trust in computer science and the semantic web. *Journal of Semantic Web*, 5(2):131–197, 2007.

2. C. Bizer and R. Cyganiak. Quality-driven information filtering using the WIQA policy framework. *Journal Web Semantics*, 7(1):1–10, Jan. 2009.
3. S. Card, T. P. Moran, and A. Newell. *The Psychology of Human Computer Interaction*. Lawrence Erlbaum Associates, 1983.
4. J. Carroll, C. Bizer, P. Hayes, and P. Stickler. Named graphs, provenance and trust. In *WWW '05*, pages 613–622. ACM, 2005.
5. D. Ceolin, P. Groth, and W. R. V. Hage. Calculating the trust of event descriptions using provenance. In *SWPM 2010*. CEUR-WS, 2010.
6. D. Ceolin, A. Nottamkandath, and W. Fokkink. Automated evaluation of annotators for museum collections using subjective logic. In *Trust Management VI*, IFIP AICT 374. Springer, May 2012.
7. C. Cortes and V. Vapnik. Support-vector networks. *M. Learn.*, 20:273–297, 1995.
8. D. Gambetta. *Can We Trust Trust?* Basil Blackwell, 1988.
9. J. Golbeck. Trust on the World Wide Web: A Survey. *Foundations and Trends in Web Science*, 1(2):131–197, 2006.
10. O. Hartig and J. Zhao. Using web data provenance for quality assessment. In *SWPM 2009*. CEUR-WS, 2009.
11. S. Javanmardi, C. Lopes, and P. Baldi. Modeling user reputation in wikis. *Stat. Anal. Data Min.*, 3(2):126–139, Apr. 2010.
12. A. Jøsang. A logic for uncertain probabilities. *Intl. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(3):279–212, 2001.
13. I. Kononenko. Naive bayesian classifier and continuous attributes. *Informatica*, 16(1):1–8, 1992.
14. H. Masum and M. Tovey, editors. *The reputation society*. MIT Press, Feb. 2012.
15. Netherlands Inst. for Sound and Vision. Waisda? `http://waisda.nl`, June 2012.
16. A. V. Pantola, S. Pancho-Festin, and F. Salvador. Rating the raters: a reputation system for wiki-like domains. In *SIN '10*, pages 71–80. ACM, 2010.
17. K. Pearson. Mathematical Contributions to the Theory of Evolution. In *Proceedings of the Royal Society of London*, pages 489–498, 1896.
18. K. Pearson. On the Criterion that a Given System of Deviations from the Probable in the Case of Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling. *Phil. Mag.*, 50:157–175, 1900.
19. S. Rajbhandari, O. F. Rana, and I. Wootten. A fuzzy model for calculating workflow trust using provenance data. In *MG'08*. ACM, 2008.
20. S. Rajbhandari, I. Wootten, A. S. Ali, and O. F. Rana. Evaluating Provenance-based Trust for Scientific Workflows. In *CCGRID 06*, volume 1, pages 365–372. IEEE, 2006.
21. J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24:33–60, 2005.
22. U.S. Institute of Museum and Library Service. Steve Social Tagging Project. `http://www.steve.museum/`, June 2012.
23. W3C. PROV-O. `http://www.w3.org/TR/prov-o//`, June 2012.
24. W3C. Resource description framework (rdf): Concepts and abstract data model. `www.w3.org/TR/2002/WD-rdf-concepts-20020829/`, June 2012.
25. T. Wien. e1071: Misc functions of the department of statistics (e1071). `http://cran.r-project.org/web/packages/e1071/`, June 2012.
26. F. Wilcoxon. Individual comparisons by ranking methods. *Biom. Bull.*, 1:80–83, 1945.
27. I. Zaihrayeu, P. da Silva, and D. L. McGuinness. IWTrust: Improving User Trust in Answers from the Web. In *iTrust2005*, volume 3477, pages 384–392. Springer, 2005.