Supporting Human Decision-Making Online using Information-Trustworthiness Metrics

Jason R.C. Nurse, Sadie Creese, Michael Goldsmith, and Syed S. Rahman

Cyber Security Centre, Department of Computer Science,
University of Oxford, Oxford, UK
{firstname.lastname}@cs.ox.ac.uk

Abstract. The vast amount of information available online places decision makers wishing to use this content in an advantageous but also very difficult position. The advantages stem from the volume of content from a variety of sources that is readily available; the difficulties arise because of the often unknown quality and trustworthiness of the information – is it fact, opinion or purely meant to deceive? In this paper we reflect on and extend current work on information trust and quality metrics which can be used to address this difficulty. Specifically, we propose new metrics as worthy of consideration and the new combinatorics required to take measurements of the various trust factors into a single score. These feed into our existing overarching policy-based approach that uses trustworthiness metrics to support decision-making online.

Key words: information trustworthiness, information quality, metrics, human decision-making, open-source content, social-media, online risks

1 Introduction

The Web is the largest source of information in the world as well as the largest open marketplace and social/political forum. A key challenge with its use as an information source for supporting human decision-making, however, is that the resources that sites and services introduce us to are often unknown to us, so raising questions about their quality and trustworthiness. This is especially true in today's world, where anyone anywhere can share content online, some useful and some meant actively to deceive; crisis situations exemplify this perfectly [1, 2]. As a result of these concerns and the potential utility of online content in a range of cases, there is an increasingly acute need to provide information-confidence and trust measures to users of online (and particularly social-media) information, to support them in making informed decisions, in light of risk.

In this paper, therefore, we reflect on some of the trust metrics currently available and those being researched, with two objectives in mind. Firstly, we aim briefly to review the state-of-the-art, identifying what trustworthiness factors can feasibly be measured and potentially utilised. This review pulls together several seminal papers (e.g., [3, 4, 5, 6]) which introduce techniques for measuring information quality, credibility and trustworthiness. Secondly, we extend

this existing work on metrics and our overarching policy-based approach [7] by proposing new metrics as worthy of consideration and the new combinatorics required to take measurements of the various factors into a single score.

This paper is structured as follows. In Section 2 we review factors that influence trustworthiness and the proposals for measuring them; this builds on our initial work in [8]. Section 3 broadens the scope beyond intrinsic trust factors to consider how to metricise the potential impact of infrastructure vulnerability upon the trustworthiness of the information being communicated over it. In Section 4 we recap our novel policy-based methodology for assigning trustworthiness measures to openly-sourced information. As part of our overall methodology presentation, we also discuss the important problem of how to combine factor measurements/scores appropriately so as to arrive at a single information-trustworthiness score (Section 5); it is this score that is typically (first) presented to a decision-maker. Finally, Section 6 reports on the practical assessment and validation that we have conducted to date on existing and proposed metrics, before concluding and presenting future work in Section 7.

2 State-of-the-art in Information-Trustworthiness Metrics

The complexity of trust is well-known, and undoubtedly derives from its social origins. To elucidate its use for our purposes, in previous work [8] we conducted a state-of-the-art review focusing on which factors influence an individual's perception of information trustworthiness online. The outcome of that study was the definition of some high-level influences, namely the provenance of information and its intrinsic quality, but also a plethora of specific trust factors within these categories. It is these specific factors that are most useful to our research now, as they are candidate properties through which trustworthiness might be measured. The remainder of this section presents several of these trust factors, and reviews proposals for their measurement; we concentrate on techniques that might be automated, and on social-media content, given its increasing prevalence. This presentation is intended to overview the state-of-the-art as it pertains to metrics and set the foundation for our work building on this baseline.

One of the most fundamental trustworthiness factors is the timeliness of the information. This factor operates on the basis that the closer information is published to a specified time, the more likely it is up-to-date and potentially accurate. Overall, the metric (a calculation of the time elapsed) has been seen to work quite well ([3, 9]) and is simple enough to be applied in most domains. Other factors that influence quality and trust include, information's completeness, complexity and relevance. To measure completeness, approaches typically conduct assessments based on expected text/features within the content (e.g., number of internal links on a Wikipedia article [3] or absence of appropriate data [10]) but there is also work that has proposed content length [11]. While not ideal techniques as they fail at judging completeness from a semantic perspective, as those articles have shown, they can be quite useful as a quick initial indicator of trustworthiness in some domains. Complexity/ease-of-understanding is also popular

in the literature (e.g., [4, 6]), undoubtedly because of its link to well-understood readability metrics (e.g., Flesch-Kincaid) which allow automated analysis based on sentences, words and syllables in content. Grammar and spelling may feed into this textual analysis as an indicator of trustworthiness [9, 12] as well. Although useful, these factors should be treated circumspectly considering that the nature of content in some social-media services lends itself to jargon or small snippets of information (e.g., Twitter), which inhibit its proper analysis. Approaches to measuring the relevance of content to the problem at hand have thus far been centred around matching topic keywords to words that appear often and/or are in notable positions in the information items [13]. This is a sensible technique, and indeed is one that search engines apply. Its weakness, however, is that because it only matches words, the semantics and context is overlooked, meaning that irrelevant information may actually be rated as relevant.

Source-related factors such as authority/reputation, recommendation and popularity can influence the trustworthiness of the information as well. To measure reputation, approaches have considered the number of edits made by other authors on a specific author's contributions [14], link analysis and ranking techniques focusing on social relationships [15] and those adapted from Web search [4], and on assessing basic features such as amount of content contributed and number of associates (followers/friends on Twitter) [5]. The metrics for recommendation may overlap with those for reputation, particularly in the social domain, where feedback on input is encouraged. For example, in the question-answer domain (e.g., Yahoo! Answers), highlighting an answer to a question as a 'best answer' affects the author's reputation as well as recommending them and the answer itself. Other work on Twitter has sought to assess retweeting, mentioning and 'favouriting' as indications of trust [16], i.e., as a form of recommendation. These approaches all provide useful techniques for measuring the factor. Finally, metrics for popularity vary across contexts, but largely the aim is to determine how frequently cited a source or page is. On the Web at large, this could mean assessing the number of times a particular site is mentioned (e.g., [17]), but in more social domains, metrics might be based on number of followers, friends or page likes. Albeit a plausible metric, two caveats with it and a few of the others mentioned above are that data is subject to manipulation in hostile situations (e.g., information warfare) and context is crucially important – a source's high popularity or good reputation in one topic may not transfer to another topic. Both are key points to note if metrics are to be broadly applied. There were several factors that we were unable to find published automatic metrics for, including: objectivity, accuracy, believability, corroboration, and location of source. In upcoming sections, we will follow up on some of these.

3 Accounting for Information Infrastructure Vulnerability

In addition to assessing intrinsic trustworthiness factors, we hypothesise that trustworthiness of information may be impacted by attempts to damage its integrity, as enabled by vulnerabilities in the technology of the infrastructure over

which it travels. Further, that assessing exposure to such risk might be a useful component of an information-trustworthiness measure, although we recognise that such a metric is likely to be context-dependent in its utility, probably only useful in cases where there is a perception of real, present threat that could target an exploitable vulnerability in the information infrastructure.

To assess the extent to which vulnerability in technologies could result in information corruption, we start by outlining an exposure model: *Exposure = Threat* $\times$ *Vulnerability*. Exposure in this context reflects the possibility that a piece of information has been corrupted, and is considered as resulting from the presence of a motivated *threat* combined with an exploitable *vulnerability*. This is based on current practice in information-risk assessment and is inspired by [18]. The Information Infrastructure Vulnerability (IIV) factor measure is this generic Exposure value detailed with the following factors.

Threats are those entities that perpetrate attacks and, in the context of our work, this means entities that deliberately seek to corrupt information or are reckless as to whether it is corrupted as 'collateral damage' in a broader attack. In general, we can gather evidence on the presence of threat by monitoring various open sources (e.g., CERT announcements, BlackHat discussion boards and security bulletins). In these cases, it is imperative to establish whether a threat actor is motivated and if they have the capability to attack. Given that we are interested in exploitable technology *vulnerabilities*, the factors to be measured by this metric should be driven by an understanding of attacks as well as nature of vulnerability. Literature has proposed several taxonomies of attacks and related vulnerabilities (see [19]). From these, we have identified particularly relevant areas for our vulnerability factor measurement to be: applications; system architectures and platforms; communication networks; and hardware aspects. Establishing the presence of vulnerabilities in these domains (and indeed, specified technology infrastructures) would then be based on a (likely automated) check against existing data stores (e.g., CERT Vulnerability Notes Database, Common Vulnerabilities and Exposures, National Vulnerability Database, BUGTRAQ).

Our general metric for IIV is therefore based on the following. Firstly, a user organisation or individual tool user will need to maintain a general threat table, which they probably already do as part of ongoing risk-management activities, which defines types of threat (t) (e.g., state-sponsored action, hackers, untargeted and opportunistic attackers) and the probability (p_t) of them being motivated to attack. This table of values is defined as p_t as t varies over threats.

Now, assume a new infrastructure vulnerability V_i is published online and has been picked up by the user's vulnerability monitoring tools: **Step 1**: Determine the likelihood that the vulnerability V_i exists within their local technology infrastructure. This can be determined by automated or manual checks against current infrastructure and configurations; the resulting value defined as $v_i : 0 < v_i < 1$. **Step 2**: Estimate which threats that can exploit V_i and their probabilities of exploiting it. This value is defined as e_{it}. **Step 3**: Calculate the probability that V_i will be exploited by the defined threats in order to conduct an attack against the tool user. This value is approximated as a_i with the for-

mula $a_i = v_i.(1 - \prod_t (1 - e_{it}.p_t))$; this treats the threats as independent and so may overestimate a_i. **Step 4**: Calculate the IIV score based on an aggregation of a_i values, i.e., one for each of the vulnerabilities identified. As such: $IIV = 1 - \prod_i (1 - a_i)$. This IIV score can then be applied to amplify/attenuate the other core trustworthiness elements; in fact, we consider the complement, Information Infrastructure Integrity (III): $III = 1 - IIV$, so as less vulnerable situations score higher. We conclude our discussion here due to space limitations but further detail is available in [19].

4 Policy-based Approach to Factor Application

To be able to appropriately apply the range of metrics discussed above to support decision-making, there is a need for a broad approach that incorporates the differing importance of factors to users, decision context and, crucially, the combinatorics behind how factors' metrics are combined into a single trustworthiness score. In previous work [7], we have introduced such a methodology based on policies by which this could be achieved. Below we briefly recap that approach and then focus on the newly detailed aspect, i.e., the factor combinatorics.

In the first instance, we assume that there is a set of information from online sources (e.g., news feeds, Twitter or Facebook posts) to be used in making a decision. The first level of processing therefore takes this information and applies a high-level policy to it to filter unwanted sources and to ensure that content from trusted sources is included. Next, users or organisations can once more specify through policy what specific provenance, quality and generally, trustworthiness factors (e.g., the competence or reputation of the source, information's recency) are to be considered. This policy also sets a basic scoring scheme defining the importance (weightings) to be given to each factor. Selection of factors will depend upon context, since their reliability as a trustworthiness indicator may vary according to their perceived inherent value (such as location in a disaster situation) as well as their vulnerability to compromise and the likelihood of a compromise (malicious or accidental) taking place. A user's own decision-making policy also impacts factors. This is a third policy level that is intended to allow the decision maker to amplify or attenuate one or more factors that may appear to be more or less important, in their opinion, by increasing or decreasing the weight given to the corresponding score amongst all information. These policies could be based on a user's experience and intuitions, but we allow for the definition of preset templates which can be shared and refined over time.

5 Factor Combinatorics

A part of properly applying the factors is considering how to appropriately combine their measurements/scores to arrive at a single trustworthiness value. For example, if we have scores for recency, popularity and III of 0.13, 0.22 and 0.88 respectively, how do we produce one value which represents the overall

trustworthiness of the associated information item? Some articles in the literature have considered this problem, but most (see [10]) seem to opt for simple weighted means, without due consideration of other approaches. Our exploration of this problem consisted of a theoretical evaluation of several methods by which values could be combined and averaged or a single representative value obtained. We considered 6 methods: Arithmetic mean (AM) ($\frac{1}{n}\sum_{i=1}^{n} x_i$), Geometric mean (GM) ($(\Pi_{i=1}^{n} x_i)^{\frac{1}{n}}$), Quadratic mean (QM) ($(\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$), Harmonic mean (HM) ($n/\sum_{i=1}^{n}\frac{1}{x_i}$), Square-Mean-Root (SMR) ($(\frac{1}{n}\sum_{i=1}^{n}\sqrt{x_i})^2$), Conjugated Root-Mean-Squared (CRMS) ($1-(\frac{1}{n}\sum_{i=1}^{n}(1-x_i)^2)^{\frac{1}{2}}$). These cover a number of popular techniques applied to similar mathematical combination problems. Simple product ($\prod_{i=1}^{n} x_i$, as used for *III*) is rejected as antitonic in n.

The criteria used for assessing these approaches included fairness (i.e., lack of bias towards higher or lower scores), rigour (working well across ranges of values while behaving in a consistent manner, e.g., in terms of ordering compared to other methods), and the ability to add weights to factors (thus allowing users to (de)emphasise factor importance in the calculation of the overall trustworthiness score). We are also interested in 0-cases (i.e., how a 0 score for a factor impacts the calculation of the overall score). We present our key findings below; readers should note that we assume normalised values ($[0, 1]$) for trustworthiness factors, as it is easier to provide homogeneous transformations on this space.

Several noteworthy points arose from our evaluation. Firstly, there were some methods that violated the fairness requirement in that they placed too great an emphasis on higher values (as compared to lower ones) or lower values (as compared to higher ones). An example of the former case is QM (emphasises higher input scores), and of the latter case is HM (emphasises lower scores compared to the other means). In terms of consistency and predictability of output, all the means did quite well. We knew from existing work [20] of the ordering of four of these techniques (i.e., $HM \leq GM \leq AM \leq QM$), but to find that the SMR tended to fit between the GM and AM was advantageous considering fairness (i.e., getting a truly representative mean score). The CRMS method did appear to give predictable preliminary results particularly in ordering (typically between GM and AM) but then produced peaks and troughs very dependent on the input data that was fed into it. For example, for some high values (0.9,0.8,1.0) it produced a mean even lower than HM.

Relating to weights, all the formulae performed in line with expectations; a range of simple weights were properly handled and increases in a weight did cause means to gravitate towards the respective input value. One observation is that calculation of Weighted GM can result in numeric underflow at high weights (e.g., greater than ∼500), thus producing an output mean of 0.0. Finally to comment on the 0-case criterion, we found that the HM was unable to gracefully handle factor scores of 0 (resulting in a divide-by-zero error). This meant that the criteria for rigour and 0 case could not be met by this method. The GM function also evaluated to 0 if there were any 0 scores present. This might or might not be preferred by users, depending on whether a terrible score in one factor meant that the values for other factors was irrelevant or compromised.

In general, it is our opinion that the choice of combination method applied should be the user's, as they may have their own perspectives or contexts. A pessimistic user, for example, may always prefer that final trustworthiness scores (i.e., mean values) are underestimated, whereas an optimist, or someone who is not especially sensitive may not be concerned about an overestimation of the trustworthiness scores. With this understanding, we summarise below cases where particular approaches may be most appropriate.

As it relates to fairness, the SMR appears the most suitable as it typically produces a value in the middle of the means considered. This positioning also makes SMR predictable and not sporadic (like CRMS which can at times result in a mean below the HM and GM). If there is a desire for underestimating trustworthiness, users could select either the HM or GM. These generally perform in a consistent manner (ordering) and cover a range of inputs. Users should however be aware of their nuances relating to 0 cases and high weights, as discussed previously. For cases where users prefer to overestimate mean values (or mask lower values), the AM and QM methods may be most appropriate. Furthermore, they stand the test of rigour and appreciation of ranges of factor weights.

Sometimes there may be cases where there are missing factor values. For example, a piece of information may lack a timestamp or there may be insufficient details to determine the *III* measure. For such situations, we allow users to specify a list of 'must have' factors, and if these are not present, we set the factor's value to 0 and assign a high weight to those information items. This allows calculations to be made as before, but ensures that a penalty is paid for missing that value; higher weighting on the 0 value pulls down the final mean. Conversely, if the factor's value is missing and it is not a required factor, it could be ignored by setting a weight of 0.

6 Experimentation and Validation

We engaged in a practical assessment and critical reflection on some trust-factor metrics. Here we discuss four of these, focusing especially on our proposed metrics (see [19] for others). For our experimentation, we used a 2011 London Riots dataset consisting of a broad range of public tweets and news reports.

Competence of Source refers to the level of expertise of an information source [21]. To measure competence, we drew on contributions from partners in the project and their approach to compare the usage of words by the unknown sources to that of a core and predefined set of competent sources. The unique perspective taken by our metric is based on the assumption that competent authors use words that are measured and appropriate in their information contributions, thus it might be possible to count the number of words from the unknown author that match words from the competent author set. A competence score is obtained by normalising the count with respect to total word count, thereby resulting in a value $[0, 1]$. The metric was set up, implemented and deployed against the dataset. Based on the initial results, it performed reasonably well at identifying the more competent sources.

To reflect critically on this metric, although it did work well there are reservations, including: (i) reliance on a core set of competent authors/texts for a topic – this would be difficult to define and challenging to prove that a word-set persists as complete; (ii) difficulty of defining the confines and granularity for a 'competent' word set per topic/field/sub-field; and (iii) simply mentioning words may not be in itself indicative of how competent a source is –an author may be constantly re-tweeting other people's content or publishing highly biased opinions. Moreover, such an approach might also fall victim to keyword stuffing attacks. Future metric refinement will need to address these concerns.

Corroboration – captures the extent to which the same information originates from various unrelated sources [22]. For our purposes, we abstracted this factor to be built on an approach that required that 'similar' information be found across different content items. To do this, we chose to apply a clustering algorithm, via the *Carrot*^{2} [23] clustering engine. This engine parses content items and creates and automatically labels clusters. We then use the labels as the comparison key to obtain similarity counts (e.g., how many other items are within the 'looting and riots in London' label), which are then normalised to provide the factor's corroboration score $[0, 1]$ for each item. From the metric's implementation and testing we were able to gather some descriptive and promising topic clusters.

There were two key weaknesses of this clustering approach to assess corroboration. Firstly, clustering does not natively accommodate negation (words like 'no', 'not') therefore proper checks should be conducted to ensure that negative words/sentiment are adequately handled and that conflicting pieces of information are not clustered together. Secondly, this approach does not conduct checks to verify that information originates from different sources. This is a problem as corroboration of information from the same or closely related sources introduces the possibility of bias therefore negatively impacting the metric.

Social-Media Jargon – this factor posits that excessive use of social-media jargon, especially emoticons and shouting (words fully and inappropriately uppercased), might indicate a lack of information trustworthiness, similar to [9]. Our algorithm emphasises simplicity and thus compares the number of characters used on jargon as opposed to real and potentially useful content. As such, the implemented metric measures two aspects, first, the percentage of characters within the content used for emoticons (assessed using regular expressions) and second the percentage of words fully capitalised (excluding well-known abbreviations). These two values are then averaged to deduce a score, which is then subtracted from 1 to get the 'goodness' score for the factor.

By this simple metric, one is able to conduct a very basic assessment of information. When we evaluated it with the dataset, the findings were encouraging but not conclusive. The issues that arose included, the dependence on an up-to-date emoticon list to compare against, in what is a fast-moving social landscape. Furthermore, there was the occasional inability to recognise the legitimate use of uppercase letters both in unknown and specialist/topic-specific fields (e.g., CH3CH2OH is not inappropriate excessive use of capital letters, it is the chemical formula for Alcohol). Future work should address these issues.

Location of a Source – this algorithm is based on the hypothesis that sources closer to an event (e.g., eye-witnesses) may know more accurate/up-to-date information about it. The metric calculates the distance between two geographical coordinates (typically, the location of the event of interest and location of a source publishing event-related content) using earth geometry and then heuristically assigns values $[0, 1]$ based on their proximity, higher values for closer proximity. Unfortunately, we were unable to thoroughly evaluate this metric because of a lack of availability of geo-tagged content. A broader investigation of the field highlighted that only a small percentage of tweets are geo-tagged [24]—this is an interesting finding in itself as it suggests that application of this factor might be limited unless novel techniques of source/content positioning are applied.

7 Conclusions and Future Work

In this paper, we have reflected on the existing research on trust and quality metrics as it specifically pertains to using them to support decision-making online. We uncovered numerous measurement techniques which advance the field, but arguably much more needs to be done to allow for a greater degree of automated trust assessment and higher quality of the factor scores output. To further progress the goal of decision-support online, this paper extended previous work on metrics to incorporate the consideration of loss of information integrity due to infrastructure compromise, and also contributed a few new factor metrics and combinatorics. The initial findings from the metrics assessment was positive but as mentioned, further refinement is needed to enhance their ultimate utility at measuring and indicating trustworthiness to users online.

Future work will focus on further development of the proposed metrics, and other techniques that may assist in measurement. This will draw from the weaknesses discovered from our metrics experimentation and ensuring they are adequately addressed. Additionally, we will seek to conduct a case-study evaluation of the III metric with real data on attacks, threats and vulnerabilities. This will be interesting as it will combine live data from online monitoring sites towards the definition of an impact score. Finally, an assessment will be done on the overall utility of the policy-based methodology in enabling users to assimilate significant amounts of online content and make well-conceived decisions in an effective and timely manner. This will interact with our other research towards optimising the communication of risk and trustworthiness in interfaces [25].

References

1. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. IEEE Intelligent Systems **99** (2012)
2. Guardian News: How riot rumours spread on twitter (2012) http://www.guardian.co.uk/uk/interactive/2011/dec/07/london-riots-twitter.

3. Stvilia, B., Gasser, L., Twidale, M., Smith, L.: A framework for information quality assessment. Journal of ASIST **58**(12) (2007) 1720–1733
4. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: Web search and web data mining Conference. (2008)
5. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: 20th International Conference on World Wide Web. (2011) 675–684
6. O'Mahony, M., Smyth, B.: Using readability tests to predict helpful product reviews. In: Adaptivity, Personalization, Fusion Heterogeneous Information. (2010)
7. Rahman, S.S., Creese, S., Goldsmith, M.: Accepting information with a pinch of salt: handling untrusted information sources. In: 7th International Workshop on Security and Trust Management (STM), Springer-Verlag (2011) 223–238
8. Nurse, J.R.C., Rahman, S.S., Creese, S., Goldsmith, M., Lamberts, K.: Information quality and trustworthiness: A topical state-of-the-art review. In: International Conference on Computer Applications & Network Security, IEEE (2011) 492–500
9. Weerkamp, W., de Rijke, M.: Credibility improves topical blog post retrieval. In: 46th Annual Meeting of the Assoc. for Computational Linguistics. (2008) 923–931
10. Helfert, M., Foley, O., Ge, M., Cappiello, C.: Limitations of weighted sum measures for information quality. In: 15th Americas Conf. on Information Systems. (2009)
11. Blumenstock, J.: Size matters: word count as a measure of quality on Wikipedia. In: 17th International Conference on World Wide Web. (2008) 1095–1096
12. Mosquera, A., Moreda, P.: Smile: An informality classification tool for helping to assess quality and credibility in web 2.0 texts. In: 6th International AAAI Conference on Weblogs and Social Media. (2012)
13. Naumann, F.: From databases to information systems-information quality makes the difference (2001) IBM Almaden Research Center.
14. Adler, B.T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I., Raman, V.: Assigning trust to Wikipedia content. In: 4th International Symposium on Wikis. (2008)
15. Al-Oufi, S., Kim, H.N., Saddik, A.E.: A group trust metric for identifying people of trust in online social networks. Expert Systems with Applications **39**(18) (2012)
16. Lumbreras, A., Gavalda, R.: Applying trust metrics based on user interactions to recommendation in social networks. In: International workshop of Social Knowledge Discovery and Utilization. (2012)
17. Ramachandran, S., Paulraj, S., Joseph, S., Ramaraj, V.: Enhanced trustworthy and high-quality information retrieval system for web search engines. International Journal of Computer Science Issues **5** (2009)
18. Ward, J., Leach, J., Creese, S.: Measuring internet threat exposure scoping out a practical approach, metrics special interest group KTN. Technical report (2008)
19. Nurse, J.R.C., Creese, S., Goldsmith, M., Rahman, S.S., Lund, D., Mourikas, G., Price, D.: TEASE Project Deliverable D3.6: Consolidated report on trustworthiness measures and overlay design. Technical report (2013)
20. Agarwal, B.: Programmed Statistics. New Age International Ltd. (2007)
21. Grandison, T., Sloman, M.: A survey of trust in internet applications. IEEE Communications Surveys and Tutorials (2000) 2–16
22. Gil, Y., Artz, D.: Towards content trust of web resources. Journal of Web Semantics: Science, Services and Agents on the World Wide Web **5**(4) (2007) 227–239
23. Carrot2: Open Source Search Results Clustering Engine http://project.carrot2.org.
24. Graham, M.: Big data and the end of theory? (Guardian News) (2012) http://www.guardian.co.uk/news/datablog/2012/mar/09/big-data-theory.
25. Nurse, J.R.C., Creese, S., Goldsmith, M., Lamberts, K.: Using information trustworthiness advice in decision-making. In: Socio-Technical Aspects in Security and Trust (STAST) Workshop at 25th IEEE CSF Symposium, IEEE (2012) 35–42