To appear in Journal of Web Semantics: Science, Services and Agents on the World Wide Web, December 2007.

Towards Content Trust of Web Resources

Yolanda Gil and Donovan Artz

Information Sciences Institute

University of Southern California

4676 Admiralty Way, Marina del Rey, CA 90292, USA

(310) 822-1511

gil@isi.edu

ABSTRACT

Trust is an integral part of the Semantic Web architecture. Most prior work on trust focuses on entity-centered issues such as authentication and reputation and does not take into account the content, i.e. the nature and use of the information being exchanged. This paper defines *content trust* and discusses it in the context of other trust measures that have been previously studied. We introduce several factors that users consider in deciding whether to trust the content provided by a Web resource. Our goal is to discern which of these factors could be captured in practice with minimal user interaction in order to maximize the quality of the system's trust estimates. We present results on a study to determine which factors were more important to capture, and describe a simulation environment that we have designed to study alternative models of content trust.

Keywords: Trust, Content Trust, Web of Trust, Semantic Web

1. INTRODUCTION

Information comes from increasingly diverse sources of varying quality. We make judgments about which sources to rely on based on prior knowledge about a source's perceived reputation, or past personal experience about its quality relative to other alternative sources we may consider. Web users make these judgments routinely, since there are often numerous sources relevant to a given query, ranging from institutional to personal, from government to private citizen, from formal report to editorial, etc. In more formal settings, such as e-commerce and e-science, similar judgments are also made with respect to publicly available data and services. All of these important judgments are currently in the hands of humans. This will not be possible in the Semantic Web [Berners-Lee 99; Berners-Lee et al 01; Berners-Lee et al 06]. Reasoners and agents alike will need to automatically make these judgments to choose a service or information source while performing a task. Reasoners will need to judge what information sources are more adequate for answering a question. In a Semantic Web where content will be reflected in ontologies and axioms, how will these automated systems choose the US census bureau over the thousands of Web pages from travel and real estate agents when searching for the population of Chicago? What mechanisms will enable these kinds of trust judgments in the Semantic Web?

In the original Semantic Web architecture design, the trust layer was envisioned to address authentication, identification, and proof checking [Berners-Lee 99; Berners-Lee et al 01; Berners-Lee et al 06], but did not mention trust in the content itself. Because the Semantic Web makes it possible to represent the content of resources explicitly, this opens the possibility of looking beyond actors and into the content when determining trust. The identity of a resource's creator is just one part of a trust decision, and the Semantic Web provides new opportunities for considering the nature of the resource's content directly.

Trust has been studied in various areas of computer science and in the context of the Web and the Semantic Web [Artz and Gil 07; Ruohomaa Kutvonen 05; Sabater and Sierra 05]. Prior work on trust has focused on issues such as reputation and authentication [Blaze et al 96; Goldbeck and Hendler 04; Rivest and Lampson 96]. Trust is an important issue in distributed systems and security. To trust that an entity is who it says it is, authentication mechanisms have been developed to check identity [Rivest and Lampson 96], typically using public and private keys [Kohl and Neuman 93; Resnick and Miller 96]. To trust that an entity can access specific resources (information, hosts, etc) or perform certain operations, a variety of access control mechanisms generally based on policies and rules have been developed [Alexander et al 98; XACML 07; SAML 07; WS-Trust 07]. Semantic representations [Nejdl et al 04; Bonatti and Olmedilla 05; Gandon and Sadeh 04; Kagal et al 03; Uszok et al 03] can be used to describe access rights and policies. The detection of malicious or otherwise unreliable entities in a network has also been studied, traditionally in security and more recently in P2P networks and e-commerce transactions [Braynov and Jadliwala 04; Kamvar et al 03; Cornelli et al 02; Damiani et al 02].

Such trust representations and metrics do not take into account how the nature of information being exchanged affects trust judgments. We distinguish between entity trust, which is concerned with trust judgments on the providers of the information, and content trust, which is concerned with the nature of the information provided. In prior work in TRELLIS, we developed an approach to derive consensus content trust metrics from users as they analyzed information from many sources, each for a different purpose and context [Gil and Ratnakar 02a; Gil and Ratnakar 02b]. However, the trust metrics were tightly coupled to the analysis structures in TRELLIS. In this paper we investigate the acquisition of content trust from users in a generic search-then-rate environment on the Web.

We begin by describing what content trust is. We identify key factors in modeling content trust in open sources and describe related work on some of these factors. We then describe a model that integrates a subset of those factors to model content trust. Finally, we show some results in a simulated environment where content trust can be derived from inputs from individual users as they search for information.

2. WHAT IS CONTENT TRUST?

Existing approaches to model trust focus on entities [Artz and Gil 07; Blaze et al 96; Braynov and Jadliwala 04; Chu et al 97; Golbeck and Hendler 04; Kamvar et al 03; Rivest and Lampson 96], but they only take into account overall interactions across entities and disregard the nature of interactions, i.e. the actual information or content exchanged. We refer to this as *entity trust*. Entity trust is a trust judgment regarding an entity based on its identity and its behavior, and is a blanket statement about the entity. This is insufficient in many situations that require making a selection among sources of information.

Content trust is a trust judgment on a particular piece of information in a given context. Content trust is related to entity trust but is different. Consider the case when many low-trust entities provide the same content independently. That content may be trusted even though the individual entities are not trusted. Consider also a case where a high-trust entity provides content that contradicts what many low-trust entities are providing independently. In this case, the content provided by the high trust entity may not be trusted. Therefore, we argue that the degree of trust in an entity is only one ingredient in deciding whether or not to trust a particular piece of information that it provided.

To illustrate the difference between content and the entities that provide it, consider an example from [Gil and Ratnakar 02a] where a document in an FA Times article from the CREUTERS agency reports on drug problems of a public figure and whose content is "At a press conference last Monday, Duckingham Palace was adamant that Prince Larry did not inhale". Many entities are originators or producers of this content. The Dublin Core standard [DublinCore 07] has well defined relations to express attribution. In this case, the CREUTERS journalist would be the creator, FA Times the publisher, the Duckingham Palace spokesperson would be a contributor, the original CREUTERS article would be the source, and the tapes of the press conference could be specified as a relation. Additional entities may be cited by this content, such as entities that sponsor the work to generate content and entities cited in a document.

When considering content trust, one must determine what defines a unit of content and how it can be described. Content is made available on Web sites managed by organizations, by services that provide information in response to specific queries, and by individuals on their personal sites or spaces. Sources can be documents that are made available on the Web, static Web pages, or dynamic Web pages created on-demand. All these formats vary greatly in nature, granularity, and lifespan. Fortunately, the Web gives us a perfect mechanism to define a unit of content: a Web *resource*. We consider content trust judgments made on specific resources, each identified by a unique URI, and the time of its retrieval. Although finer-grain trust decisions can be made, for example on each individual statement, we consider here a Web resource as a basic unit for content trust on the Web. Ultimately, each statement can be assigned its own URI.

Content trust is often subjective, and there are many factors that determine whether content could or should be trusted, and in what context. Some resources are preferred to others depending on the specific context of use of the information (e.g., students may use different sources of travel information than families or business people). Some resources are considered very accurate, but they are not necessarily up to date. Content trust also depends on the context of the information sought. Information may be considered sufficient and trusted for more general purposes. Information may be considered insufficient and distrusted when more fidelity or accuracy is required. In addition, specific statements (content) by traditionally authoritative entities can be proven wrong in light of other information. The entity's reputation and trust may still hold, or it may diminish significantly. Finally, resources may specify the provenance of the information they provide, and by doing so may end up being more trusted if the provenance is trusted in turn.

Content trust is a new area of research that we foresee will take on a prominent role in Semantic Web community. Addressing content trust presents many challenges. What factors that influence content trust can be captured in practice? How can these factors be combined into an overall content trust value? Can users supply valuable information about trust as they analyze web resources? This paper reports on our work to date to address these open questions.

3. FACTORS THAT INFLUENCE CONTENT TRUST

We conducted an exhaustive literature review of trust research to investigate issues relevant to content trust [Artz and Gil 07]. We also analyzed many scenarios of content trust decisions in a variety of domains and contexts. We identified several salient factors that affect how users determine trust in content provided by Web information sources:

1. **Topic**. Resources that would be trusted on certain topics may not be trusted for others. We may trust a critic's movie site for director's information but not for market prices for movies.
2. **Context and criticality**. The context in which the information is needed determines the criteria by which a user judges a source to be trustworthy. If the need for information is critical and a true fact needs to be found with high precision, the amount of effort placed in comparing, contrasting, ranking, and disproving information is much higher.
3. **Popularity**. If a resource is used or referenced by many people, it tends to be more trusted.
4. **Authority**. A resource describing an exchange rate is more trusted if it is created by financial news source, as opposed to by an anonymous Internet user.
5. **Direct experience**. The direct interaction of a user with a resource provides reputation information, a record of whether or not trust was well-placed in the past.
6. **Recommendation**. Referrals from other users for a resource or its associations provide indirect reputation information.
7. **Related Resources**. Relations to other entities which allow (some amount of) trust to be transferred from those resources (e.g., citations and Web hyperlinks)
8. **Provenance**. Trust in the entities responsible for generating a unit of content may transfer trust to the content itself.

9. **User expertise**. A user with expertise in the information sought may be able to make better judgments regarding a resource's content, and conclude whether or not it is to be trusted. For example, residents of a city have may have more expertise in knowing which resources are authorities on local demographics.
10. **Bias**. A biased source may convey certain information that is misleading or untrue. For example, a pharmaceutical company may emphasize trial results and omit others with respect to a certain type of treatment. Bias is often not only subtle, but also very hard to determine without deep expertise in the subject matter.
11. **Incentive**. Information may be more believable if there is motivation for a resource or its associations to provide accurate information.
12. **Limited resources**. The absence of alternate resources may result in placing trust in imprecise information. Some resources may end up being trusted only because no other options are available.
13. **Agreement**. Even if a resource does not engender much trust in principle, a user may end up trusting it if several other resources concur with its content.
14. **Specificity**. Precise and specific content tends to engender more trust than abstract content that is consistent with true facts.
15. **Likelihood**. The probability of content being correct, in light of everything known to the user, may be determined with an understanding of the laws and limits of the domain.
16. **Age**. The time of creation or lifespan of time-dependent information indicates when it is valid. For example, a detailed weather report that is updated weekly may be trusted the day it is posted, but other sources may be used during the week, even if less detailed.
17. **Appearance**. A user's perception of a resource effects the user's trust of the content. For example, the design and layout of a site and the grammar and spelling of the content may both be used to judge content accuracy, and whether it should be trusted.
18. **Deception**. Some resources may have deceptive intentions. Users should always consider the possibility that a resource may not be what it appears to be, and that the stated associations may not be recognized by the sources they reference.
19. **Recency**. Content, associations, and trust change with time. For example, a resource that had a very good reputation a few months ago may degrade its behavior and have a worse reputation.

Table 1 summarizes the factors discussed. Some of these factors are related. Topics and criticality specify the context of trust and therefore restrict the scope of trust, allowing for more accurate determination. Direct experience and recommendations capture reputation by using a resource's history in determining if it should be trusted now or in the future. Limited resources and agreement are relative trust judgments, made when an absolute trust decision is not possible. Associations (e.g., authority and resource associations) allow the trust on some entities to be transferred to a resource associated with those entities. Conversely, once a trust judgment is made about a resource, that trust may be propagated out to a resource's associations, or otherwise related resources. Many of those factors are heuristic in nature, for example incentive and likelihood may be estimated using general knowledge about the world. An important challenge is to determine which of these factors can be captured in practice.

The next section presents an overview of previous research that addresses some of these factors.

4. RELEVANT RESEARCH TO CONTENT TRUST

Popularity is often correlated with trust but not necessarily. One measure of popularity in the Web is the number of links to a Web site, and is the basis for the widely used PageRank algorithm [Brin and Page 98]. Popular sources are often deserving of higher trust, but this is not always the case. For example, blogs can be ranked high because of the popularity of certain bloggers and their higher degree of linking by others, even though the value of some of the information they provide and comment on is not necessarily trustworthy. Another problem with the PageRank algorithm is that it does not capture the negative references to a linked source. For example, a link to a source that is surrounded by the text "Never trust the Web site pointed to by this link" is counted as a positive vote of the source's popularity, just as positive as a link surrounded by the text "I always trust the Web site at this link" [Massa and Hayes 05]. This problem is often discussed in the context of spam [Gyongyi et al 04], but not in terms of the content provided by the sources.

Authority is an important factor in content trust. Authoritative sources on the Web can be detected automatically based on identifying bipartite graphs of "hub" sites that point to lots of authorities and "authority" sites that are pointed to by lots of hubs [Kleinberg 99]. This mechanism can be used to complement our approach by weighing associations based on their authority. Many Web resources lack authoritative sources. Preferences among authoritative sources within a topic still need to be captured.

1.	Topic considered
2.	Context and criticality of the need for information
3.	Popularity of the resource
4.	Recognized authority of associations
5.	Reputation by direct experience
6.	Referrals by other users
7.	Association by other trusted resources (eg citations)
8.	Provenance and pedigree
9.	Expertise of the user
10.	Perceived bias of source
11.	Perceived incentive in providing accurate information
12.	Absence of other alternative resources
13.	Agreement with other resources
14.	Precise and specific content
15.	Likelihood of content being correct given what is known
16.	Time of creation of the content
17.	Professional appearance
18.	Likelihood of deceptive behavior
19.	Recency of factors under consideration

Table 1. Factors that influence content trust decisions.

Reputation of an entity can result from direct experience or recommendations from others. Reputation may be tracked through a centralized authority or through decentralized voting [Blaze et al 96; Chu et al 97]. The trust that an entity has for another is often represented in a web of trust, where nodes are entities and edges relate a trust value based on a trust metric that reflects the reputation one entity assigns to another. A variety of trust metrics have been studied, as well as algorithms for transmission of trust across individual webs of trust [Goldbeck and Hendler 04; Kamvar et al 03]. Semantic representations [Goldbeck and Hendler 04; Chirita et al 04] of webs of trust and reputation are also applied in distributed and P2P systems.

There are manual and automatic mechanisms to define provenance of resources. The Dublin Core [Dublin Core 07] defines a number of aspects related to provenance. Provenance can be captured using semantic annotations of results inferred by reasoners [Zaihrayeu et al 05], including explanations of reasoning steps and axioms used as well as descriptions of original data sources.

All related work described so far focuses on trusting entities rather than trusting content. In prior work we developed TRELLIS [Gil and Ratnakar 02a; Gil and Ratnakar 02b], a system that allows users to make trust related ratings about sources (entities) based on the content provided. Users can specify the source attributions for information extracted during a search and information analysis process to describe the sources. As users specify ratings, they are used to automatically derive a measure of collective trust based on the trust metrics from individual users. In TRELLIS, a user can add semantic annotations to qualify the sources of a statement by its reliability and credibility. Reliability is typically based on credentials and past performance of the source. Credibility specifies the user's view of probable truth of a statement, given all the other information available. Reliability and credibility are not the same, as a completely reliable source may provide some information that may be judged not credible given other known information. This is an approach to distinguish between entity trust and content trust. However, in TRELLIS the derived consensus trust was only applicable to analyses that were created with TRELLIS. Some later work was done on representing TRELLIS structures in the Semantic Web [Blythe and Gil 04], but the algorithms for deriving content trust were not fully integrated.

In summary, there are techniques to address some of the factors that we outlined as relevant to content trust, such as popularity, authority, reputation, and provenance. The challenge is how to integrate these techniques and incorporate the other remaining factors to enable content trust on the Web.

5. ACQUIRING CONTENT TRUST FROM USERS

There are no mechanisms in today's Web to represent or capture content trust. Although millions of Web users make content trust decisions on a daily basis, all those decisions do not leave behind information. For example, current search engines do not capture any information about whether or not a user "accepts" the information provided by a given Web resource when they visit it, nor is a click on a resource an indicator of acceptance, much less trust by the users that have visited it. We wish to capture, in the least intrusive way, some information about why any content provided by a resource is trusted. This information can be used to decide what resources should be more highly ranked in terms of trust. Starting from a baseline system that ranks search results by topic and popularity, our goal is to develop new techniques to re-rank search results using additional trust factors so that more trustworthy resources appear higher in the results list.

At the same time, users are unlikely to invest the time to record their content trust decisions and the rationale that led to them. Therefore, it is crucial to determine (1) what factors have most utility in determining content trust, (2) what information can be captured in practice from users regarding content trust decisions as they use the Web, (3) how a user's information can be complemented by automatically extracted information, and (4) how to use this information to derive content trust. Next, we present our work to date to study and model the acquisition of content trust from users as they perform Web searches. The purpose of this model is to study different approaches to collect and learn content trust.

6. CENTRAL FACTORS IN CONTENT TRUST DECISIONS

We conducted a detailed analysis to prioritize our research by identifying some of the main factors that support the rationale for content trust decisions. This analysis was designed to be a formative analysis that would provide a set of uniform set of criteria and decisions on trust content to guide our initial work. This section describes the details of this analysis and the findings that drove the focus of our subsequent work.

The goal of this analysis was to determine what contributing factors to content trust decisions discussed in Section 3 were more prominent in making decisions. We chose 5 topics, and considered 20 resources (the top 20 returned by a search engine) and annotate an assessment for each of the 19 factors. For each of the 5 topics, we created a matrix of 20 rows corresponding to the 20 resources examined, and 19 columns corresponding to the 19 factors for content trust. This gave us a total of 100 content trust decisions on that many sources based on a total of 1,900 factors considered.

We considered the context of the content trust decision to be conducting a web search for a well-defined research purpose. The five topics chosen were:

E1: Ground turkey cholesterol.

E2: Staffordshire hotels.

E3: Remaining rainforests.

E4: Giraffe lifespan.

E5: Al Quaeda headquarters.

Each resource was marked as “irrelevant”, “trust”, “limited trust”, and “distrust”. In all cases, the context was clearly stated. For E1, E3, and E4, we assumed the user was looking for information that was trusted enough to include in a course project. For E2, we assumed a user would be looking for a hotel to stay in a holiday in that area. For E5, a user would be interested in finding the location.

For all the sources marked as relevant, we considered all 19 factors in making their trust decision of “trust”, “limited trust”, or “distrust”. Each of the 19 factors were marked as “+”, “-“, “?”, and “not considered”. The plus and minus indicated a positive and negative influence of that factor in the overall trust rating for that source. Note that the strength of a given factor and its weight in the overall trust decision were not captured. The question mark included cases where the factor was of unknown value, or it was not possible to find information that enabled assessment of that factor, or factors that simply did not apply to that particular source.

We first discuss qualitative findings of the study, and then present more quantitative results.

A subset of the factors did not play a central role in the queries considered in this analysis. Three of the factors were built into the scenarios. The topic considered (Factor 1) is given, and the context and criticality of the need for information (Factor 2) is determined by the goals of the search (for a course, a holiday, or to answer a question). We decided to build these into the scenarios, as addressing them would require a more complex user model that is also absent in current web search engines and therefore in our assumed baseline system. The expertise of the user (Factor 9) was none in all cases. This is a good default assumption, since users with expertise are more likely to rely on their own knowledge to make content trust decisions, rather than trust metrics provided by the system. In addition, four of the factors did not intervene in the decisions for these particular scenarios. Reputation by direct experience (Factor 5) was not available for any of the sources in the scenarios, neither was referral by other users (Factor 6). As above, incorporating these factors would require user models that are absent from current Web search engines and our baseline model. The baseline search engine we used ranks results higher based on popularity (Factor 3), so that factor was not analyzed. The time of creation of the content (Factor 16) and the recency of factors under consideration (Factor 19) were not examined for the queries considered.

Of the remaining factors, five were found to have minimal or no influence in the overall trust decisions. The absence of alternative resources (Factor 12) did not increase the trust more on the sources at hand. This may be due to the context set for the analysis, where the quality of the result of the analysis matters and therefore trust is not placed lightly. Agreement with other resources (Factor 13) was also not found to influence trust. The specificity of the content (Factor 14) and the professional appearance (Factor 17) we not considered either in the decision. In these four cases, we believe these factors were not influential because of the perceived importance of trust in the context and use of the content that were given. The fifth factor not considered was the likelihood of the content being correct (Factor 15), which we believe was due to the lack of expertise assumed in the topics searched (Factor 9).

There were six factors that were considered important in making content trust decisions. The first three were directly related to the origins of the information:

- Recognized authority of associations (Factor 4)
- Related resources (Factor 7)
- Provenance and pedigree (Factor 8)

In the remainder of this section, we refer to these factors as Authority (A), Related Resources (R), and Provenance (P) respectively.

The other three factors were:

- Perceived bias of the source (Factor 10)
- Perceived incentive in providing accurate information (Factor 11)
- Likelihood of deceptive behavior (Factor 18)

These three are all concerned with bias. We found that these three factors were analyzed together, so we conflated them into one. We refer to these three factors as Bias (B).

In summary, the qualitative findings of our study is that the principal factors out of the nineteen considered were Authority (A), Related Resources (R), Provenance (P), and Bias (B). We now present quantitative results on these main factors.

Table 2 shows the overall source ratings. Of the 100 resources, 39 were considered irrelevant. Relevance was not one of our original factors because we assumed that search engines would retrieve mostly relevant sources. However, a large amount of sources were considered not relevant. As it turned out, this was not necessarily due to shortcomings in the search engine's algorithms, but often we found the sources themselves advertised their content in a misleading way.

Table 2 also shows that an overwhelming majority of the resources (66%) were rated as limited trust. A few sosources were rated as trusted (10%), and a few others as distrusted (14%). This highlights the difficulties for a particular user with no expertise in a topic to determine whether to trust a source, and only about a third of the resources consulted can be classified as trusted or as distrusted, two thirds fell in the middle of the road category.

Table 3 presents data on authority (A), provenance (P), related resources (R), and bias (B). We discuss first the first three factors. These three factors are analyzed by searching for *associations* of the resource, i.e., entities or other resources that were linked to the origins of the information contained. We noted that many associations could not be determined and therefore A, P, and R were hard to judge. Table 3 shows for each of the three factors the proportion of associations investigated that were actually determined. The data indicates that the majority of associations (66%) were not determined, an even larger proportion in the case of A. This suggests that users may find challenging to assess authority. Fortunately, there are effective algorithms to find authoritative sites on the web [Kleinberg 99] that could be used to assess this particular factor automatically for the user.

We noted that at least one of these three factors was considered for any given resource. We also noted that only in very few cases (3/61) the association-related ratings were opposite (e.g., negative in authority and positive in provenance). This suggests that a content trust model should always consider associations, and at least one association per resource.

Table 3 also includes data on the bias (B) factor, and shows that in the majority of the cases (72%) the user was able to determine bias. This is a high rate if compared with the 33% rate for associations. Since bias appeared to be easier to determine, we analyzed whether bias could be used as a predictor of trust and distrust decisions.

Table 4 shows the data on bias, both positive and negative, with respect to trust and distrust decisions. We found that bias is highly correlated with trust and distrust decisions. At the same time, we found that bias alone yields too many false positives when used a predictor of trust and distrust decisions.

We noticed that many associations determined bias. Table 5 shows an analysis on whether bias was apparent in the resource itself (which we termed content bias), or if one if the resource's associations were perceived as biased (which we termed association bias). The data show that associations are the overwhelming indicator used to determine bias.

In summary, our analysis showed that the main factors that influence trust decisions on the content of a web resource are authority, related resources, provenance, and bias. The first three factors are determined by examining associations of the resource. The fourth factor, bias, appears to be overwhelmingly determined by examining associations as well, rather than being apparent in the resource content itself. Our analysis highlighted the importance of associations to determine content trust for any given resource. We concluded that the core of our model for content trust would represent and propagate trust on the associations of individual resources. The next section presents the details of our model for content trust.

7. MODELING THE ACQUISITION OF CONTENT TRUST FROM USERS

Our analysis suggests that content trust models should center on associations and the transfer of trust through associations. Associations are central to the Semantic Web. RDF was originally designed to represent information about associations of resources on the Web. Because associations facilitate the transfer of existing trust, they serve as an explicit source of trust information, unlike the many other trust factors (e.g., time of creation, likelihood, appearance, etc.).

Scenario	*Relevant*	*Distrust*	*Limited Trust*	*Trust*
E1	5	3	2	0
E2	15	1	13	1
E3	11	2	7	2
E4	15	1	12	2
E5	15	6	7	2
Total	61	13	41	7

Table 2. Source ratings in our study, marked as relevant, distrust, limited trust, and trust.

Scenario	*Authority*	*Provenance*	*Related Resources*	*Bias*
E1	2/5	3/5	0/4	1/4
E2	11/15	15/15	13/15	1/15
E3	10/11	9/11	7/11	4/11
E4	10/14	12/15	11/15	8/11
E5	13/15	13/15	8/15	2/15
Total	46/60	42/61	39/60	16/56

Table 3. Source associations that could not be determined in analyzing authority, provenance, related resources, and bias.

Scenario	-$\mathcal{B}$ & $\mathcal{D}$ / $\mathcal{D}$	+$\mathcal{B}$ & $\mathcal{T}$ / $\mathcal{T}$
E1	3/3	0/0
E2	1/1	1/1
E3	2/2	2/2
E4	0/1	2/2
E5	6/6	2/2
Total	13/14	7/7

Scenario	-$\mathcal{B}$ & $\mathcal{D}$ / -$\mathcal{B}$	+$\mathcal{B}$ & $\mathcal{T}$ / +$\mathcal{B}$
E1	3/4	0/0
E2	1/13	1/1
E3	2/5	2/2
E4	0/0	2/3
E5	6/6	2/7
Total	12/31	7/13

Table 4. Bias as a predictor of trust/distrust decisions.

Scenario	*Content*	*Associations*	*Both*
E1	0	3	0
E2	0	14	0
E3	1	5	1
E4	1	1	1
E5	2	11	0
Total	4	34	2

Table 5. Bias in content and bias in associations.

Once we identify a unit of content, many associations related to it can influence content trust. One important set of *associations* is the group of entities responsible for the information within a resource. Moreover, the roles of those associated entities further specify the context of trust. We mentioned in Section 2 that there may be a variety of relationships between content and entities. For example, a Web page that contains an article can be associated with "Joe Doe" as one author, *newstoday.com* as a publisher, and "Charles Kane" as the editor. The types proposed in the Dublin Core [Dublin Core 07] provide a reasonable set of roles for this kind of information. There are other kinds of associations possible. For example, a resource may be endorsed by an entity, or a resource may cite another resource as evidence for the content it provides.

The types of associations of resources mentioned so far are strongly correlated to trust, but there are many other types of associations that are used only selectively. Consider, for example, a Web resource that recommends a set of readings in the history of astronomy, and is maintained by an astronomy department on a university Web site. If the Web page is authored by a faculty member in the astronomy department, then a user would make a strong association between trust in the content and trust in the university, the department, and the authoring professor. If the Web page is authored by a student on a temporary internship, who happens to like astronomy as a hobby, the user would not put as much weight in the association of the resource with the astronomy department or the university. In general, a Web page's main site is an associated entity that should not be assumed to be highly weighted when determining trust.

In our initial work, we assume that each association has a single overall trust value. We do not address how that trust value is derived, possibly as a combination of its popularity, reputation, and authority. There is ongoing work in capturing associations for Web resources, particularly provenance and authority, and we may use these associations to transfer trust from entities to resources. We believe our framework can be extended to incorporate those factors explicitly in future work.

The rest of this section describes our model for studying the use and acquisition of content trust. This model is used later in the paper to create simulations of users providing and using content trust metrics. Note that this is the first model proposed that reflects the acquisition of user feedback on content trust. Existing models address entity trust, and could be integrated with the model proposed here to create more comprehensive trust models.

A resource, $r \in R$, is our basic unit of content to which trust can be applied. A resource can be a Web site or service, and in this work, is anything that can be referenced by a URI. The URI serves as a resource's unique identifier, and this identifier is returned by the function *ID(r)*. A resource also has a time at which it was retrieved, which is returned by the function *time(r)*. An association is anything having a relationship to a resource, such as an author, a sponsor, or a service provider. Each resource is represented by a subset of the set of all associations A. Each member of A is an association tuple, $<ar , ae>$, which contains an association relation, *ar*, and an association entity, *ae*. Association entities may be anything that can be trusted (or distrusted), including people, businesses, governments, or other resources (including services). A single association entity may participate in multiple possible types of relations. For example, the entity "Noam Chomsky" may be an author, a subject, or even a critic of any given resource: ("author", "Noam Chomsky"), ("subject", "Noam Chomsky"), or ("critic", "Noam Chomsky"). Figure 1 illustrates our model of resources and associations examples. We assume the associations for each resource are given.

We will study trust over a fixed time where a set of users, U, make a subset of queries from a set of possible queries, Q. A user, $u \in U$, queries an information system and analyzes the results to determine content trust. The set of users who make query q is U^q, a subset of U. The result returned for q is a sequence of resources, R^q. The baseline system returns resources ordered by relevance as current search engines do, without taking trust into account. The resource rq i is the i^{th} resource in R^q. When using this model for simulation, we assume that the queries, users, resources, and associations are given.

We define several functions, each returning a value representing trust. All functions that return trust have τ as a range. τ can be discrete or continuous. For example, it could be a discrete set with the values *trust*, *distrust*, and *neutrality* (i.e., neither trust nor distrust).

Users make trust decisions for a resource by combining trust in that resource's individual associations. As a starting point, we assume that users will provide the system with an overall trust value on a given resource without going into any details on why and what produced that trust value. A user's trust in an association for a given query is the *user association trust*, mapped by the function $UAT: Q,A,U \rightarrow \tau$. This function is given to the simulation, and we assume it does not change over time. *UAT* is derived by the user from various forms of entity trust already mentioned, such as reputation and authority. A user's trust decision for a resource is computed from trust decisions for that resource's associations for a given query. This is the *user resource trust*, and is mapped by the function $URT: Q,R,U \rightarrow \tau$. Examples of methods for computing the *URT* include the sum, the mean, or the maximum of the *UAT* for all of a resource's associations. Note that each user may have a unique function to determine trust, and we incorporate this by including the user as an input to the single function, *URT*. It is our expectation in real systems that the output of *URT* will be easier to capture than *URT* itself. However, for our simulation, we model users by implementing *URT*. We assume for this paper that users provide *URT* for some (not all) query results, since specifying *UAT* is more intrusive.

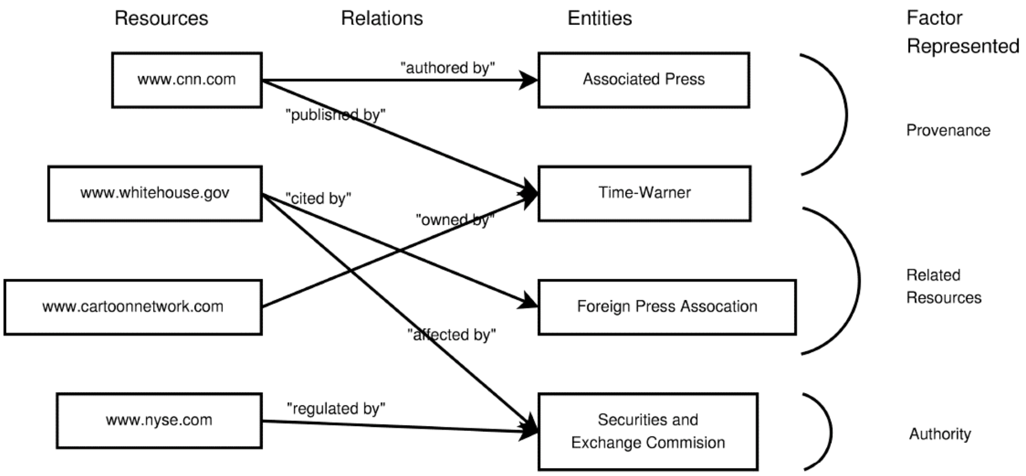


Figure 1: A resource may have multiple associations, and an entity can be related to multiple resources with different relationships.

The *association trust*, $AT: Q,A \rightarrow \tau$, is the global trust of an association, derived from the *UAT* of individual users. The *resource trust*, $RT: Q,R \rightarrow \tau$, is the global trust of a resource, derived from the result of *URT* for all users. It is possible to derive RT if *AT* is known, using a given function similar to that used to compute the output of *URT* from *UAT*. However, in real systems, neither the outputs for *RT* or *AT* are known, as it is not possible to ask each user for a trust decision for each resource or association for each possible query.

We propose the *RT* for any resource and the *AT* for any association can be estimated using only the user inputs (*URT*) from a sample of users who have made a given query (which is assumed to be significantly less than the cardinality of U). The *estimated resource trust*, mapped by the function *ERT*: $Q,R \rightarrow \tau$, may be any function of the *URT* for all users in U^q, such as the sum, average, or mode. An estimate of *AT* is the *estimated association trust*, mapped by the function *EAT*: $Q,A \rightarrow \tau$, and could be derived from the *ERT* over all resources that have the association in question. We do not use the *EAT* in this work, but will in future work exploiting the transitivity of trust over associations to other resources.

Each resource has a relevance score, returned by the function s^q: $R \rightarrow O$, where O is a set of values that can be used to order (rank) resources (e.g., consider $O = \{0, 1\}$, and if $s^q(r) = 1$, then it is listed before any resource *r0* where $s^q(r0) = 0$). The *trust re-rank* function ρ: $O, \tau \rightarrow O$, maps an order value and a trust value to a new order value. We can apply this function to re-rank a sequence of query results, using the result of combining the relevance score function, *s*, and the *ERT* for each resource. An example of ρ may be a linear combination of the relevance and trust inputs. The re-ranked sequence of results, T^q, contains the elements of R^q sorted by the output of ρ.

Figure 2 illustrates the initial use of our model, starting with the original sequence of results, and ending with the re-ranked sequence. Given a query, q, the set of users who make that query, U^q, and the sequence of resources returned for that query, R^q, we obtain the *URT* from the users in U^q, and use those trust values to compute the ERT for all resources in R^q. The *ERT* is combined with the relevance score, s, using the trust rerank function, ρ, and T^q receives the elements of R^q sorted by both trust and relevance.

Note that our model considers global trust metrics for all users, and could be extended to compute local or customized trust metrics for individual users or specific groups.

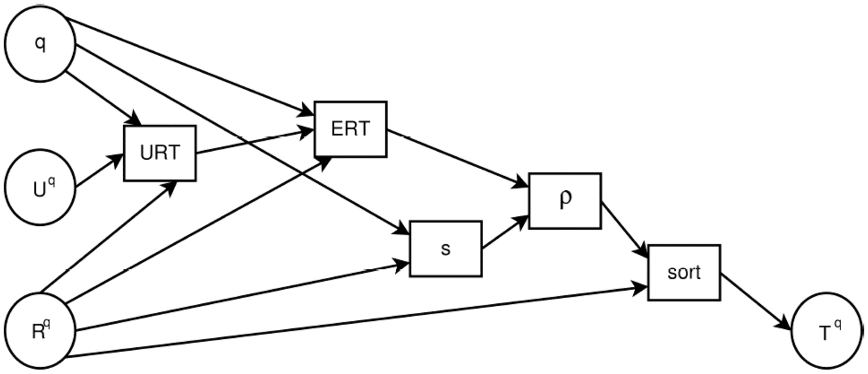


Figure 2: Model of trust to re-rank resources, where arrows denote input dependencies.

8. MODELING USE CASE SCENARIOS

Our long term plan is to use the model presented to: (1) study alternative approaches to collect content trust from individual users, learning trustworthiness over time, and to (2) help design a system that will collect content trust values from real Web users interacting with real Web search engines, and make predictions about the nature and utility of the trustworthiness values that are learned. Our first steps toward this plan are to explore how our model can represent different situations with varying amounts of information and trust values, and to study whether trustworthiness can be learned and estimated as proposed. This model will be the basis for future work to analyze through simulations alternative forms of user trust feedback and other model variants.

To illustrate how our model effectively captures content trust, we show the model used to simulate three nominal use case scenarios that are representative of the range of decisions users make regarding content trust. As we mentioned above, we assume for these simulations that the queries, users, resources, and associations are given. The initialization of the simulations is described in detail in this section.

8.1 Use Case Scenarios

We selected the following scenarios to illustrate some common issues we have encountered in our studies of using trust to choose information sources on the Web. In each scenario, some distrusted resources have higher relevance rankings than trusted resources, and if information about users' trust decisions were captured, it could be used to learn ERT and rank more trusted resources first.

8.*1.1 Trust and Distrust*

A user searches the Web for "ground turkey cholesterol", to learn how much ground turkey she can eat in her cholesterol limited diet. Out of hundreds of results, the user selects 5 candidates, and in examining these, she finds conflicting answers, even between sites that cite the same source. The first site is sponsored by the "Texas Beef Council", which compares ground turkey to ground beef. The second site belongs to a group of turkey farmers in British Columbia, Canada. The third site provides medical advice attributed to a "Dr. Sears", which the user trusts when she is seeking medical advice, but not for nutrition data. The fourth site provides an answer contributed by an anonymous person with no credentials or sources cited. The fifth site is the nutrition facts database created and published by the U.S. Department of Agriculture (USDA), the source cited by the "Texas Beef Council" site. Most users may agree, that the creators of first two sites hold a bias against and for turkey, respectively. The creators of the third site may be trusted by users in a medical context, but not as much for nutrition data. The fourth site may be dismissed, lacking a source or identifiable creator. The fifth site may be accepted by users, as they may already trust its associations (i.e., the USDA and the U.S. government).

In this scenario, the user is able to determine both trust and distrust using associations between the sites and the users' broad range of existing trust and distrust. Assuming many users make similar judgments, capturing their trust and distrust would allow the government site to be listed first, and the first four distrusted sites to be listed last.

8.*1.2 Distrust Only*

A user searches the Web for "remaining rainforests", seeking the specific number of acres left worldwide. Considering four candidates that appear to provide results, the user notes that all the sites provide a reasonable answer, but none provide a citation or other verifiable source. Moreover, the user is unable to find any associations where there is existing trust for this query, only distrust. The first site sells products made from plants and animals found in rain forests. The second site notes emphatically that human kind will perish completely by 2012 if the destruction of rain forests is not stopped immediately. The third site belongs to an organization known by user, the World Wildlife Federation. The fourth site considered, is intended for children, and includes a source, but the source cannot be found or verified. Except for the World Wildlife Federation (WWF), none of the results have clearly demonstrated their authority to answer the question, and even the WWF is biased with its ecological agenda. Without being able to identify trust over associations, users may at best be able to identify distrust.

This is a scenario showing how users could determine distrust in sites using existing distrust, but are not able to associate sites with any existing trust. Sites that have not been considered may have more potential to be trustworthy, and would be listed before unequivocally distrusted sites.

8.*1.3 Sparse Trust and Distrust*

A user wants to visit his friend in Staffordshire county, England, and searches the Web for "Staffordshire hotels". Out of many relevant results, all appearing equally likely to provide trustworthy information, 5 candidates are selected, each providing a tremendous amount of information. The first site provides a long list with a comprehensive set of details, but the source behind this information is unknown, and there is no indication of how the list has been generated. The second site is run by a company, Priceline, whose American operation is trusted by American users, but the UK division is largely unknown to Americans. The third site has a small and informative list with pictures, but again, no associations can be made to anything most users already trust. The fourth site collects and publishes user-submitted photographs of locations in England, and is funded by providing links to hotels that are nearby the locations pictured in the photos. The fifth site collects the opinions of travelers who have visited hotels in England, but does not restrict who may submit opinions.

This scenario illustrates that in cases of sparse existing trust and distrust, most users are not be able to make a trust or distrust decision for any of these sites. However, having asked a sufficiently large number of users, the few who have existing trust or distrust may be able to provide trust decisions. If there are a small group of users who know and trust the UK Priceline site, this site would be listed first if we are able to capture enough trust decisions.

8.2 Simulating the Use Case Scenarios

Recreating and simulating the use case scenarios with our model requires us to generate a large amount of data which represents the qualities described in each scenario. In this section, we describe what parameters we use, how we pick distributions to generate the necessary random data to populate the model, and what algorithms are used for the model's trust functions.

8.2.1 Initialization

We began by choosing the population sizes for each set, a set of order values, and a representation of trust. We adopted Marsh's [Marsh 94] range of trust values, $\tau = [-1, 1]$, where -1 is maximum distrust and 1 is maximum trust. Not all research agrees with this representation, but it provides a simple starting point for demonstrating our model. We defined the set of possible order values for relevance to be a singleton, $O = \{1\}$, such that in these examples, all query results are assumed to be equally relevant. However, O is equivalent to τ for the output of ρ, a trust-reranked ordering. For each use case scenario (a unique query), we examined 1000 random instances ($|Q| = 1000$), each instantiated randomly from a pool of 1000 resources ($|R| = 1000$), 10000 associations ($|A| = 10000$), and 1000 users ($|U| = 1000$). Each instance of a query was randomly assigned 20 resources ($|R^q| = 20$), and was executed by a default of 50 randomly assigned users ($|U^q| = 50$). The number of users executing a query is a parameter we varied in simulation. These values are arbitrarily chosen to be as large as possible while still allowing fast simulation in software.

We initialized the simulation by (1) generating resources and associations, (2) generating the existing trust of users, (3) generating subsets of query results and users.

We used a standard normal distribution, by default, to assign trust values to each member of A, where all random numbers less than -1 or greater than 1 are replaced with these limits, respectively. The standard deviation of this distribution changes between use case scenarios, and we refer to this parameter as ρ. The larger ρ is, the greater the contrast between trust and distrust in the population of resources. Next, we randomly assigned associations to resources, where the number of assignments to each resource is a random number chosen from a normal distribution with an arbitrarily chosen mean of 6.0 and standard deviation of 5.0 We ensured each resource has at least one association, and each association is chosen randomly, with replacement, using a uniform distribution over A. Using AT and the association assignments, we computed RT for each resource as the mean AT over all of a resource's assigned associations.

For more meaningful results, we select many random samples of R^q and U^q to evaluate. We assign a random subset of U to each U^q, as not all users make all queries, and we assign a random subset of R to each R^q. Both assignments are performed using random selection, with replacement, from uniform distributions over the respective sets.

We derive values for UAT for each user, by selecting which associations each user knows, and what trust a user has in those associations. Not all users have existing trust for all associations, nor do all users have the correct existing trust for the associations they do know. The number of associations a user has existing trust (or distrust) in is a random number selected from a pareto distribution, with a default location of 1.0 and a default shape (power) of $|A|/20 = 500$, offset by a default minimum amount of known associations $|A|/100 = 100$ (note that we are selecting percentages of A, such that $|A|/20$ is 5% of all associations). This distribution is selected with the assumption that most users know a little and some users know a lot, and the offset ensures that each user has prior trust in at least 1% of all associations. As the amount of existing trust users have changes between use case scenarios, we characterize this using the parameters α for the distribution shape and δ for the offset. Given the number of associations each user knows, that number of associations are randomly assigned to each user, with replacement, using a uniform distribution over A. Next we determine the amount of existing trust a user has in each of his known associations. We also use a pareto distribution to determine the "accuracy" of a user's existing trust (how close the user's value is to the "correct" value returned by AT). We have selected a location of 1.0 and a shape of 0.1 for this distribution, making the assumption that most users have existing trust close to the value returned by AT, but some do not. The random value assigned to each user from this distribution is used as the standard deviation in the distribution of Gaussian noise added to the value of AT for each known association. For example, if a user's "accuracy" is chosen to be 0.5 from the pareto distribution, the user's trust in each known association assigned to him would be computed as the value of *AT* for that association plus a random value selected from a normal distribution with mean 0 and standard deviation 0.5. The resulting trust value is restricted to the range [-1, 1]. Given *UAT*, we compute *URT* as the mean *UAT* over all of a resource's associations. If the *UAT* is undefined for a given association, it is not included in the mean. If none of a resource's associations had a *UAT* defined, the resulting *URT* is 0.

8.*2.2 Parameters for Modeling Scenarios*

In each use case scenario, the significant qualities that vary are the distribution of trust over the resources returned, characterized by the parameter σ, and the distribution of existing trust held by users who make the query, characterized by the parameters α and δ. Table 6 shows the parameter values and constraints used to generate data for modeling each of the use case scenarios We set the "trust and distrust" and "distrust only" scenarios so that most users have existing trust for less then 5% of associations. The "sparse trust and distrust" scenario is set so most users have existing trust for less than 1% of the associations. The spread between trust and distrust is set to be greater in the "trust and distrust" scenario than in the others (with a higher standard deviation in the distribution of *AT*), and the "distrust only" scenario has the constraint that users only have existing distrust, and no existing trust. These parameters affect distributions which correspond to the *RT* and the *URT* functions in the model. We have selected very specific and arbitrary ways to compute *RT* and *URT* for our selected use case scenarios, but we believe this is still useful to illustrate our work, which focuses on utilizing trust derived from associations. We note that there are many other ways to compute *RT* and *URT*, which our model can also accommodate.

Use Case Scenario	σ	α	δ
Trust and distrust	3.0	|A|/20	|A|/100
Distrust Only	1.0	|A|/20	|A|/100
Sparse Trust and Distrust	1.0	|A|/100	|A|/500

Table 6: Parameter values used in generating data for simulation of each use case scenario.

8.*2.3 Simulation*

After generating the data described in the above steps, we may execute the simulation. For each pair of R^q and U^q, we computed the *ERT* for each resource (the other trust functions, *RT* and *URT*, were computed during initialization). We used the mean *URT* over all users who executed that query instance (i.e., who are members of U^q) to find the *ERT* of a resource. By this method, the *ERT* is a sample mean, and the *RT* is a population mean. We do not examine the *EAT* in this work, but one way to compute it is finding the mean *ERT* over all resources that have been assigned a given association.

We have performed several simulations to show that the scenarios had been modeled, and that the estimation of trust varies with the qualities of the use case scenario and the number of users. We recall our application of trust in this work: to re-rank query results so that resources which are trusted and relevant (and not just relevant) appear first, and distrusted resources appear last. With this goal in mind, we evaluate the simulated *ERT* by examining:

- **MSE:** the mean squared error, where the error is (*ERT* - *RT*), over all resources in a query result,
- **k-sum:** the sum of the *RT* ("correct" trust) in the first k resources in a result sequence, and
- **ED:** the edit distance of result sequences, original and re-ranked, to the ideal re-ranking.

The MSE provides a measure of how well the *ERT* predicts the *RT* in a given use case scenario, and we use the mean MSE, over all instances of a query, as a single value that characterizes the success of the *ERT* in a specific simulation scenario. Our baseline measure for *ERT* is the error between *RT* and the expected trust value for any resource (which is 0 due to our choice of distribution).

The k-sum is computed for the original query result sequence (R^q), the re-ranked result sequence (T^q) found using the *ERT* as the trust input to ρ, and the ideal result sequence found using the RT as the trust input to ρ. These three values allow us to compare ERT-based re-ranking to the baseline (i.e., no trust-based re-ranking) and the optimal case (i.e., using the unobtainable "correct" trust, *RT*, to re-rank results). We report the mean k-sum over all query instances. The ED is computed for the original result sequence (R^q) and the re-ranked result sequence (T^q), and shows the improvement in re-ranking independent from the magnitude of trust (i.e., the ED is computed using sequence positions, not trust values). We report the mean ED over all query instances for both the original and re-ranked sequences. Our baseline measure for k-sum and ED is to use the original ranking, without any trust-based re-ranking. Lower MSE values suggest more accuracy in predicting trustworthiness, higher k-sum values suggest more trusted resources are being listed first, and lower ED values suggest the re-ranking is closer to ideal.

8.3 Results

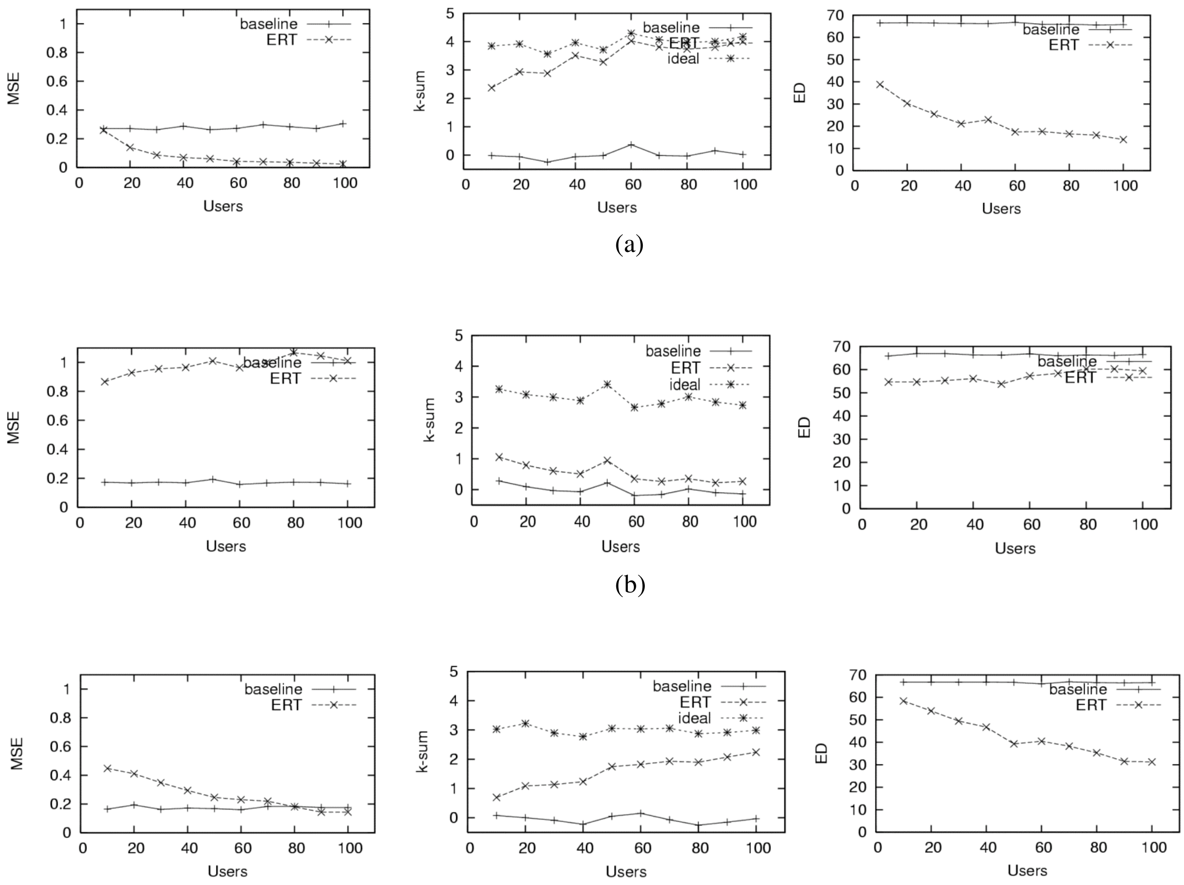
We simulated each of the use case scenarios using our model as described in the previous section. In addition to evaluating the *ERT* in each of the use cases, we also examine the effect of different types of user feedback.

Specifically, we simulate users providing a binary trust decision (rounding the output of URT to either -1 or 1), and we simulate users providing real numbers for trust decisions (keeping the output of URT unchanged). For each use case simulation with binary user feedback, we show the change in our evaluation metrics as the number of users providing trust feedback increases. For brevity, we give only one simulation result where continuous user feedback is used: the k-sum of use case 2.

These results show that we are able to use the model to simulate each of the use cases, and that we can use the model to explore varying user feedback and the success of *ERT* in re-ranking resources with trust. The MSE is given in trust units squared, and due to our choice of τ ([-1, 1]), the maximum possible error is 4.0. The k-sum is also given in trust units, and with k = 10 and our choice of τ, this value falls in the range [-10, 10]. The ED is given in rank units, where a distance of 1 means a resource is off one rank position from its target (i.e., listed 5th instead of the ideal ranking of 6th).

Figure 3 shows the results of the simulation for the three scenarios considered: (a) Trust and Distrust scenario, (b) Distrust Only scenario, (c) Sparse Trust and Distrust scenario. For each scenario, we show the MSE, k-sum, and ED metrics. The first trust and distrust scenario has success in predicting trust with *ERT*, as the MSE decreases quickly as the amount of user feedback increases. This is in contrast to the MSE for the distrust only scenario, where the *ERT* does worse than the baseline (only distrust feedback), and the MSE for the sparse trust and distrust scenario, where the *ERT* starts worse than baseline, and finally improves after enough users provide feedback (sparse existing trust). The k-sum in the trust and distrust scenario rapidly approaches the ideal value. In the distrust only scenario, the k-sum has no significant change with the amount of user feedback, and in the sparse trust and distrust scenario, the k-sum starts close to baseline and gradually approaches ideal with increased user feedback. We observe the same effect for ED, where the trust and distrust scenario starts well and quickly improves, the distrust only scenario starts poorly and does not change significantly, and the sparse trust and distrust scenario starts poorly and improves gradually with more feedback.

Regarding the type of user feedback, our simulation showed in all scenarios and all metrics that continuous user feedback does at least as well as binary user feedback, and mostly does better. Figure 4 shows the results for the distrust only scenario, and only for the k-sum. When using continuous feedback, shown in Figure 4(a), the estimate is always closer to the ideal value than when using binary user feedback Figure 4(b). In all simulations executed, even when the ERT is worse than baseline, the ED always shows improvement over baseline using ERT-based re-ranking.

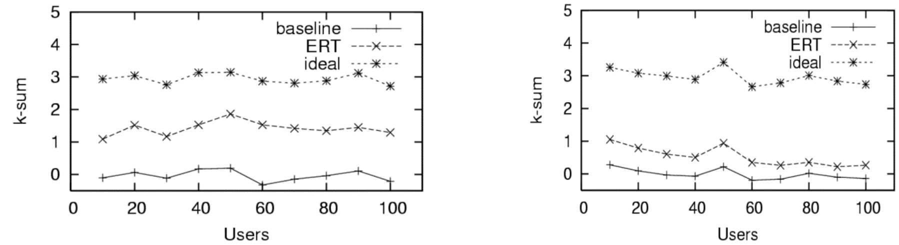


(a)

(b)

(c)

Figure 3: Estimating trust values as user feedback increases in (a) Trust and Distrust scenario, (b) Distrust Only scenario, (c) Sparse Trust and Distrust scenario.



(a)

(b)

Figure 4: Continuous user feedback is better than binary user feedback. Distrust Only scenario, k-sum in trust units using (a) continuous user feedback and (b) binary user feedback.

In summary, these results show that we are able to model the scenarios under the simulation parameters we have selected. We do not know if these parameters accurately reflect the Web, but the simulation still allows us to study the effects of user feedback and different approaches to combining various factors of content trust. We intend to incorporate real-world characteristics of the Web in our simulator in future work.

9. CONCLUSIONS

This paper defined content trust as a new trust metric based on the nature of the information to be trusted. This is in contrast with entity trust, where an entity has a trust metric for any content provided by it regardless of the nature of the information being exchanged. Assessing whether to trust any content is a complex process affected by many factors. Identifying and correlating the factors that influence how trust decisions are made in information retrieval, integration, and analysis tasks becomes a critical capability in a world of open information sources such as the Web. We presented a model for analyzing content trust, its acquisition from users, and its use in improving the ranking of resources returned from a query, and we described important factors in determining content trust. The model was illustrated in the context of three use cases, and the results of model-based simulations of these use cases are presented. We show that the model can be applied to some representative scenarios for Web search, and that the effects of varying types and quantities of user feedback can be explored in the simulation framework.

This work provides a starting point for further exploration of how to acquire and use content trust on the Web. Richer and more comprehensive factors of trust may be included in the model, and integration of existing work in other factors of trust (e.g. recommendations, authority) may be explored. Work in the transitivity of trust may be applied to evaluate the trustworthiness of resources never evaluated by users. More detailed simulations may be performed, leading to the development of a real system for the acquisition and application of content trust on the Web. Additional types of user feedback can be tested, along with the effect of malicious users. Real-world characteristics and qualities of the Web may be incorporated to enable more meaningful exploration of content trust in simulation. Starting with more detailed development and simulations with this model, we plan to design tools to collect information from Web users that will be valuable to estimate content trust.

More research is needed on better mechanisms that could be supported on the Web itself. First, accreditation and attribution to any Web resource supplying content could be captured more routinely. RDF was initially designed to describe this kind of relation among Web resources. Ontologies and more advanced inference could be used to represent institutions, their members, and possibly the strength of these associations. For example, a university could declare strong associations with opinions expressed by its faculty, and less strength in associations with undergraduate students.

In many situations, trust is a judgment on whether something is true and can be corroborated. For example, when agents or services exchange information or engage in a transaction, they can often check if the result was satisfactory, and can obtain feedback on the trust of that entity. In the Web, content trust occurs in an "open loop" manner, where users decide what content to trust but never express whether that trust was well placed or not. Further research is needed on mechanisms to capture how much trust users ultimately assign to open Web sources, while balancing the burden from eliciting feedback during regular use of the Web. There may be very transparent mechanisms based on studying regular browsing and downloading habits.

Users will not be the only ones making trust decisions on the Semantic Web. Reasoners, agents, and other automated systems will be making trust judgments as well, deciding which sources to use when faces with alternatives. Semantic representations of Web content should also enable the detection of related statements and whether they are contradictory. Further research is needed on how to discern which source a reasoner should trust in case of contradictions or missing information. Content trust is a key area of future research for the Semantic Web.

10. ACKNOWLEDGMENTS

We gratefully acknowledge support from the US Air Force Office of Scientific Research (AFOSR) with grant number FA9550-06-1-0031.

11. REFERENCES

Alexander, D. S., Arbaugh, W. A., Keromytis, A. D., and Smith, J. M. (1998). A secure active network environment architecture: Realization in switchware. IEEE Network Magazine 12, 3.

Artz, D., and Gil, Y. (2007). A survey of trust in computer science and the semantic web. Journal of Web Semantics, Volume 5, Issue 2.

Berners-Lee, T. (1999). Weaving the Web. Harper.

Berners-Lee, T. (2000). Semantic Web on XML. Presentation at XML 2000, available from http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html.

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. Scientific American, May 2001.

Berners-Lee, T., Hall, W., Hendler, J., O'Hara, K., Shadbolt, N., Weitzner, D. (2006). A Framework for Web Science. Foundations and Trends in Web Science, Vol 1, No 1, 2006.

Blaze, M., Feigenbaum, J., and Lacy, J. (1996). Decentralized trust management. In Proceedings of the 17th Symposium on Security and Privacy.

Blythe, J., and Gil, Y. (2004). Incremental formalization of document annotations through ontology-based paraphrasing. In Proceedings of the 13th International World Wide Web Conference.

Bonatti, P. and Olmedilla, D. (2005). Driving and monitoring provisional trust negotiation with metapolicies. In POLICY '05: Proceedings of the Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05), pages 14–23, Washington, DC, USA. IEEE Computer Society.

Braynov, S., and Jadliwala, M. (2004). Detecting malicious groups of agents. In Proceedings of the 1st IEEE Symposium on Multi-Agent Security.

Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In Proceedings of the 7th International World Wide Web Conference.

Chirita, A., Nejdl, W., Schlosser, M., and Scurtu, O. (2004). Personalized reputation management in P2P networks. In Proceedings of the Trust, Security, and Reputation Workshop held at the 3rd International Semantic Web Conference.

Chu, Y., Feigenbaum, J., LaMacchia, B., Resnick, P., and Strauss, M. (1997). Referee: Trust management for web applications. World Wide Web Journal.

Cornelli, F., Damiani, E., and Capitani, S. D. (2002). Choosing reputable servents in a P2P network. In Proceedings of the 11th International World Wide Web Conference.

Damiani, E., di Vimercati, D. C., Paraboschi, S., Samarati, P., and Violante, F. (2002). A reputation-based approach for choosing reliable resources in peer-to-peer networks. In CCS '02: Proceedings of the 9th ACM conference on Computer and communications security, pages 207–216, New York, NY, USA. ACM Press.

Dublin core metadata intiative. (2007). http://www.dublincore.org.

Gandon, F. L. and Sadeh, N. M. (2004). Semantic web technologies to reconcile privacy and context awareness. In UbiMob '04: Proceedings of the 1st French-speaking conference on Mobility and ubiquity computing, pages 123–130, New York, NY, USA. ACM Press.

Gil, Y., and Ratnakar. (2002). V. TRELLIS: An interactive tool for capturing information analysis and decision making. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management.

Gil, Y., and Ratnakar, V. (2002). Trusting information sources one citizen at a time. In Proceedings of the 1st International Semantic Web Conference.

Golbeck, J., and Hendler, J. (2004). Inferring reputation on the semantic web. In Proceedings of the 13th International World Wide Web Conference.

Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. (2004). Combating web spam with TrustRank. Technical Report, Stanford University.

Kagal, L., Finin, T. W., and Joshi, A. (2003). A policy based approach to security for the semantic web. In Proceedings of the 2nd International Semantic Web Conference, Lecture Notes in Computer Science, pages 402–418. Springer.

Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. (2003). The EigenTrust algorithm for reputation management in P2P networks. In Proceedings of the 12th International World Wide Web Conference.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM 46, 5.

Kohl, J. and Neuman, B. C. (1993). The kerberos network authentication service. IETF RFC 1510.

Marsh, S. (1994). Formalising Trust as a Computational Concept. PhD thesis, University of Stirling.

Massa, P. and Hayes, C. (2005). Page-rerank: Using trusted links to re-rank authority. In WI ’05: Proceedings of the The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI’05), pages 614–617, Washington, DC, USA. IEEE Computer Society.

Miller, S. P., Neuman, B. C., Schiller, J. I., and Saltzer, J. H. (1987). Kerberos authentication and authorization system. Tech. rep., MIT.

Nejdl, W., Olmedilla, D., and Winslett, M. (2004). Peertrust: Automated trust negotiation for peers on the semantic web. In Proceedings of Secure Data Management, pp. 118–132.

Resnick, P., and Miller, J. (1996). ICS: Internet access controls without censorship. Communications of the ACM.

Rivest, R. L., and Lampson, B. (1996). SDSI: A simple distributed security infrastructure. Technical Report, Laboratory for Computer Science, Massachusetts Institute of Technology. Available from http://theory.lcs.mit.edu/~cis/sdsi.html.

Ruohomaa, S. and Kutvonen, L. (2005). Trust management survey. In Proceedings of iTrust 2005, Lecture Notes in Computer Science, pages 77–92. Springer.

Sabater, J. and Sierra, C. (2005). Review on computational trust and reputation models. Artif. Intell. Rev., 24(1):33–60.

SAML (2007). SAML. www.oasis-open.org/committees/security.

Uszok, A., Bradshaw, J., Jeffers, R., Suri, N., Hayes, P., Breedy, M., Bunch, L., Johnson, M., Kulkarni, S., and Lott, J. (2003). Kaos policy and domain services: Toward a description-logic approach to policy representation, deconfliction, and enforcement. policy, 00:93.

WS-Trust (2007). WS-Trust. http://www.ibm.com/developerworks/library/specification/ws-trust.

XACML (2007). XACML. www.oasis-open.org/committees/xacml.

Zaihrayeu, I., da Silva, P. P., and McGuinnes, D. L. (2005). IWTrust: Improving user trust in answers from the web. In Proceedings of 3rd International Conference on Trust Management.