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
+1 310-822-1511, +1 310-448-9197
{gil,dono}@isi.edu

ABSTRACT

Trust is an integral part of the Semantic Web architecture. While most prior work focuses on entity-centered issues such as authentication and reputation, it does not model the content, i.e. the nature and use of the information being exchanged. This paper discusses *content trust* as an aggregate of other trust measures that have been previously studied. The paper introduces several factors that users consider in deciding whether to trust the content provided by a Web resource. Many of these factors are hard to capture in practice, since they would require a large amount of user input. Our goal is to discern which of these factors could be captured in practice with minimal user interaction in order to maximize the system's trust estimates. The paper also describes a simulation environment that we have designed to study alternative models of content trust.

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms: Design, Reliability, Human Factors, Languages

Keywords: 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. Agents 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 Malibu? What mechanisms will enable these kinds of trust judgments in the Semantic Web?

WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
ACM 1-59593-323-9/06/0005.

Prior work on trust has focused on issues such as reputation and authentication [2, 4, 13, 21]. Such trust representations and metrics do not take into account content trust, i.e. how the nature of information being exchanged affects trust judgments. 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 [11, 12]. However, the approach was tightly coupled to the analysis structures the users were creating with the TRELLIS system.

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 how related work has investigated some of these factors in isolation. 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. CONTENT TRUST IN INFORMATION SOURCES

In the original Semantic Web architecture design, the trust layer was envisioned to address authentication, identification, and proof checking [3], but did not mention trust in the content itself. The Semantic Web makes it possible to represent the content of resources explicitly. This opens the possibility of looking beyond the actors to 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 content directly.

Existing approaches to model trust focus on entities [6, 4, 9, 13, 15, 2], but they only take into account overall interactions across entities and disregard the nature of interactions, i.e. the actual information or content exchanged. This is insufficient in many situations that require making a selection among sources of information. For example, if n entities have low trust, but give a similar answer to a question, one may trust that answer. Conversely, an entity with very high trust may give an answer that contradicts all answers from the n entities with low trust, causing the answer from the

entity with high trust to be distrusted. Therefore, we argue that the degree of trust in an entity is only one ingredient in deciding whether or not to trust the information it provides.

We distinguish between *entity trust* and *content 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. *Content trust* is a trust judgment on a particular piece of information or some specific content provided by an entity in a given context.

Content trust is often subjective, and there are many factors that determine whether content could or should be trusted, and in what context. Some sources 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 sources 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 by traditionally authoritative sources can be proven wrong in light of other information. The source's reputation and trust may still hold, or it may diminish significantly. Finally, sources 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. There is a finer grain of detail in attributing trust to a source with respect to specific statements made by it.

3. FACTORS THAT INFLUENCE CONTENT TRUST

Before describing important factors that influence content trust, we make some useful distinctions regarding what defines a unit of content and how it can be described.

Information sources range from Web sites managed by organizations, to services that provide information in response to specific queries. Sources can be documents that are made available on the Web, static Web pages, or dynamic Web pages created on-demand. These information sources differ 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.

Once we identify a unit of content, many entities 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. 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 [10] 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 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 which should not be assumed to be highly weighted when determining trust.

There are many 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 a financial news source, as opposed to the personal page of an anonymous Internet user.
5. **Direct experience.** The direct interactions 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 entities to a resource (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 has a vested interest in conveying certain information that may be 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 bad reputation a few months ago, may improve its behavior and have earned a better reputation.

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 transfered 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.

Some of these factors cannot be easily captured, such as the context of the need for information, or the bias of a source in certain topics. An important challenge is to determine which of these factors can be captured in practice.

Next, we present an overview of previous research that addresses some of these factors.

4. RELATED WORK

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 [21], typically using public and private keys [18, 20]. 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 [1]. Semantic representations [19] 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 [6].

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 [7]. Popular sources are often deserving of higher trust, but this is not always the case. For example, blogs were ranked high in a number of cases 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". This problem is often discussed in the context of spam [14], 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 [16]. 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.

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 [4, 9]. 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 [13, 15]. Semantic representations [13, 8] of webs of trust and reputation are also applied in distributed and P2P systems.

There are manual and automatic mechanisms to define provenance with resources. The Dublin Core [10] defines a number of aspects related to provenance. Provenance can be captured using semantic annotations of results inferred by reasoners [22], 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 [11, 12], a system that allows users to make trust-related ratings about sources based on the content provided. Users can specify the source attribution for information ex-

tracted during a search and information analysis process to describe the source. 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 source 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 not applicable to Web searches, but only to searches and analyses that followed the structure of TRELLIS. Some later work was done on turning TRELLIS statements into Semantic Web languages [5], but the algorithms mentioned for 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 remaining factors to enable content trust on the Web.

5. ACQUIRING CONTENT TRUST FROM USERS

We have given an overview of factors that users consider when making a trust decision. Many of the trust factors listed (e.g., authority, reputation, popularity, etc.) are being addressed by other research, and we can build on that research, as it provides basic trust values for a resource's associated entities. Other factors are not currently addressed by existing work, and may require capturing additional input from users (e.g., bias, incentive, likelihood, etc.). However, work does exist in computing associations (esp. provenance and authority), and we may use these associations to transfer trust from entities to resources. This approach also allows us to utilize existing trust judgments that do consider other factors. Associations are also central to the Semantic Web, and 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 trust heuristics (e.g., time of creation, bias, appearance, etc.). Moreover, when content trust cannot be determined directly (which is common when searching the Web), associations are the only mechanism through which a trust decision can be made. Therefore, we believe the best place to begin exploring content trust, is through the transfer of trust using a resource's associations. In our initial work, we assume that each association has a single overall trust value, and do not address how that trust value is derived (e.g., possibly as a combination of its popularity, reputation, authority, etc.). We believe our framework can be extended to incorporate those factors explicitly in future work.

Currently, 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,

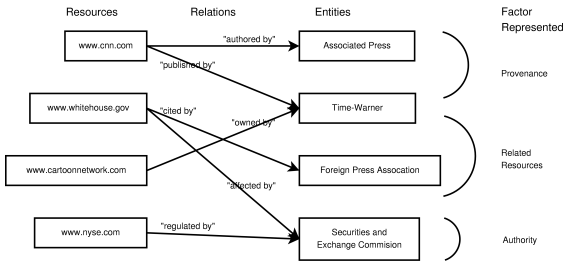

Figure 1: A resource may have multiple associations, and an entity can be related to multiple resources with different relationships.

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. We assume a baseline of topic and popularity to rank search results, and we believe results can be reranked using additional trust factors so that more trustworthy resources appear higher in the results list.

Our next challenge is to determine (1) what information can be captured in practice from users regarding content trust decisions as they perform Web searches, (2) how a user's information can be complemented by automatically extracted information, (3) how is all the information related to the factors outlined above, and (4) how to use this information to derive content trust. Next, we present our approach to model and study 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. MODELING THE ACQUISITION OF CONTENT TRUST FROM USERS

In this section, we describe our model for simulating and studying the use and acquisition of content trust.

A resource, $r \in \mathcal{R}$, 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 $\mathcal{A}$. Each member of $\mathcal{A}$ is an association tuple, $\langle a_r, a_e \rangle$, which contains an association relation, a_r, and an association entity, a_e. 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" ⟩. See Figure 1 for an illustration of resources and associations. We assume the associations for each resource are given.

We will study trust over a fixed time where a set of users, $\mathcal{U}$, make a subset of queries from a set of possible queries, $\mathcal{Q}$. A user, $u \in \mathcal{U}$, queries an information system and analyzes

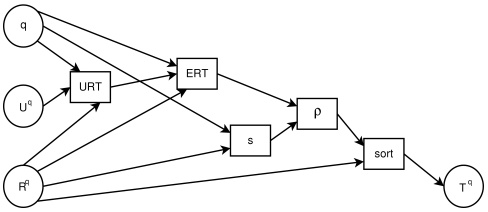

Figure 2: Model of trust to rerank resources, where arrows denote input dependencies.

the results to determine content trust. The set of users who make query q is U^q, a subset of $\mathcal{U}$. 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 r_i^q 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 $UAT : \mathcal{Q}, \mathcal{A}, \mathcal{U} \rightarrow \tau$. 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 $URT : \mathcal{Q}, \mathcal{R}, \mathcal{U} \rightarrow \tau$. 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.

The *association trust*, $AT : \mathcal{Q}, \mathcal{A} \rightarrow \tau$, is the global trust of an association, derived from the UAT of individual users. The *resource trust*, $RT : \mathcal{Q}, \mathcal{R} \rightarrow \tau$, is the 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 or 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 $\mathcal{U}$). The *estimated resource trust*, mapped by the function $ERT : \mathcal{Q}, \mathcal{R} \rightarrow \tau$, 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 $EAT : \mathcal{Q}, \mathcal{A} \rightarrow \tau$, 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 $s^q : \mathcal{R} \rightarrow \mathcal{O}$, where $\mathcal{O}$ is a set of values that can be used to order (rank) resources (e.g., consider $\mathcal{O} = \{0, 1\}$, and if $s^q(r) = 1$, then it is listed before any resource r' where $s^q(r') = 0$). The *trust rerank* function $\rho : \mathcal{O}, \tau \rightarrow \mathcal{O}$, maps an order value and a trust value to a new order value. We can apply this function to rerank 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 reranked sequence of results, T^q, contains the elements of R^q sorted by the output of ρ. Starting with the original sequence of results, and ending with the reranked sequence, Figure 2 illustrates the initial use of our model. 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.

7. 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.

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.

7.1 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.

7.1.1 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.

7.1.2 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.

7.1.3 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.

7.2 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.

7.2.1 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 [17] 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, $\mathcal{O} = \{1\}$, such that in these examples, all query results are assumed to be equally relevant. However, $\mathcal{O}$ 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 ($|\mathcal{Q}| = 1000$), each instantiated randomly from a pool of 1000 resources ($|\mathcal{R}| = 1000$), 10000 associations ($|\mathcal{A}| = 10000$), and 1000 users ($|\mathcal{U}| = 1000$). 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 $\mathcal{A}$, 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 σ. The larger σ 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 $\mathcal{A}$. 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 $\mathcal{U}$ to each U^q, as not all users make all queries, and we assign a random subset of $\mathcal{R}$ 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 $|\mathcal{A}|/20 = 500$, offset by a default minimum amount of known associations $|\mathcal{A}|/100 = 100$ (note that we are selecting percentages of $\mathcal{A}$, such that $|\mathcal{A}|/20$ 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 $\mathcal{A}$. 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.

7.2.2 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 1 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	$\|\mathcal{A}\|/20$	$\|\mathcal{A}\|/100$
Distrust Only	1.0	$\|\mathcal{A}\|/20$	$\|\mathcal{A}\|/100$
Sparse Trust and Distrust	1.0	$\|\mathcal{A}\|/100$	$\|\mathcal{A}\|/500$

Table 1: Parameter values used in generating data for simulation of each use case scenario.

7.2.3 Execution

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.

7.2.4 Evaluation

We have performed several evaluations 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 rerank 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:

1. the mean squared error, where the error is $(ERT - RT)$, over all resources in a query result,
2. the sum of the RT ("correct" trust) in the first k resources in a result sequence, and
3. the edit distance of result sequences, original and reranked, to the ideal reranking.

We refer to these metrics as the MSE, the *k-sum*, and the ED, respectively. 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).

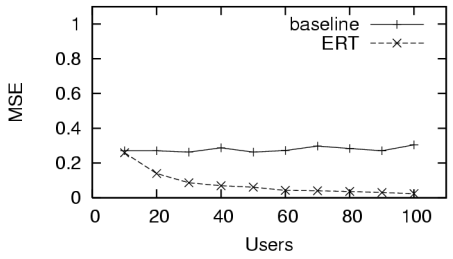

Figure 3: Trust and Distrust MSE, in trust units squared, using binary user feedback.

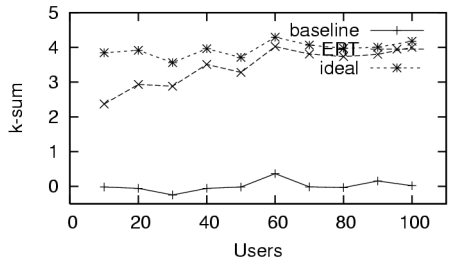

Figure 4: Trust and Distrust, k-sum in trust units, using binary user feedback.

The k-sum is computed for the original query result sequence (R^q), the reranked 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 reranking to the baseline (i.e., no trust-based reranking) and the optimal case (i.e., using the unobtainable "correct" trust, RT, to rerank results). We report the mean k-sum over all query instances. The ED is computed for the original result sequence (R^q) and the reranked result sequence (T^q), and shows the improvement in reranking 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 reranked sequences. Our baseline measure for k-sum and ED is to use the original ranking, without any trust-based reranking.

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 reranking is closer to ideal.

7.3 Results

We have simulated each of the use case scenarios using our model as described in the previous section. In addi-

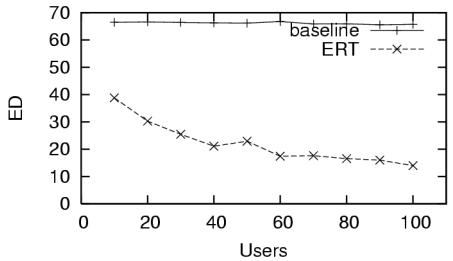

Figure 5: Trust and Distrust, ED in rank units, using binary user feedback.

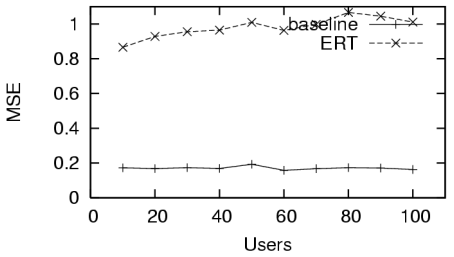

Figure 6: Distrust Only MSE, in trust units squared, using binary user feedback.

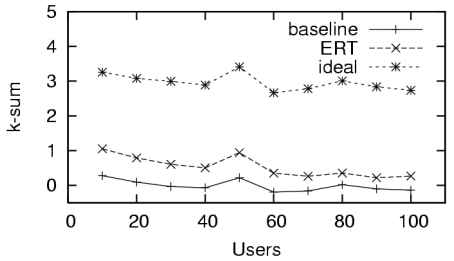

Figure 7: Distrust Only, k-sum in trust units, using binary user feedback.

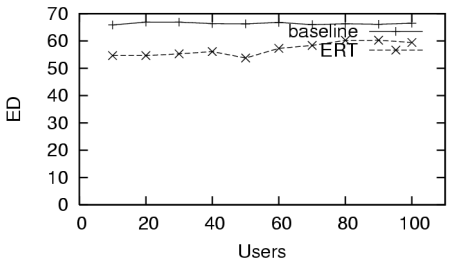

Figure 8: Distrust Only, ED in rank units, using binary user feedback.

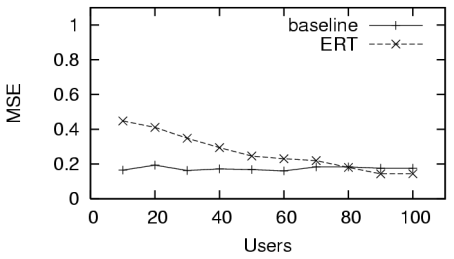

Figure 9: Sparse Trust and Distrust MSE, in trust units squared, using binary user feedback.

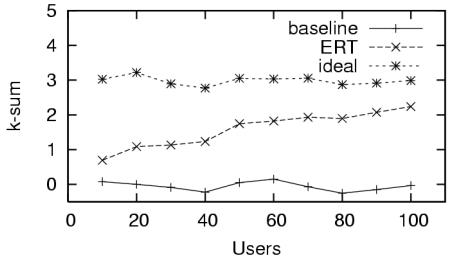

Figure 10: Sparse Trust and Distrust, k-sum in trust units, using binary user feedback.

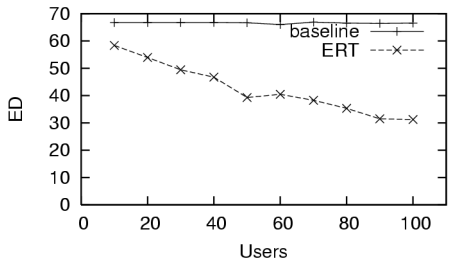

Figure 11: Sparse Trust and Distrust, ED in rank units, using binary user feedback.

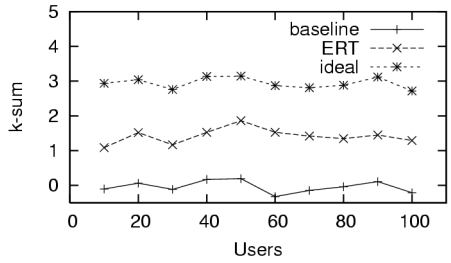

Figure 12: Distrust Only, k-sum in trust units, using continuous user feedback.

tion 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 reranking 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 an resource is off one rank position from its target (i.e., listed 5^{th} instead of the ideal ranking of 6^{th}).

In Figure 3, we see 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 (Figure 6), where the ERT does worse than the baseline (only distrust feedback), and the MSE for the sparse trust and distrust scenario (Figure 9), 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 (Figure 4) rapidly approaches the ideal value. In the distrust only scenario (Figure 7), the k-sum has no significant change with the amount of user feedback, and in the sparse trust and distrust scenario (Figure 10), 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 (Figure 5) starts well and quickly improves, the distrust only scenario (Figure 8) starts poorly and does not change significantly, and the sparse trust and distrust scenario (Figure 11) starts poorly and improves gradually with more feedback. Regarding the type of user feedback, it is consistent in all scenarios and all metrics that continuous user feedback does at least as well as binary user feedback, and mostly does better. For example, the k-sum for the distrust only scenario when using continuous feedback, shown in Figure 12, is always closer to the ideal value than when using binary user feedback (Figure 7). In all simulations executed, even when the ERT is worse than baseline, the ED always shows improvement over baseline using ERT-based reranking.

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.

8. CONCLUSIONS

Assessing whether to trust any information or content provided by a source 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 chart a path to designing 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. New 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. New 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 research area for the Semantic Web.

9. ACKNOWLEDGMENTS

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

10. REFERENCES

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

[2] Artz, D., and Gil, Y. A survey of trust in computer science and the semantic web. *Submitted for publication* (2006).

[3] Berners-Lee, T. *Weaving the Web*. Harper, 1999.

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

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

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

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

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

[9] Chu, Y., Feigenbaum, J., LaMacchia, B., Resnick, P., and Strauss, M. Referee: Trust management for web applications. *World Wide Web Journal 2* (1997).

[10] Dublin core metadata intiative. WWW, 2005. `http://www.dublincore.org/`.

[11] Gil, Y., and Ratnakar, 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* (October 2002).

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

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

[14] Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. Combating web spam with TrustRank. Tech. rep., Stanford, March 2004.

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

[16] Kleinberg, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM 46*, 5 (September 1999), 604–632.

[17] Marsh, S. *Formalising Trust as a Computational Concept*. PhD thesis, University of Stirling, 1994.

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

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

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

[21] Rivest, R. L., and Lampson, B. SDSI: A simple distributed security infrastructure. Tech. rep., MIT, 1996. `http://theory.lcs.mit.edu/~cis/sdsi.html`.

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