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

A Survey of Trust in Computer Science and the Semantic Web

Donovan Artz and Yolanda Gil

Information Sciences Institute
University of Southern California
4676 Admiralty Way, Marina del Rey CA 90292
+1 310-822-1511
gil@isi.edu

March 15, 2007

Abstract

Trust is an integral component in many kinds of human interaction, allowing people to act under uncertainty and with the risk of negative consequences. For example, exchanging money for a service, giving access to your property, and choosing between conflicting sources of information all may utilize some form of trust. In computer science, trust is a widely-used term whose definition differs among researchers and application areas. Trust is an essential component of the vision for the Semantic Web, where both new problems and new applications of trust are being studied. This paper gives an overview of existing trust research in computer science and the Semantic Web.

Keywords: Trust, Web of Trust, Policies, Reputation

1 Introduction

Trust is a central component of the Semantic Web vision (Berners-Lee 1999; Berners-Lee et al 2001; Berners-Lee et al 2006). The Semantic Web stack (Berners-Lee 2000; Berners-Lee et al 2006) has included all along a trust layer to assimilate the ontology, rules, logic, and proof layers. Trust often refers to mechanisms to verify that the source of information is really who the source claims to be. Signatures and encryption mechanisms should allow any consumer of information to check the sources of that information. In addition, proofs should provide a tractable way to verify that a claim is valid. In this sense, any information provider should be able to supply upon request a proof that can be easily checked that certifies the origins of the information, rather than expect consumers to have to generate those proofs themselves through a computationally expensive process. The web motto "Anyone can say anything about anything" makes the web a unique source of information, but we need to be able to understand where we are placing our trust.

Trust has another important role in the Semantic Web, as agents and automated reasoners need to make trust judgements when alternative sources of information are available. Computers will have the challenge to make judgements in light of the varying quality ande truth that these diverse "open" (unedited, uncensored) sources offer. Today, web users make judgments routinely about which sources to rely on since there are often numerous sources relevant to a given query, ranging from institutional to personal, from government to private citizen, from objective report to editorial opinion, etc. These trust judgements are made by humans based on their prior knowledge about a source's perceived reputation, or past personal experience about its quality relative to other alternative sources they may consider. Humans also bring to bear vast amounts of knowledge about the world they live in and the humans that populate the web with information about it. 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 trust judgments are currently in the hands of humans. This will not be possible in the Semantic Web, where humans will not be the only consumers of information. Agents will need to automatically make trust judgments to choose a service or information source while performing a

task. Reasoners will need to judge which of the many information sources available, at times contradicing one another, are more adequate for answering a question. In a Semantic Web where content will be reflected in ontologies and axioms, how will a computer decide what sources to trust when they offer contradictory information? What mechanisms will enable agents and reasoners to make trust judgments in the Semantic Web?

Trust is not a new research topic in computer science, spanning areas as diverse as security and access control in computer networks, reliability in distributed systems, game theory and agent systems, and policies for decision making under uncertainty. The concept of trust in these different communities varies in how it is represented, computed, and used. While trust in the Semantic Web presents unique challenges, prior work in these areas is relevant and should be the basis for future research.

This paper provides an overview of trust research in computer science relevant to the Semantic Web. We focus on relating how different areas define and use trust in a variety of contexts. The paper begins with a general discussion and definitions of trust from the literature. It describes reputation and policies as two broad categories of research to model trust. It then discusses a third category of trust research in designing general computational models of trust. The fourth and final category of research surveyed is trust in information sources. Along the way, we discuss the relevance of the work presented to ongoing and future Semantic Web research.

2 Modeling and Reasoning about Trust

Many have recognized the value of modeling and reasoning about trust computationally. A wide of variety of literature now exists on trust, ranging from specific applications to general models. However, as many authors in the field have noted, the meaning of trust as used by each researcher differs across the span of existing work. In order to give the reader a reference point for understanding trust, we offer three general definitions from existing research. The first definition, from (Mui et al., 2002), refers to past encounters, and may be thought of by some as "reputation-based" trust:

> "[Trust is] a subjective expectation an agent has about another's future behavior based on the history of their encounters."

The next definition, from (Grandison and Sloman, 2000), introduces context and is unique in referring to the "competence" to act (instead of actions, themselves):

> "[Trust is] the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context."

The third definition, from (Olmedilla et al., 2005), applies to many cases in this survey, and it refers to actions and not competence like the previous definition:

> "Trust of a party A to a party B for a service X is the measurable belief of A in that B behaves dependably for a specified period within a specified context (in relation to service X)."

A unifying theme is that trust is only worth modeling when there is a possibility of deception, that is, when there is a chance of a different outcome than what is expected or has been agreed upon.

Two common ways of determining trust are through using *policies* or *reputation*. We adopt these categories from (Bonatti et al., 2005), as they best describe the distinction we observe between the "hard evidence" used in policies, and the estimation of trust used in reputation systems. Policies describe the conditions necessary to obtain trust, and can also prescribe actions and outcomes if certain conditions are met. Policies frequently involve the exchange or verification of *credentials*, which are information issued (and sometimes endorsed using a digital signature) by one entity, and may describe qualities or features of another entity. For example, having the credential of a university degree means its holder has been recognized by the issuing university as having a certain education level. This associates the holder with the university and to those educated in his field. Credentials can be used when trust in the entity itself is unknown, but there is existing trust in what is associated through the entity's credentials.

Reputation is an assessment based on the history of interactions with or observations of an entity, either directly with the evaluator (personal experience) or as reported by others (recommendations or third party verification). How these histories are combined can vary, and recursive problems of trust can occur when using information from others (i.e., can I trust an entity's recommendation about another entity?). At a basic level, both credentials and reputation involve the transfer of trust from one entity to another, but each approach has its own unique problems which have motivated much of the existing work in trust.

Table 1 is a roadmap for this survey, and gives an overview of research areas and references. We organize trust research in four major areas:

1. **Policy-based trust.** Using *policies* to establish trust, focused on managing and exchanging credentials and enforcing access policies. Work in policy-based trust generally assumes that trust is established simply by obtaining a sufficient amount of credentials pertaining to a specific party, and applying the policies to grant that party certain access rights. The recursive problem of trusting the credentials is frequently solved by using a trusted third party to serve as an authority for issuing and verifying credentials.
2. **Reputation-based trust.** Using *reputation* to establish trust, where past interactions or performance for an entity are combined to assess its future behavior. Research in reputation-based trust uses the history of an entity's actions/behavior to compute trust, and may use referral-based trust (information from others) in the absence of (or in addition to) first-hand knowledge. In the latter case, work is being done to compute trust over social networks (a graph where vertices are people and edges denote a social relationship between people), or across paths of trust (where two parties may not have direct trust information about each other, and must rely on a third party). Recommendations are trust decisions made by other users, and combining these decisions to synthesize a new one, often personalized, is another commonly addressed problem.
3. **General models of trust**. There is a wealth of research on modeling and defining trust, its prerequisites, conditions, components, and consequences. Trust models are useful for analyzing human and agentized trust decisions and for operationalizing computable models

of trust. Work in modeling trust describes values or factors that play a role in computing trust, and leans more on work in psychology and sociology for a decomposition of what trust comprises. Modeling research ranges from simple access control polices (which specify who to trust to access data or resources) to analyses of competence, beliefs, risk, importance, utility, etc. These subcomponents underlying trust help our understanding of the more subtle and complex aspects of composing, capturing, and using trust in a computational setting.

4. **Trust in information resources.** Trust is an an increasingly common theme in Web related research regarding whether Web resources and Web sites are reliable. Moreover, trust on the Web has its own range of varying uses and meanings, inluding capturing ratings from users about the quality of information and services they have used, how web site design influences trust on content and content providers, propagating trust over links, etc.. With the advent of the Semantic Web, new work in trust is harnessing both the potential gained from machine understanding, and addressing the problems of reliance on the content available in the web so that agents in the Semantic Web can ultimately make trust decisions autonomously. Provenance of information is key to support trust decisions, as is automated detection of opinions as distinct from objective information.

In the rest of the paper, we devote a section to each of the categories in turn, and we provide a section each on related surveys and concluding remarks. We begin with policies, followed by reputation, due to dependencies in some of the concepts explained. Likewise, the section on general models uses concepts from both policies and reputation research. We cover information sources and the Web last, as we believe research in this area is best explained with knowledge of the previous sections. In categorizing existing work, we do not focus on the individual key contributions, but instead on how trust is used and defined. Many papers may fit under multiple categories, but we have organized references in a way we think is most useful to readers.

Table 1: A categorization of major areas of trust research.

Policy-Based Trust

Network security credentials
(Kohl and Neuman 1993)

Trust negotiation
(Yu et al 2001)
(Yu and Winslett 2003)
(Winslett et al 2002)
(Li et al 2003)
(Nejdl et al 2004)
(Bonatti and Olmedilla 2005)
(Gandon and Sadeh 2004)
(Winsborough et al 2000)
(Seigneur and Jensen 2004)

Security policies and trust languages
(Tonti et al 2003)
(Uszok et al 2003)
(Kagal et al 2003)
(Nielsen and Krukow 2003)
(Carbone et al 2003)
(EHR Policy 2001)
(XACML 2005)
(SAML 2005)
(WS-Trust 2005)
(Becker and Sewell 2004)
(Leithead et al 2004)

Distributed trust management
(Blaze et al 1996)
(Blaze et al 1999)
(Chu et al 1997)
(Kagal et al 2002)

Effect of credential type
(Zheng et al 2002)

Reputation-Based Trust

Decentralization and referral trust
(Abdul-Rahman and Hailes 1997a)
(Abdul-Rahman and Hailes 1997b)
(Yu and Singh 2000)
(Yu and Singh 2002)
(Yu and Singh 2003)
(Sabater and Sierra 2002)
(Beth et al 1994)
(Xiao and Benbasat 2003)
(O'Donovan and Smyth 2005)

Trust metrics in a web of trust
(Goldbeck and Hendler 2004a)
(Goldbeck and Hendler 2004b)
(Stewart 1999)
(Stewart and Zhang 2003)
(Richardson et al 2003)
(Masa and Avesani 2005)
(Guha et al 2004)
(Advogato 2000)
(Chirita et al 2004)
(Ding et al 2004)

Trust in P2P networks and grids
(Kamvar et al 2003)
(Cornelli et al 2002)
(Aberer and Despotovic 2001)
(Damiani et al 2002)
(Olmedilla et al 2005)

Application-specific reputation
(Pirzada and McDonald 2004)
(Dash et al 2004)
(Josang and Ismail 2002)

General Models of Trust

General characteristics of trust
(McKnight and Chervany 1996)
(Gefen 2002)
(Acrement 2002)
(Mui et al 2002)
(Staab et al 2004)

Computational and online trust models
(Marsh 1994)
(Ziegler and Lausen 2005)
(Resnick et al 2000)
(Friedman et al 2000)
(Falcone and Castelfranchi 2004)
(Jonker et al 2004)

Game theory and agents
(Buskens 1998)
(Brainov and Sandholm 1999)
(Ashri et al 2005)
(Ramchurn et al 2003)
(Huynh et al 2004)

Software engineering
(Viega et al 2001)

Trust in Information Resources

Trust concerns in the Web
(Khare and Rifkin 1997)
(Grandison and Sloman 2000)

Trust concerns in the Semantic Web
(Bizer and Oldakowski 2004)
(Berners-Lee 1999)
(O'Hara et al 2004)

Trust Using Hyperlinks
(Gyongy et al 2004)
(Massa and Hayes 2005)
(Brin and Page 1998)
(Kleinberg 1999)

Filterning information based on trust
(Ciolek 1996)
(Clarke et al 2001)
(Downey et al 2005)

Filtering the Semantic Web
(Bizer et al 2005)
(Ding et al 2003)
(Ding et al 2005)
(Ziegler 2004)

Subjectivity analysis
(Riloff et al 2005)
(Stoyanov et al 2005)
(Cardie et al 2004)

Provenance information
(McGuinness 2005)
(Golbeck 2006)
(Zhao et al 2004)
(Wong et al 2005)
(Kim et al 2007)

Content trust
(Gil and Ratnakar 2002)
(Chklovski et al 2003)
(Castelfranchi et al 2003)
(Gil and Artz 2006)

Site design and human factors
(Silence et al 2004)
(Stephens 2004)
(Corritore et al 2001)

3 Policy-based Trust

This section summarizes work using policies to establish trust. Policies allow the expression of when, for what, and even how to determine trust in an entity.

3.1 Network Security Credentials

The application of a policy is performed by considering some set of information about an entity with regard to trust, and this information is commonly a credential. Although the word "credential" is frequently used to refer to "signed" statements about an entity, it lacks a precise common definition across existing work. Many policies rely on credentials, but in general they may utilize a broader range of information that can be used to make trust decisions. An illustrative example of a common alternative to a signed credential occurs in the process of logging into a computer. A valid user name with a correct password must be given to gain access. According to the system's policy, this information "proves" the user is trusted by the computer's administrator. At the same time, a user must keep his password secret, as revealing it to anything other than the computer system will allow others to use the same credential. In more complex examples, it may be undesirable to reveal credentials to another party. When revealing a credential, an entity may sacrifice privacy and reveal information that may be used by others to the entity's disadvantage. For example, most users implicitly trust the computer they log into, but the need to establish trust in both directions is essential for entities providing services on the Web. Evolving work in policies highlight a more complex problem in trust: how much to trust another entiry to see your own credentials when you wish to earn that entity's trust.

Credentials are sometimes implemented using security certificates with digital signatures. Typically in research, a security certificate has the primary role of having one entity vouch for the identity of another, but does not necessarily include credential information. A certificate can be used as a credential if it includes properties about an entity.

The well-known Kerberos protocol (Kohl and Neuman, 1993) is used to exchange credentials. The Kerberos system uses a third party to facilitate the exchange of credentials (digital signatures) between a user and a computer. Kerberos does not determine access rights, but instead enables two parties to securely exchange verifiable credentials.

3.2 Trust Negotiation

An important problem in establishing trust is that revealing a credential may incur a loss of privacy or control of information. Winslett and colleagues (Yu et al., 2001, Yu and Winslett, 2003, Winsborough et al., 2000) have focused on the trade-off between privacy and earning trust. In this work, trust in a particular context is earned by revealing a certain number and type of credentials, and privacy of credential information is lost as the credentials are revealed. An implemented architecture based on these principles is *TrustBuilder* (Winslett et al., 2002), which provides mechanisms for addressing this trade-off. This work builds on a "hard security" view of trust, which means trust is established using traditional security techniques (e.g., authentication, access control, encryption, etc.). In TrustBuilder, trust is earned when sufficient credentials are revealed (but not too many to sacrifice privacy). Making trust decisions requires understanding the risk of revealing a credential, and the benefit of earning trust. Also in TrustBuilder, is the concept of a "credential chain", where trust is transfered transitively through credentials (e.g., if *A* trusts the credentials of *B*, and *B* trusts the credentials of *C*, then *A* may have some trust in the credentials of *C*). The trust management language RT_0 (Li et al., 2003) is designed explicitly to perform credential chaining, and allows for an efficient distributed search to find such chains. Another system is *PeerTrust* (Nejdl et al., 2004), a more recent policy and trust negotiation language that facilitates the automatic negotiation of a credential exchange. Following PeerTrust, is PROTUNE (Bonatti and Olmedilla, 2005), a provisional trust negotiation framework. PROTUNE allows policies with "provisional predicates", where actions may be specified that will satisfy (currently unsatisfied) conditions. In a more specific view, (Gandon and Sadeh, 2004) have proposed using ontologies to enable context-aware applications on the Semantic Web. Context-aware applications will only reveal credentials in the correct context. Others working in this area have contributed ideas on client-server credential exchange (Winsborough et al., 2000), and protecting privacy through generalizing or categorizing credentials (Seigneur and Jensen, 2004).

3.3 Security Policies and Trust Languages

Security research is responsible for many of the first models and descriptions of trust in computer science. Trust is frequently motivated by work in security and policy representation, and trust and

security are related, interdependent concepts with different purposes. In (Tonti et al., 2003), several current policy languages, designed for use in the Semantic Web, are compared and contrasted. A key point in this work is that policy specification for negotiating interactions is essential for building trust, as the rules of negotiation determine how and if trust is achieved. In most trust-related policy languages, the type of trust in mind is typically related to access control. A notable system designed originally for agents, (Uszok et al., 2003) describes the *KAoS* policy language and KAoS "services" used to enforce its policies. The major drive for KAoS has been to enable the use of the same policy in distributed heterogeneous environments and to enable dynamic policy changes. In (Kagal et al., 2003), a policy language (subsequently known as *Rei*) is described which addresses security and privacy issues in the semantic Web, while allowing each entity to specify their own policy. The Rei language uses semantic representations to separate policy from implementation, and models "speech acts" (to programmatically "discuss" a policy at runtime) as a means of negotiation and dynamic policy manipulation.

Several recent efforts in creating security policies have considered how to represent and express trust. In (Nielsen and Krukow, 2003), the authors propose trust replaces key-based security, based on the fact that we can't ever know everything about everyone. Trust in this work is comprised by observations of a user, recommendations from others about that user, and references to other sources of trust on that user. Access control is determined by a user's level of trust, and this work provides a formal policy language in which trust can be proved. In (Carbone et al., 2003), trust is decomposed into different types and qualities, yielding a policy language that allows fine-tuned control over trust decisions using lattices of relative trust values. One example of trust in policy form is the electronic health records policy (EHR Policy, 2001) generated for use with Cassandra (Becker and Sewell, 2004). This policy exemplifies Cassandra's role-based access control approach to trust. Keeping trust and security separate, some policy languages, such as the OASIS extensible access control markup language (XACML, 2005), still assume trust is established through some external system. The OASIS security assertion markup language (SAML, 2005), provides a means for authentication and authorization, but is not able to represent or suggest trust. As a consequence, SAML has the prerequisite that some external system is trusted.

To facilitate the exchange of credentials, several standards for representation of policies and credentials have been proposed. *WS-Trust* (WS-Trust, 2005), an extension of *WS-Security*, specifies how trust is gained through proofs of identity, authorization, and performance. This work literally views trust from a hard security perspective, issuing a "security token" when trust is earned. WS-Trust does not address the trust negotiation process, only its representation.

The Cassandra system (Becker and Sewell, 2004) uses a policy specification language that enforces how trust may be earned through the exchange of credentials. This work is inspired by role-based access control, a context-based system for authorization. (Leithead et al., 2004) uses ontologies to flexibly represent trust negotiation policies (rules used to negotiate trust). Ontologies have more flexibility than set standards, they simplify policy specification, and they enable more information to be specified to control privacy during trust negotiation.

(Olmedilla, 2006) provides a comprehensive overview and comparison of policy languages.

3.4 Distributed Trust Management

A problem in using credentials, is that they are also subject to trust decisions (i.e., can you believe a given credential to be true?). A trusted third party may sign credentials if it has verified or issued them, and in practice, certificate authorities are used to verify signatures. Even with this limited capability, it can be undesirable to have a single authority responsible for deciding who and when someone is trusted. This problem is broadly described as *trust management*. Early work on this problem is found in *PolicyMaker* (Blaze et al., 1996), which called for the separation of security and trust, recognizing the problems allowing individual systems to have separate and different trust policies separate from the common, global authentication and security system. Following PolicyMaker, (Blaze et al., 1999) presents a system called *KeyNote*, which provides a standard policy language which is independent of the programming language used. KeyNote provides more application features than PolicyMaker, and the authors compare their idea of trust management with other existing systems at the time, including *REFEREE* (Chu et al., 1997). However, as seen in more recent work (Kagal et al., 2002), some researchers in security still take a hard security approach to trust (i.e., trust is completely present or absent).

Trust in this work is defined as what is earned after identity and authorization are verified, or rather, after credentials and their claimed association is verified.

(Ruohomaa and Kutvonen, 2005) provides a detailed survey and discussion of alternative approaches for trust management.

3.5 Effect of Credential Type

Some types of credentials affect trust more than others in certain scenarios, and this phenomenon is examined by (Zheng et al., 2002) for agents playing in a variation of the prisoner's dilemma. Trust is measured as the amount of cooperation between two users, and the types of credentials include resumes, text-chats, and pictures of players. The results of this study show that the type of credential affects the amount of trust or distrust received.

4 Reputation-based Trust

Reputation-based trust uses personal experience or the experiences of others, possibly combined, to make a trust decision about an entity. This section explores work in reputation-based trust, a well-defined area of trust research in computer science.

4.1 Decentralization and Referral Trust

Just as in policy-based trust, one solution to obtaining trustworthy reputation information is to consult a central, trusted third party that has had prior experience with the entity in question and can provide an assessment of its reputation. The majority of existing work avoids this solution, and most research focuses explicitly on decentralization for reputation management. Citing the problems with hard security in traditional mechanisms, (Abdul-Rahman and Hailes, 1997a, Abdul-Rahman and Hailes, 1997b) focus on providing a system in which individuals are empowered to make trust decisions rather than rely on a centralized process. The main contribution of this work is to describe a system where it can be acknowledged that malicious entities coexist with the innocent, achieved through a decentralized trust decision process. (Yu and Singh, 2000, Yu and Singh, 2002, Yu and Singh, 2003) describe a decentralized solution to *reputation management*, which allows agents to actively determine trust using reputation

information they receive from other agents. Reputation management avoids a hard security approach by distributing reputation information, allowing individuals to make trust decisions instead of a single, centralized trust management system making the decisions for them. Singh and Yu have provided approaches to using reputation information from external sources, weighting it by the reputation of those sources for providing good information. In this work, a peer that provides trust information about another peer is referred to as a *witness*, and this type of information is more commonly referred to as *referral trust*. (Sabater and Sierra, 2002) also give an approach on how to combine reputation information from the individual and from others while paying attention to context. This enables an agent to specify both who can be trusted and for what they can be trusted. The idea of using referral trust is presented early in trust work in "open networks" by (Beth et al., 1994). This work provides methods for computing degrees of trust in the presence of conflicting information, also departing from the view of hard security. Other work with referral trust includes (Xiao and Benbasat, 2003) and (O'Donovan and Smyth, 2005) for describing how reputation is applied to and affects recommenders.

4.2 Trust in P2P Networks and Grids

A target application of reputation-based trust is to address problems of data quality in peer-to-peer (P2P) networks. There may be no barriers or requirements to publish a file in a P2P network, thus allowing anyone to publish anything under any name with any level (or lack) of quality. Moreover, the availability and reliability of any given node in the network is not guaranteed, thus possibly precluding reliable transfer of data. In the wake of the *PageRank* algorithm (Brin and Page, 1998) for ranking Web sites by authority, the *EigenTrust* algorithm (Kamvar et al., 2003) computes a global reputation value (using PageRank) for each entity. Reputation in this work is the quality of a peer's uploads (e.g., did the file successfully upload?) within a peer-to-peer network. The *P2PRep* system (Cornelli et al., 2002) gives protocols and an algorithm for sharing reputation information with peers in a peer-to-peer network. This work also uses the idea of referral trust in its approach.

Contrasting with the work of Singh and Yu, (Aberer and Despotovic, 2001) claim a more scalable approach, as other reputation-based approaches require the maintenance of a growing

performance history to maintain reputation information. While still using reputation information, this approach uses statistical analysis to characterize trust and reputation so that the computation remains scalable. Embracing the qualities of a peer-to-peer network to provide a more robust method of reputation management, (Damiani et al., 2002) present the *XRep* protocol, which allows for an automatic vote using user's feedback for the best host for a given resource.

(Olmedilla et al., 2005) describes the requirements in supporting trust in "virtual organizations of shared resources", discusses the limitations of existing work on trust in the context of grid computing, and argues that semantic representations can address the requirements outlined.

4.3 Trust Metrics in a Web of Trust

A trust decision can be a transitive process, where trusting one piece of information or information source requires trusting another associated source. For example, one might trust a book and its author because of the publisher, and the publisher may be trusted only because of the recommendation of a friend. Winslett's work in policy-based trust uses (or refers to) "credential chains" (the issuer of one credential is the subject of another), the majority of transitive trust computation has been focused on using reputation. A key recent example of this approach is (Golbeck and Hendler, 2004a, Golbeck and Hendler, 2004b), which describe how trust is computed for the application *TrustMail*. Reputation is defined as a measure of trust, and each entity maintains reputation information on other entities, thus creating a "web", that is called a *web of trust*. The work by Golbeck and Hendler uses ontologies to express trust and reputation information, which then allows a quantification of trust for use in algorithms to make a trust decision about any two entities. The quantification of this trust and associated algorithms are called *trust metrics*.

Given an existing quantification of trust, approaches exist to transfer that trust to other entities which may not have been evaluated for trust. One area of research assumes we are given a web of trust, where a link between two entities mean a trust decision has been made and the value of that trust is known. How trust decisions are made do not matter, as long as the resulting trust values can be quantified. If there is no link between a pair of entities, it means no trust

decision has yet been made. This is the case in which *trust transitivity* can be applied, a simplified example being if *A* trusts *B* and *B* trusts *C*, then *A* trusts *C*. Building on work in reputation management (described earlier as empowering individual agents to make trust decisions instead of a single, central authority making decisions for them), multiple researchers are exploring ways to transfer trust within a web of trust. In (Stewart, 1999, Stewart and Zhang, 2003), a set of hypotheses and experiments are described for testing how trust is transfered between hyperlinks on the Web. Specifically, this work examines how much trust (in the context of a consumer trusting a business for purchasing a product) is transferred from a trusted Web resource to an unevaluated one. The transfer is evaluated considering differing types of links, types of resources, and types of trust in the known source. Other more recent work looks at how to compute trust transitivity given actual quantities for trust or distrust. A key work in this area is (Richardson et al., 2003), whose goal is to provide a means of merging trust that is robust to noise. Emphasizing personalized trust, as opposed to globally computed values, this approach is described as a generalization of PageRank (Brin and Page, 1998) to the Semantic Web. In contrast to the EigenTrust approach described earlier, (Richardson et al., 2003) avoids computing global values by altering the algorithm to produce personalized results for each entity. Likewise, EigenTrust uses specifically computed reputation values, and not with an arbitrarily given quantification of trust. In (Massa and Avesani, 2005), the problem of *controversial users* (those who are both trusted and distrusted) is addressed. This work shows that the globally computed trust value (in a web of trust) for a controversial user may not be as accurate as a locally computed value due to the global disagreement on trust for that user. Golbeck and Hendler's TrustMail also performs a local computation of reputation within a web of trust. A difficult problem addressed in (Guha et al., 2004) is the transitivity of distrust, the main problem being if *A* distrusts *B* and *B* distrusts *C*, we cannot say if *A* trusts *C*. This work also evaluates and ranks several methods for propagating trust and distrust in a given web of trust. Evaluation is performed using data from *Epinions.com*, a common data set used in trust research, where users have provided trust or distrust information about each other's ability to write reviews. Another approach to computing trust transitivity is (Advogato, 2000), in which maximum network flow is computed over a web of trust to find trust between any pair of entities. An advantage to this

approach, is that it is very robust to noise and even attacks altering the given web of trust. In (Chirita et al., 2004), the authors present a method that performs a global computation on reputation values (like EigenTrust) but considers the individual's input to the evaluation as well. This approach uses "personalized page ranks" to disseminate reputation information from individuals while considering referral trust (like P2PRep).

All of these approaches to computation over a web of trust do not consider context, and as a result do not differentiate between "topic specific trust" and referral trust. In contrast, (Ding et al., 2004), presents a method of computing within a web of trust that also considers the domain of knowledge (context), and does so separately from referral trust. This work enumerates several kinds of referral (trust in ability to recommend) and associative (two agents being similar) trust as a result: domain expert (trust in an agent's domain knowledge), recommendation expert (trust in an agent's ability to refer other agents), similar trusting (two agents having similar trust in other agents), and similar cited (two agents being similarly trusted by other agents).

4.4 Application-specific Reputation

Some applications allow for unique ways to harness or use reputation. For the application of routing in ad-hoc networks where some nodes may be more trustworthy for routing packets than others, (Pirzada and McDonald, 2004) present a reputation-based system for deciding which nodes in a network to use for routing traffic. Nodes in the network can indirectly monitor the performance of other nodes nearby, and in this application, a node will only ever need to select a nearby host to trust. This is a good example of a case to apply local computation of reputation. Another specific application is (Dash et al., 2004) for allocating tasks to the best performing agent (instead of agent with best specifications, noting the difference). Using statistics to determine reputation from past performance history, (Josang and Ismail, 2002) present a method to combine reputation feedback data using a beta probability distribution.

5 General Models of Trust

This section summarizes work that presents a broader view on models of trust and the properties of trust. Work in multiple, differing fields is presented, as it is relevant to and frequently cited by computer scientists.

5.1 General Considerations and Properties of Trust

Several papers in social sciences, similar to this survey, have put forth an interpretation of existing research in trust. A frequently cited work is (McKnight and Chervany, 1996), which is noted for its effort to integrate existing work and for its resulting classification of types of trust. The goal of this work was to highlight and find common ground between the many different uses of the word "trust" in social sciences research. Of key importance, are the four qualities that McKnight and Chervany identify as being significant when making a trust decision: competence (ability to give accurate information), benevolence (willingness to expend the effort), integrity (adherence to honest behavior), and predictability (evidence to support that the desired outcome will occur). Alternatively cited, is (Gefen, 2002), which simplifies the trust decision to three of these qualities, leaving out predictability and keeping the others. Gefen stresses the importance of these dimensions in different uses of trust online (e.g. how vulnerable is the agent: is he just window-shopping, or is he a serious buyer), citing a definition from relevant research in management: "trust is a willingness to be vulnerable to the actions of another person or people." In (Acrement, 2002), seven qualities of trust are given from a business management perspective. These qualities share predictability and integrity with McKnight and Chervany's set, and add five more new characteristics specific to the management domain: congruity (actions match claims), reliability, openness (don't keep secrets), acceptance (equal respect among diversity), and sensitivity (pay attention to individuals). An "integrated account" of trust and reputation across disciplines is given in (Mui et al., 2002), which explicitly focuses on deriving a computational model accounting for current work. A key concept used is *reciprocity*: "be nice to others who are nice to you". This work also differentiates trust and reputation, describes how trust can be inferred from reputation, and proposes a probabilistic mechanism for inferring trust given

reputation and reciprocity. (Staab et al., 2004) is an edited series of short articles about different ways to represent, manage, and manipulate the properties of trust.

5.2 Computational and Online Trust Models

The widely-cited 1994 Ph.D. dissertation by Stephen Marsh (Marsh, 1994) is considered the first prominent, comprehensive, formal, computational model of trust. His intent was to address "an imperfect understanding, a plethora of definitions, and informal use in the literature and in everyday life" with regard to trust. Marsh proposed a set of (subjectively set) variables, and a way to combine them to arrive at one continuous value of trust in the range [-1,1]. While the intuitive explanation of this range may be complete distrust to full trust, Marsh actually argues against these meanings at the extremes, saying neither full trust or distrust is actually possible. Marsh identified three types of trust: *basic*, over all contexts; *general*, between two people and all their contexts occurring together; and *situational*, between two people in a specific context. In addition to context, Marsh also identified time as being relevant to each of the variables used to comprise trust. Authors who cite Marsh frequently use a simplification of his work (e.g., trust is a continuous value, and its composition is not of concern) or do not follow his model due to the difficulty of finding values for some variables used to compute trust (e.g., importance, utility, competence, risk, etc.).

Many researchers have endeavored to model and explain the properties of trust and reputation in a computational setting. Different trust metrics are compared against several features in (Ziegler and Lausen, 2005), where the concept of "local group trust computation" is advocated (a compromise between local and global trust computation). The authors make the claim that trust is a "subjective expectation". A method for performing local group trust computation, *Appleseed*, is proposed, and the authors also discuss the meaning and propagation of distrust. Creating a clearer picture for reputation, (Resnick et al., 2000) describes reputation as "important for fostering trust among strangers". This work outlines the qualities of reputation that make it valuable for us on the Internet, and identifies issues in applying reputation (e.g., what reputation does a new user have?). In (Friedman et al., 2000), a general discussion of trust on the Internet is given, outlining ten characteristics of trust in an online interaction. A key point presented is that simply

performing a task is not the same as providing good service or being high quality, which is a problem with automated reputation systems that fail to capture this subtle difference. Also made prominent is the idea that people trust people, not technology, which itself earns (or loses) our trust as an extension of trust in people. In (Falcone and Castelfranchi, 2004), a key idea is dealing with the dynamic nature of trust, and making the realization that an agent that knows he's trusted may act differently from one who does not know his level of trust. As a result, this work attempts to show that "good" reputation is useless without knowledge of the context in which that reputation was earned (e.g., was the agent behaving just to "look good"?). Looking at another aspect of trust dynamics, (Jonker et al., 2004) reports on human experiments showing how positive and negative experiences can change negative and positive trust, respectively. Key results from this work suggest that trust does change with different experiences, and that distrust may be harder to overcome than one would expect.

5.3 Game Theory and Agents

Autonomous agents and multi-agent systems have several uses for trust, and one perspective in related research is game theory. In (Buskens, 1998), the author is a sociologist using a game theoretic approach to show that his proposed heuristics can measure a type of trust from the graph of a social network. Buskens uses a variant of the *Trust Game*, which is analogous to the prisoner's dilemma, but set in a market scenario. Another work using game theory is (Brainov and Sandholm, 1999), which shows that underestimating trust hurts all agents involved, and utility is maximized if the level of trust is mutual. The game defined by this work is again a market-based scenario, where the players are a buyer and a seller. This is another work in which trust is claimed to be a way to deal with uncertainty. Using relationships between agents, (Ashri et al., 2005) claims that rules of trust can be determined from the context and the roles of interacting agents. Specifically mentioned are the general relationships of trade, dependency, competition, and collaboration. In this work, trust exists when it is believed that one agent will not gain at the disadvantage of another agent. Trust in (Ramchurn et al., 2003) is an expectation of agents to exhibit a specific behavior in an interaction based on reputation from various sources. The main focus of this work is combining the sources of reputation, and they refer to

direct experience as *confidence*. The *FIRE* model, presented in (Huynh et al., 2004), is designed for combining multiple sources of trust (reputation, context-based rules, and credentials) in an agent system. A key part of this model is the use of "references" (endorsements of trust from other agents), in cases where no reputation or other sources of trust exist. This feature enables FIRE to provide a trust metric in cases where other models fail due to ignorance about an agent.

(Sabater and Sierra, 2005) and (Ramchurn et al., 2004) both provide excellent in-depth surveys of trust in multi-agent systems. (Josang et al., 2006) provides an overview of trust in web communities interacting through market-like systems and services.

5.4 Software Engineering

In the domain of software engineering, (Viega et al., 2001) declares that trust is a critical consideration citing the trust assumptions (e.g., that a user will enter a certain input) commonly made when developing software. This work also notes that trust is used to deal with uncertainty, when specific requirements are unknown, and the contribution is to describe where requirements can fail to make trust explicit.

6 Trust in Information Resources

This section summarizes relevant work in web and document retrieval, information filtering, representing the sources of information as its provenance trail, and other factors in trusting content of information resources.

6.1 Trust Concerns on the Web

"Trust on the Web" may refer to several different problems, and one perspective on this is given in (Khare and Rifkin, 1997). This work begins by noting a flawed assumption that cryptography provides trust, and continues to point out various applications on the Web that require different kinds of trust. The main contribution of Khare and Rifkin is identifying the distinctions between types of agents, policies, and applications with regard to trust management on the Web. Focusing on trust in Internet applications, which exchange or display information, (Grandison and Sloman, 2000) give a provisional definition and discussion of trust across a wide set of literature, and

explore solutions and applications of trust management (which in their work essentially means the implementation of security policies). The authors make an interesting deviation from the definition of trust we offered in the introduction; they define trust as a belief in an entity's ability and not directly a belief in how an entity will perform.

6.2 Trust Concerns on the Semantic Web

Declaring that there is more to trust than reputation, (Bizer and Oldakowski, 2004) make several claims with the Semantic Web in mind. First, any statements contained in the Semantic Web must be considered as claims rather than facts until trust can be established. Second, this work makes the case that it is too much of a burden to provide trust information that is current. Third, context-based trust matters; in this case, context refers to the circumstances and associations of the target of the trust decision. An example of context is an agent providing a description for an item, where the agent may be a vendor selling that item, or as a consumer advocate reporting on that item. Fourth, it is possible to use "content-based trust", using common sense rules of the world to make a trust decision (e.g., do no trust prices below 50 percent of the average price). Finally, Bizer and Oldakowski recall Tim Berners-Lee's "Oh yeah? " button (Berners-Lee, 1999), where he envisioned functionality in Web browsers that when invoked would give reasons why a Web page or service should be believed. Bizer and Oldakowski build on this idea to provide the justification for trust which will be needed on the Semantic Web. Noting that "trust is at the heart of the Semantic Web vision", (O'Hara et al., 2004) name five trust strategies for agents using the Semantic Web: optimism, pessimism, centralized, investigation, and transitivity. Optimism is to assume trust, pessimism is to assume distrust, centralized is to trust through a single third party, investigation is to collect trust information from others, and transitivity is to use a web of trust. This work refers to trust generally as a "method of dealing with uncertainty".

6.3 Trust Using Hyperlinks

Work exists in learning users' trust in Web sources using the link structure of the Web to transfer trust. Given a small data set of decisions made by users about whether or not Web sites are spam, *TrustRank* (Gyongyi et al., 2004) uses the link structure to other pages to determine whether or not they are also spam. The decision can be interpreted as a trust decision in the context of

finding true and accurate sources of information. (Massa and Hayes, 2005) address the problem of assuming that all Web links are positive endorsements (and indications of trust). Algorithms such as Google's PageRank (Brin and Page, 1998) make this assumption, which does not always hold true. Massa and Hayes propose a minor addition to HTML, enabling the author to specify whether a link is positive, negative, or neither. (Kleinberg, 1999) makes the observation that links encode a human judgment that one page is related to another. Kleinberg describes the concepts of a *hub* and an *authority*, the former being a page that points to many authorities, and the latter being a page that is pointed to by many hubs. The PageRank algorithm exploits Kleinberg's ideas of using links as human encoded judgments of relevance and uses the concept of authorities to compute a heuristic of popularity.

6.4 Filtering Information Based on Trust

Work in information filtering has addressed some of the same problems as work in trust. The concept of *quality* is a common goal (i.e., "high quality information"), as quality often correlates with trust. Quality on the Web is discussed in detail in (Ciolek, 1996), highlighting that massive amounts of Web content are becoming outdated with the rapid pace of change in the Web. More recently, the field of question answering is an area of research that may use the Web as a source of answers for given queries. In (Clarke et al., 2001), it is noted that many answers may be returned for a given query, and one of these must be selected as the answer. While trust is not mentioned, the problem can be characterized as determining which answer to trust. The proposed solution is to assume the answer occurring most frequently is correct. In the field of information extraction, the goal is to extract information from unlabeled text (i.e., without semantic markup). One question arising from this work is "can an automatically determined label be trusted? ". A model given in (Downey et al., 2005) shows that the magnitude of redundancy of information (i.e., the frequency of occurrence) can be used as metric for the accuracy (or trustworthiness) of a computed label.

6.5 Filtering the Semantic Web

Information filtering is becoming an increasingly significant area of research as the amount of information available, specifically on the Web, continues to grow. After relevant information is

filtered, there is still a question of whether that information can be trusted. In many cases, filtering still results in too much relevant information, and the most trusted source or content is desired. For the Semantic Web, (Bizer et al., 2005) have created a browser which filters content based on a user specified policy. These policies, written in the *TriQL.P* query language, allow specification of requirements for the context, content, and source of information. The implementation of this browser includes a mechanism that displays to the user justification of why a Web site should be trusted. In (Ding et al., 2003), agents are enabled to use both context and reputation to determine what information to trust in the Semantic Web. This work employs referral trust to collect reputation, and it relies on the richness of the Semantic Web to determine context. The result is the ability to ask another agent "which agent can I trust to get the weather? ". In related work, (Ding et al., 2005) provides a method for picking information sources using both provenance and computation over a web of trust. Assuming that provenance can be determined, a method is given for using this information to filter more trusted sources. The work from this group also incorporates the concept of ignorance (i.e., not having any information about trust). Recommender systems are common on the Web, and may filter information based on recommendations and/or trust ratings. An example considering the Semantic Web is (Ziegler, 2004), where a "taxonomy" is used to score the similarity between profiles of users' interests. Trust values, or recommendations, are computed within a group of "similar" users, and the resulting information is filtered accordingly.

6.6 Subjectivity Analysis

Although information retrieval pioneered some of the approaches used now on the Web for locating relevant information sources, trust-based retrieval is a relatively recent focus in that area of research. Trust in information retrieval is motivated by the need for not just relevant documents, but high-quality documents as well (Zhu and Gauch, 2000). One approach to this is *subjectivity analysis*, which aims at distinguishing true facts from subjective opinions (Riloff et al., 2005).

Trust is also an important area in question answering, since contradictory answers can be obtained from diverse sources in answer to a question. Sometimes opinions are often filtered out

in question answering tasks so that only objective facts are returned as answers (Stoyanov et al., 2005). In other contexts, detecting opinions is useful when no single ground truth can be provided in answer to a question, and instead multiple perspectives are summarized as the answer provided (Cardie et al 2004).

6.7 Provenance of Information

The details regarding the sources and origins of information (e.g., author, publisher, citations, etc.) are referred to as *provenance*, and they serve as a means to evaluate trust. Provenance representation and tracking has been studied in the context of information sources. (McGuinness, 2004) uses semantic annotations to represent the provenance of any results inferred by reasoners, including explanations of reasoning steps and axioms utilized, as well as descriptions of the original data sources. In (Golbeck, 2006), both provenance and the semantic Web are used to infer trust relationships. Provenance establishes a relationship between people and information, and the semantic Web contains social network data used to compute trust between people.

Provenance has been studied in the context of scientific data analysis, especially when generated by simulation and computation. Semantic web technologies have shown to be effective in representing application-relevant provenance information that explains how results are obtained through workflows of computations (Zhao et al., 2004, Wong et al., 2005, Kim et al 2007). (Simmhan et al., 2005) and (Moreau and Ludaescher 2007) provide overviews of provenance research in scientific applications.

6.8 Content Trust

Trust on the Web is needed to make decisions when information conflicts or is non-authoritative. In (Gil and Ratnakar, 2002a, Gil and Ratnakar, 2002b, Chklovski et al., 2003), a system called *Trellis* is introduced, which derives consensus trust on information resources in a community of users as they use or dismiss sources for information analysis tasks. Also examining trust in information sources, (Castelfranchi et al., 2003) proposes a model for making trust decisions about sources, differentiating internal and external attributes affecting trust in a source. The authors note that the composition of inputs to a trust decision affects the outcome of the decision,

and thus the decision itself can not be characterized by a final probability. This observation might be restated that the inputs together form part of the context in which trust is being determined. Also acknowledged is that "attribution of trust is a very complex task", a problem that is exemplified on the Web, as the sources behind information are not always clear or correct. Specifically for trust in information sources, four types of inputs to a trust decision are given: direct experience, categorization (generalization to or from something known), reasoning (application of common sense or rules to verify truth), and reputation. (Gil and Artz, 2006) differentiates trust in a given source from trust in a specific piece of content provided by that source, where trust in one does not always indicate trust in the other. For example, a trusted source may inadvertently issue a patently false statement, or a typically distrusted source may post information that is trustworthy. A key focus of this work is *content trust*, how it may be derived, and how it may be captured and used on the Semantic Web.

6.9 Site Design and User Interaction

The elements of a Web site considered by users when making trust decisions on the Web is explored in (Sillence et al., 2004). A key finding is that in spite of the personal risk and instructions to do otherwise, users in this study consistently examined design factors when making a trust judgment. Focusing on small hotels, (Stephens, 2004) performs experiments to test a proposed "Integrated Trust Model" for Web sites, which includes multiple design elements (e.g., page layout, style, graphics, etc.). Several factors are shown to be more important for earning trust, at least in the context of gaining customers (seeking a small hotel). Related to Web design, researchers in human-computer interaction have outlined in (Corritore et al., 2001) the importance of establishing trust with users online. In this context, the authors observe that trust is multi-dimensional, which is a cause of a current lack of agreement on trust issues, citing trust research across multiple fields. The authors note that trust is used to decrease complexity, and identify existing work in human factors that points to trust as necessary for users to believe computers.

7 Discussion

Trust has been studied in social sciences, business and management, and psychology, before it became central to computer science research. Considering the research we have reviewed, there are several dimensions to describe trust:

1. *Target*: The entity which is being evaluated or given trust varies with the perspective of the problem. Users are the target of trust in access control systems. Networks are trusted by agents or users when using communication channels. When seeking a reliable service, agents or services become the target of trust. On the Web, we can trust agents providing content, on even make trust judgements on the content itself.
2. *Representation*: There are many ways that trust can be digitally encoded. Credentials include digital signatures and tokens. Agents may carry histories of past interactions with other agents. Users may employ social networks, or webs of trust, to determine trust in an unknown correspondent. Semantic Web work includes detailed ontologies for trust policies, trust negotiation, access control, and data provenance.
3. *Method*: Determining trust can be accomplished through many methods. Hard security uses identification and authorization alone to decide complete trust in a user. Many Internet applications use the exchange of credentials (i.e., digital signatures) to establish trust before engaging in a transaction. Agents may use their histories of past interactions, or other agents' histories to determine trust through reputation. In many applications, including information retrieval, trust may be determined through transfer of trust from associated entities.
4. *Management*: The entity or entities that determine trust can vary with the application. In many traditional systems, a single service acts as a trusted third party to mediate the establishment of trust between two unknown agents or users. In more recent work, there is a push for decentralization of control of the trust decision, including the enablement of individual agents to make their own trust decisions. For system-wide or global trust, voting mechanisms or other forms of consensus may be used to collect individual trust decisions.

5. *Computation*: Trust may be quantified and computed in many ways. Some approaches, including those harnessing the Semantic Web, choose discrete trust values (e.g., *trust*, *distrust*, or *neutral*), while others, especially when computation is needed, choose a continuous numerical range. Algorithms for how trust is transferred, combined, or resolved can range from a simple average, to computing eigenvalues on graph adjacency matrices. Many approaches compute trust assuming time is static, while others may account for the changes in trust over time. In cases where trust information is large or always changing, several approaches argue for local computation of trust, rather than a globally consistent value.
6. *Purpose*: The need for trust spans all aspects of computer science, and each situation places different requirements on trust. Human users, software agents, and increasingly, the machines that provide services all need to be trusted in various applications or situations. The communication channels between computers and users, and the content exchanged between computers and users also require trust, in both directions, for real world use. Trust can be used to protect data, to find accurate information, to get the best quality service, and even to bootstrap other trust evaluations.

Trust may be better seen as a motivating concept underlying many problems and contexts rather than as a precise idea to be studied under a uniform framework.

8 Conclusions

Trust research in the Semantic Web poses new challenges that can be better met by building on the diverse but significant body of work in modeling trust in computer science. In this paper, we have identified four broad categories of existing work in trust and given a brief overview of literature in each category. We have discussed the relevance of each of these areas to important aspects of ongoing and future Semantic Web research.

Acknowledgements

We would like to thank the anonymous reviewers for their valuable comments and feedback on this work. We gratefully acknowledge support from the US Air Force Office of Scientific Research (AFOSR) with grant number FA9550-06-1-0031.

References

Abdul-Rahman, A. and Hailes, S. (1997a). A distributed trust model. In *Proceedings of the New Security Paradigms Workshop*, pages 48–60. ACM.

Abdul-Rahman, A. and Hailes, S. (1997b). Using recommendations for managing trust in distributed systems. In *Proceedings of IEEE International Conference on Communication*.

Aberer, K. and Despotovic, Z. (2001). Managing trust in a peer-2-peer information system. In Paques, H., Liu, L., and Grossman, D., editors, *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM01)*, pages 310–317. ACM Press.

Acrement, B. (2002). Elements for building trust: Do your management skills measure up? `http://www.imakenews.com/smei/e_article000051474.cfm`.

Advogato (2000). Advogato's trust metric. `http://www.advogato.org/trust-metric.html`.

Ashri, R., Ramchurn, S. D., Sabater, J., Luck, M., and Jennings, N. R. (2005). Trust evaluation through relationship analysis. In *Proceedings of the 4th International Joint Conference on autonomous Agents and Multi-Agent Systems*, pages 1005–1012.

Becker, M. Y. and Sewell, P. (2004). Cassandra: Distributed access control policies with tunable expressiveness. In *Proceedings of the 5th IEEE International Workshop on Policies for Distributed Systems and Networks*.

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.

Beth, T., Borcherding, M., and Klein, B. (1994). Valuation of trust in open networks. In *Proceedings of the 3rd European Symposium on Research in Computer Security*, pages 3–18.

Bizer, C., Cyganiak, R., Gauss, T., and Maresch, O. (2005). The TriQL.P browser: Filtering information using context- , content- and rating-based trust policies. In *Proceedings of the Semantic Web and Policy Workshop at the 4th International Semantic Web Conference*.

Bizer, C. and Oldakowski, R. (2004). Using context- and content-based trust policies on the semantic web. In *WWW Alt. '04: Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, pages 228–229, New York, NY, USA. ACM Press.

Blaze, M., Feigenbaum, J., Ioannidis, J., and Keromytis, A. D. (1999). The role of trust management in distributed system security. *Lecture Notes in Computer Science*, 1603:185–210.

Blaze, M., Feigenbaum, J., and Lacy, J. (1996). Decentralized trust management. In *Proceedings of IEEE Symposium on Security and Privacy*, pages 164–173.

Bonatti, P., Duma, C., Olmedilla, D., and Shahmehri, N. (2005). An integration of reputation-based and policy-based trust management. In *Proceedings of the Semantic Web Policy Workshop*.

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.

Bonatti, P. A., Shahmehri, N., Duma, C., Olmedilla, D., Nejdl, W., Baldoni, M., Baroglio, C., Martelli, A., Patti, V., Coraggio, P., Antoniou, G., Peer, J., and Fuchs, N. E. (2004). Rule-based policy specification: State of the art and future work. Technical Report Project deliverable D1, Working Group I2, EU NoE REWERSE.

Brainov, S. and Sandholm, T. (1999). Contracting with uncertain level of trust. In *EC '99: Proceedings of the 1st ACM conference on Electronic commerce*, pages 15–21, New York, NY, USA. ACM Press.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.

Buskens, V. (1998). The social structure of trust. *Social Networks*, 20:265–289.

Carbone, M., Nielsen, M., and Sassone, V. (2003). A formal model for trust in dynamic networks. In *Proceedings of International Conference on Software Engineering and Formal Methods*. IEEE Computer Society.

Cardie, C., Wiebe, J., Wilson, T., and Litman, D. (2004) Low-Level Annotations and Summary Representations of Opinions for Multiperspective Question Answering. In M. Maybury (Ed), New Directions in Question Answering , AAAI Press/MIT Press.

Castelfranchi, C., Falcone, R., and Pezzulo, G. (2003). Trust in information sources as a source for trust: a fuzzy approach. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 89–96, New York, NY, USA. ACM Press.

Chirita, P.-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*.

Chklovski, T., Gil, Y., Ratnakar, V., and Lee, J. (2003). Trellis: Supporting decision making via argumentation in the semantic web. In *Proceedings of the 2nd International Semantic Web Conference*.

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

Ciolek, M. T. (1996). The six quests for the electronic grail: Current approaches to information quality in WWW resources. *Review Informatique et Statistique dans les Sciences humaines (RISSH)*, 1–4:45–71.

Clarke, C. L. A., Cormack, G. V., and Lynam, T. R. (2001). Exploiting redundancy in question answering. In *SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 358–365, New York, NY, USA. ACM Press.

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*.

Corritore, C. L., Wiedenbeck, S., and Kracher, B. (2001). The elements of online trust. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 504–505, New York, NY, USA. ACM Press.

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.

Dash, R. K., Ramchurn, S. D., and Jennings, N. R. (2004). Trust-based mechanism design. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 748–755, Washington, DC, USA. IEEE Computer Society.

Ding, L., Kolari, P., Finin, T., Joshi, A., Peng, Y., and Yesha, Y. (2005). On homeland security and the semantic web: A provenance and trust aware inference framework. In *Proceedings of the AAAI Spring Symposium on AI Technologies for Homeland Security*. AAAI Press.

Ding, L., Kolari, P., Ganjugunte, S., Finin, T., and Joshi, A. (2004). Modeling and evaluating trust network inference. In *Proceedings of the 7th International Workshop on Trust in Agent Societies at AAMAS*.

Ding, L., Zhou, L., and Finin, T. (2003). Trust based knowledge outsourcing for semantic web agents. In *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence*.

Downey, D., Etzioni, O., and Soderland, S. (2005). A probabilistic model of redundancy in information extraction. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*.

EHR Policy (2001). Electronic health records policy. `http://www.show.scot.nhs.uk/sehd/publications/DC20011220IMTEHRPol.pdf`.

Falcone, R. and Castelfranchi, C. (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 740–747, Washington, DC, USA. IEEE Computer Society.

Friedman, B., Peter H. Khan, J., and Howe, D. C. (2000). Trust online. *Commun. ACM*, 43(12):34–40.

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.

Gefen, D. (2002). Reflections on the dimensions of trust and trustworthiness among online consumers. *SIGMIS Database*, 33(3):38–53.

Gil, Y. and Artz, D. (2006). Towards content trust of web resources. In *Proceedings of the 15th International World Wide Web Conference*.

Gil, Y. and Ratnakar, V. (2002a). Trellis: An interactive tool for capturing information analysis and decision making. In *EKAW '02: Proceedings of the 13th International*

Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, pages 37–42, London, UK. Springer-Verlag.

Gil, Y. and Ratnakar, V. (2002b). Trusting information sources one citizen at a time. In *ISWC '02: Proceedings of the First International Semantic Web Conference on The Semantic Web*, pages 162–176, London, UK. Springer-Verlag.

Golbeck, J. (2006). Combining provenance with trust in social networks for semantic web content filtering. In *Proceedings of the International Provenance and Annotation Workshop*.

Golbeck, J. and Hendler, J. (2004a). Accuracy of metrics for inferring trust and reputation. In *Proceedings of the 14th International Conference on Knowledge Engineering and Knowledge Management*.

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

Grandison, T. and Sloman, M. (2000). A survey of trust in internet applications. *IEEE Communications Surveys and Tutorials*, 4(4):2–16.

Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA. ACM Press.

Gyongyi, Z., Garcia-Molina, H., and Pedersen, J. (2004). Combating web spam with trustrank. In *Proceedings of the 30th International Conference on Very Large Data Bases*, pages 271–279.

Huynh, T. D., Jennings, N. R., and Shadbolt, N. R. (2004). FIRE: An integrated trust and reputation model for open multi-agent systems. In *Proceedings of the 16th European Conference on Artificial Intelligence*.

Jonker, C. M., Schalken, J. J., Theeuwes, J., and Treur, J. (2004). Human experiments in trust dynamics. *Lecture Notes in Computer Science*, 2995:206–220.

Josang, A. and Ismail, R. (2002). The beta reputation system. In *Proceedings of the 15th Bled Conference on Electronic Commerce*.

Josang, A., Ismail, R., and Boyd, C. (2006). A survey of trust and reputation systems for online service provision. *Decision Support Systems*. in press.

Kagal, L., Finin, T., and Joshi, A. (2002). Developing secure agent systems using delegation based trust management. In *Workshop on Security of Mobile MultiAgent Systems held at Autonomous Agents and MultiAgent Systems*.

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 *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 640–651, New York, NY, USA. ACM Press.

Khare, R. and Rifkin, A. (1997). Weaving a web of trust. *Journal of the World Wide Web*, 2(3):77–112.

Kim, J., Deelman, E., Gil, Y., Mehta, G. and Ratnakar, V. (2007). Provenance Trails in the Wings/Pegasus workflow system. *Journal of Computation and Concurrency: Practice and Experience,* Special issue on the First Provenance Challenge, L. Moreau and B. Ludaescher (Eds).

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

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

Leithead, T., Nejdl, W., Olmedilla, D., Seamons, K. E., Winslett, M., Yu, T., and Zhang, C. C. (2004). How to exploit ontologies for trust negotiation. In *ISWC Workshop on Trust, Security, and Reputation on the Semantic Web*, volume 127 of *CEUR Workshop Proceedings*, Hiroshima, Japan. Technical University of Aachen (RWTH).

Li, N., Winsborough, W. H., and Mitchell, J. C. (2003). Distributed credential chain discovery in trust management. *Journal of Computer Security*, 11(1):35–86.

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

Massa, P. and Avesani, P. (2005). Controversial users demand local trust metrics: an experimental study on epinions.com community. In *Proceedsings of the 25th American Association for Artificial Intelligence Conference*.

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.

McGuinness, D. L. (2004). Question answering on the semantic web. *In IEEE Intelligent Systems*, 19(1).

McKnight, D. H. and Chervany, N. L. (1996). The meanings of trust. Technical Report 94-04, Carlson School of Management, University of Minnesota.

Moreau, L., and Ludaescher, B (Eds). (2007). Special issue on the First Provenance Challenge, *Journal of Computation and Concurrency: Practice and Experience,*

Mui, L., Mohtashemi, M., and Halberstadt, A. (2002). A computational model of trust and reputation. In *Proceedings of the 35th International Conference on System Science*, pages 280–287.

Nejdl, W., Olmedilla, D., and Winslett, M. (2004). Peertrust: Automated trust negotiation for peers on the semantic web. In *Proceedings of Workshop on Secure Data Management in a Connected World in conjunction with the 30th International Conference on Very Large Data Bases*, pages 118–132.

Nielsen, M. and Krukow, K. (2003). Towards a formal notion of trust. In *PPDP '03: Proceedings of the 5th ACM SIGPLAN international conference on Principles and practice of declaritive programming*, pages 4–7, New York, NY, USA. ACM Press.

O'Donovan, J. and Smyth, B. (2005). Trust in recommender systems. In *IUI '05: Proceedings of the 10th international conference on Intelligent user interfaces*, pages 167–174, New York, NY, USA. ACM Press.

O'Hara, K., Alani, H., Kalfoglou, Y., and Shadbolt, N. (2004). Trust strategies for the semantic web. In *Proceedings of Workshop on Trust, Security, and Reputation on the Semantic Web, 3rd International Semantic Web Conference*.

Olmedilla, D. (2006). Security and privacy on the semantic web. In Petkovic, M. and Jonker, W., editors, *Security, Privacy and Trust in Modern Data Management*. Springer.

Olmedilla, D., Rana, O., Matthews, B., and Nejdl, W. (2005). Security and trust issues in semantic grids. In *Proceedings of the Dagsthul Seminar, Semantic Grid: The Convergence of Technologies*, volume 05271.

Pirzada, A. A. and McDonald, C. (2004). Establishing trust in pure ad-hoc networks. In *CRPIT '04: Proceedings of the 27th conference on Australasian computer science*, pages 47–54, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Ramchurn, S. D., Huynh, D., and Jennings, N. R. (2004). Trust in multi-agent systems. *Knowledge Engineering Review*, 19(1):1–25.

Ramchurn, S. D., Sierra, C., Godo, L., and Jennings, N. R. (2003). A computational trust model for multi-agent interactions based on confidence and reputation. In *Proceedings of the 6th International Workshop of Deception, Fraud and Trust in Agent Societies*, pages 69–75.

Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. (2000). Reputation systems. *Commun. ACM*, 43(12):45–48.

Richardson, M., Argawal, R., and Domingos, P. (2003). Trust management for the semantic web. In *Proceedings of the Second International Semantic Web Conference*, pages 351–368. Spring-Verlag.

Riloff, E., Wiebe, J., and Phillips, W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proceedings of the 20th National Conference on Artificial Intelligence*.

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. (2002). Reputation and social network analysis in multi-agent systems. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 475–482, New York, NY, USA. ACM Press.

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

SAML (2005). SAML. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security.

Seigneur, J.-M. and Jensen, C. D. (2004). Trust enhanced ubiquitous payment without too much privacy loss. In *SAC '04: Proceedings of the 2004 ACM symposium on Applied computing*, pages 1593–1599, New York, NY, USA. ACM Press.

Sillence, E., Briggs, P., Fishwick, L., and Harris, P. (2004). Trust and mistrust of online health sites. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 663–670, New York, NY, USA. ACM Press.

Simmhan, Y., Plale, B., and Gannon, D. (2005). A survey of data provenance in e-science. *Special Interest Group on Management of Data Record*, 34(3):31–36.

Staab, S., Bhargava, B., Lilien, L., Rosenthal, A., Winslett, M., Sloman, M., Dillon, T. S., Chang, E., Hussain, F. K., Nejdl, W., Olmedilla, D., and Kashyap, V. (2004). The pudding of trust. *IEEE Intelligent Systems*, 19(5):74–88.

Stephens, R. T. (2004). A framework for the identification of electronic commerce design elements that enable trust within the small hotel industry. In *ACM-SE 42: Proceedings of the 42nd annual Southeast regional conference*, pages 309–314, New York, NY, USA. ACM Press.

Stewart, K. J. (1999). Transference as a means of building trust in world wide web sites. In *ICIS '99: Proceeding of the 20th international conference on Information Systems*, pages 459–464, Atlanta, GA, USA. Association for Information Systems.

Stewart, K. J. and Zhang, Y. (2003). Effects of hypertext links on trust transfer. In *ICEC '03: Proceedings of the 5th international conference on Electronic commerce*, pages 235–239, New York, NY, USA. ACM Press.

Stoyanov, V., Cardie, C., and Wiebe, J. (2005). Multi-perspective question answering using the opqa corpus. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language*.

Tonti, G., Bradshaw, J. M., Jeffers, R., Montanari, R., Suri, N., and Uszok, A. (2003). Semantic web languages for policy representation and reasoning: A comparison of kaos, rei, and ponder. In *Proceedings of the 2003 International Semantic Web Conference*, pages 419–437.

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.

Viega, J., Kohno, T., and Potter, B. (2001). Trust (and mistrust) in secure applications. *Commun. ACM*, 44(2):31–36.

Winsborough, W. H., Seamons, K. E., and Jones, V. E. (2000). Automated trust negotiation. In *Proceedings of the DARPA Information Survivability Conference and Exposition*, pages 88–102. IEEE Press.

Winslett, M., Yu, T., Seamons, K. E., Hess, A., Jacobson, J., Jarvis, R., Smith, B., and Yu, L. (2002). Negotiating trust on the web. *IEEE Internet Computing*, 6(6):30–37.

Wong, S. C., Miles, S., Fang, W., Groth, P., and Moreau, L. (2005). Provenance-based validation of e-science experiments. In *Proceedings of the 4th International Semantic Web Conference*, volume 3729 of *Lecture Notes in Computer Science*, pages 801–815.

WS-Trust (2005). WS-Trust. http://www-128.ibm.com/developerworks/library/specification/ws-trust/.

XACML (2005). XACML. http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml.

Xiao, S. and Benbasat, I. (2003). The formation of trust and distrust in recommendation agents in repeated interactions: a process-tracing analysis. In *ICEC '03: Proceedings of the 5th international conference on Electronic commerce*, pages 287–293, New York, NY, USA. ACM Press.

Yu, B. and Singh, M. P. (2000). A social mechanism of reputation management in electronic communities. In *CIA '00: Proceedings of the 4th International Workshop on Cooperative Information Agents IV, The Future of Information Agents in Cyberspace*, pages 154–165, London, UK. Springer-Verlag.

Yu, B. and Singh, M. P. (2002). An evidential model of distributed reputation management. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 294–301, New York, NY, USA. ACM Press.

Yu, B. and Singh, M. P. (2003). Detecting deception in reputation management. In *AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, pages 73–80, New York, NY, USA. ACM Press.

Yu, T. and Winslett, M. (2003). Policy migration for sensitive credentials in trust negotiation. In *WPES '03: Proceedings of the 2003 ACM workshop on Privacy in the electronic society*, pages 9–20, New York, NY, USA. ACM Press.

Yu, T., Winslett, M., and Seamons, K. E. (2001). Interoperable strategies in automated trust negotiation. In *CCS '01: Proceedings of the 8th ACM conference on Computer and Communications Security*, pages 146–155, New York, NY, USA. ACM Press.

Zhao, J., Wroe, C., Goble, C., Stevens, R., Quan, D., and Greenwood, M. (2004). Using semantic web technologies for representing e-science provenance. In *Proceedings of the 3rd International Semantic Web Conference*.

Zheng, J., Veinott, E., Bos, N., Olson, J. S., and Olson, G. M. (2002). Trust without touch: jumpstarting long-distance trust with initial social activities. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 141–146, New York, NY, USA. ACM Press.

Zhu, X. and Gauch, S. (2000). Incorporating quality metrics in centralized/distributed information retrieval on the world wide web. In *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 288–295, New York, NY, USA. ACM Press.

Ziegler, C.-N. (2004). Semantic Web recommender systems. In Lindner, W., Mesiti, M., Türker, C., Tzitzikas, Y., and Vakali, A., editors, *EDBT 2004 Workshops (PhD, DataX, PIM, P2P&DB, and ClustWeb)*, volume 3268 of *LNCS*, pages 78–89, Heraklion, Greece. Springer-Verlag.

Ziegler, C.-N. and Lausen, G. (2005). Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, 7(4-5):337–358.