"TRUST ON THE WEB: A MODEL FOR ONLINE DISCUSSIONS"

Ioannis Pagkalos
Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
ion@pagkalos.com

Mark J. Weal
School of Electronics and Computer Science
University of Southampton
mjw@ecs.soton.ac.uk

ABSTRACT

Trust can be a valuable ally when filtering information provided by the users of online forums or other tools of social interaction in the web, but it is also an implicit procedure performed by humans. This work studies the notion of Trust in online discussions and defines a generic model to describe, measure and evaluate it. An electronic survey was performed in order to understand how humans perceive Trust when interacting with a Forum. The combined results of this survey along with previous work in the area and initial observations provided a set of quantifiable Trust Metrics that was used as the basis for developing a generic model and a prototype web application (TrustBB) to describe and measure Trust. The intention is not to provide an all-in-one solution but to specify a framework of configurable metrics to describe Trust in forums and express it via an abstract rating. Administrators of TrustBB are then able to plug-in and edit their own metrics according to different research scenarios and requirements. The resulting Trust rating is useful to users/administrators of a bulletin board, who have a metric to consult when filtering information, but also to Social Scientists in their studies of the social interactions that take place in online discussions. Furthermore, not only is the model extensible and adaptable, but the combination of the model and an existing Forum technology could be used as a basis for further, more specialised experiments.

KEYWORDS

Trust, Trust Metric, Online Discussions, Forum

1. INTRODUCTION

Since its inception in 1989, the World Wide Web has changed our lives in ways that its designers and creators could never have predicted. The Web, with its highly decentralised architecture, has allowed for the creation and proliferation of new forms of social interaction among its users. Among those, online forums provide a mechanism for conversations and discussions on the Web. Web users regularly visit forums where they are faced with communicating with complete strangers, whatever the purpose of the visit may be. When attempting to filter such information, humans try to determine the trustworthiness of the source. These measures of Trust are implicit, and the patterns followed are part of human nature. Unfortunately, our understanding of many of its implications has not kept up with the Web's impressive growth. This work aims to help understand, study and model the implications of the Web in common social interactions and is inspired by the Web Science Research Initiative (WSRI[1]). More specifically, it studies the notion of Trust in online discussions and defines a generic model to describe, measure and evaluate it. In other words, it proposes a way to explicitly model the implicit concept of Trust in online discussions.

[1] http://www.webscience.org

Of specific interest to this work is how people use the Web daily. Kraut et al (2002) identify two separate perspectives: social and non-social use of the Web. Non social use of the Web includes solitary activities such as Web "surfing", news reading and single-player gaming. Social use on the other hand includes direct contact with other people – friends, acquaintances, or even total strangers (Weiser 2001). The opportunities that the Web provides for interpersonal communication over its infrastructure have led to a new global paradigm – Web Communication.

Online discussion is a relatively novel form of communication (compared to traditional means such as the telephone), facilitated by computer networks. The first such communication systems were built on mainframes in the early 1970s. By the mid 1980's, "BBSs" (dial-up Bulletin Board Systems) run by hobbyists on personal computers began to host online discussions as well. As networks became more sophisticated, and access to them became easier (through universities and dial-up modem access), networks like USENET hosted a substantial range of discussions (~650 million messages in 2003 (Ridings et al. 2004)). In the early 1990s, Internet Relay Chat (IRC), a chat-room style form of communication became (and remains) popular. With the advent of the Web, millions of users started using Web-based discussion boards (Forums). In the era of Web 2.0, these are joined by discussions on blogs (Web diary-style pages) and comment-enabled websites.

Virtual Communities (groups of people interacting using computer networks) soon formed around the first implementations of online discussion media. In every form of virtual community, whether they hold online discussions or not, there is an established group of leaders joined by numerous communicators (or "posters") which may log-in frequently or sparingly. The leader group (also called "moderators" or "administrators") is responsible for maintaining the community in terms of both technical and social balance. Posters who cause trouble are often banned –temporarily or permanently– from participating in the discussions. Virtual communities and the discussions that form around them are definitely one of the landmarks of the new Web era and portray the implications of the Web in the new form of social interaction.

For the purposes of this work, the chosen Online Discussion medium is the Forum.

2. TRUST ON THE WEB: A MODEL FOR ONLINE DISCUSSIONS

2.1 Trust & Trust Metrics

Trust, according to the Oxford English Dictionary, is defined as "confidence in or reliance on some quality or attribute of a person or thing, or the truth of a statement". Trust has existed as long as the history of human beings and the existence of human social interactions. Naturally, different scientific disciplines which have studied trust over the years give different definitions of trust. Furthermore, trust is often conceptualized by researchers according to the features of a particular context (Feng et al. 2004). This is to be expected as trust is primarily a subjective matter – it is directly related to and affected by individual differences as well as situational factors.

Different people view the role of trust differently in different scenarios and have different magnitudes of trust towards different trustees (Wang & Emurian 2005). It is, thus, very difficult to provide one single definition of human trust. On the other hand, the value of such an endeavour is debatable as providing an all-purpose definition to such a complicated notion is a great risk.

Computer scientists have studied Trust from a multitude of different viewpoints. Even though there's sufficient work in Machine trust, there is a lack of explicit models to describe Trust in Online Interactions and, generally, in cases where the "actor" is human. The importance of Trust in human interactions, other than being a natural component of human behaviour, has been proven in the research community by publishings as early as 1967 (Rotter). The importance of Trust has been identified however far earlier than this, even by the Ancient Greek Philosophers.

Of specific interest to this work is the concept of *Trust Metric*. A Trust Metric is a measure of how a member of a group is trusted by the other members. This notion is inter-disciplinary but, naturally, every discipline defines different metrics and different variables to affect them. There are two basic variations of Trust Metrics[2]: a Local trust metric predicts trust scores that are personalized from the point of view of every

[2] Trustlet WIKI, http://www.trustlet.org

single user. For example a local trust metric might predict "Alice should trust Bob as 0.9" and "Bob should trust Carol as 0.1". In direct contrast, a Global trust metric computes a single global trust value for every single user. For example, a global trust metric might predict that “Alice should be trusted {by everybody} as 0.9”.

2.2 Related Work

In Computer Science, researchers have applied the work of Philosophers, Sociologists, and Psychologists in cases where machine (computer) behaviour closely resembles that of human nature. Work in Trust related to this work can be found in the areas of Artificial Intelligence (Intelligent Agents), the Semantic Web, e-commerce as well as other areas of Computer Science.

Systems to measure and evaluate Trust in Agent interactions can be divided into two categories: *Centralised Systems* where a central authority is in charge of gathering trust and reputation information for all agents and *Decentralised Systems* where every agent carries out trust evaluations on its own (Huynh et al. 2006). Examples of such systems include SPORAS (Zacharia & Maes 2000), TRAVOS (Teacy et al. 2005) and FIRE (Huynh et al. 2006). In such systems Trust is measured by very specific and precise criteria – unfortunately many of the criteria chosen are either agent-specific and do not account for the improbabilities of human nature or are very machine-specific, using variables that depend on the way systems are built and the protocols that agents use to communicate between them. Nevertheless, some of the criteria, like Role-Based Trust, which is present in FIRE apply to the research scenario of this work and have been considered in the design process.

The Semantic Web (SW) is an evolving extension of the World Wide Web in which the semantics of information and services on the web is defined, making it possible for the Web to understand and satisfy the requests of people and machines to use the Web’s content (Berners-Lee et al. 2001). When deciding whether to trust a result provided by an unknown source, the SW client may want to ask a simple question: “How did you come to this conclusion” or “How did you get this data?” The reply to this question, a "proof", can be defined as a valid justification for a produced result. This justification can, for example, be described as a sequence of information manipulations used to generate an answer (Silva et al. 2006). Proofs can be a direct set of statements but may also include provenance information (information about how the proof was generated). Trust can then be established, for example, through proofs that are believed. This link between Proofs & Trust shows how the Semantic Web handles Trust – with a solid, grounded system of machine-understandable explanations. Unfortunately, even though this may be sufficient for agent-to-agent interactions, when human behaviour is factored in, “explanations” need to be very complex and are many times hard to translate into machine language.

In e-commerce, a well-known example of a website dealing with online auctions is eBay[3]. In such an environment, there must be mechanisms in place to ensure the trustworthiness of the individual as well as the transaction context in general. To accomplish this, eBay uses feedback ratings: After each transaction both buyer and seller can leave positive, neutral or negative feedback. The total value of these scores is presented as a measure of "trust". For example, a seller with 2000 ratings of which 99% are positive would be considered highly trustworthy while a seller with 10 ratings some of which are positive and some negative would be considered untrustworthy. EBay’s feedback rating is a simple yet immediately useful implementation of a global trust metric.

Even though there are many algorithms to compute Trust in these contexts, aggregating Trust in Forums is a difficult venture but also one that can be approached through careful work and consideration. There is the need for calculations to be as neutral, general and configurable as possible. Thus, the research question of this work is: ‘Trust is a major component of human interaction. Can we create a configurable tool to measure Trust in online discussions that a) automatically gathers and evaluates data from discussions and b) explicitly models the implicit way that humans perceive trust?’

In essence, the idea is inspired from eBay’s feedback value which was the first commercial visible Trust metric as well as existing trust models. By extracting information and variables from data already existing in a Forum, it becomes possible to define a range of metrics (classes of metrics) to describe Trust. This should

[3] http://www.ebay.com

not be a canonical, one-off "solution" to Trust computation, and thus these metrics must be completely configurable and the administrators of such a tool must have the ability to create and delete them at will.

2.3. Research Methodology

As seen already, Trust is being studied by numerous disciplines and work is progressing from different viewpoints. It is also clear that Trust is an important factor of every online interaction. The purpose of this work is two-fold. First, to design a Trust Model to describe Trust specifically in Online Discussions and second, to design a prototype tool that implements this model in order to test and evaluate it. The prototype tool can then be used to refine the model and vice-versa.

The research method used in this work was Prototyping and the research methodology that was followed is briefly described below (a detailed description can be found in (Pagkalos 2008)):

Phase A: A questionnaire was designed and distributed to forum users of a specific forum. The purpose of the questionnaire was to understand how humans perceive Trust as well as the interest in having a Trust Metric in a Forum. The questionnaire's results along with extensive literature review and initial observations provided a quantifiable metrics list, which can be extracted from board usage statistics and content analysis.

Phase B: A prototype tool was designed & developed to act as an interface over an existing forum technology. Codenamed "TrustBB" and built in PHP over a phpBB Forum (http://www.phpbb.net), this prototype is a configurable rule-based system that calculates an aggregated trust index on the Forum it is attached to. Administrators of TrustBB can create and edit their own rules to calculate Trust as well as group rules in rulesets to form different testing environments.

Phase C: The list of metrics designed in Phase A were ranked in order of importance by a Psychologist, who also provided qualitative feedback on the testing environments created thus far and provided suggestions for improvement of the Prototype. According to these, a specific ruleset was created for purposes of evaluation and its results were presented in the original group of users, along with a final questionnaire.

2.4. Metrics used

Explaining what Trust is is a difficult and risky endeavour. Trying to model Trust is equally, if not more difficult – care has to be taken to ensure that the resulting model is as abstract, general and configurable as possible, as Trust is a varied and ever-changing aspect of human behaviour. Also, it is not the job of a computer scientist to determine what constitutes the best way to compute Trust – this falls outside the discipline and is best left to experts of fields such as Sociology and Psychology. Computer science can and should be the discipline to create the tools and expose the data that will lead to context-dependant, configurable Trust aggregation.

During the course of this work, the combination of the initial questionnaire, literature review and initial observations resulted in 10 proposed metrics to describe Trust. These have been further divided into 5 groups according to their nature (as seen in *Table 1*). Those are Counters, Language Analysis, Behaviour Analysis, Miscellaneous, and, Plug-in-dependant metrics. This creates a varied range of metrics and sets the groundwork for the addition of more in the future.

Group	Group Name	Metrics
A	Counters	Number of Posts/Threads, Forum Age
B	Language Analysis	Keyword & Language Analysis, Writing Style
C	Behaviour	Forum Activity, Posting/Replying Habits, Success of Posted Threads, Significance in Forum
D	Miscellaneous	Role-based Trust
E	Plug-in-dependant metrics	"Thanks"

Table 1: Grouping of Metrics

Group A: Counters

1. Number of Posts/Threads: Perhaps the most known counter in Forums is the amount of posts a user has. This is usually readily available and serves as a quick indicator of the user's contribution to the community.

In our scenario, it is also useful to include the number of threads that the user has started as this can be used to create more specialized rules.
2. Forum Age (Member Since/Joined): Another one of a user's most visible and simple counter variables is the date he/she joined. This is also readily available and indicates how long the user has been a member of this community. This is something that can suggest how trustworthy a user is, as elder members of a community are traditionally more respected.

Group B: Language Analysis
3. Keywords & Language: This is potentially the most complex family of metrics. Depending on the context and the type of forum, keyword & language analysis can provide very interesting Trust calculations. For example, a user that constantly uses offensive or racist keywords should have a lower Trust rating than a person who speaks politely. As always, this is subject to the designer of the rules and should be as customisable as possible. The best results will come from a system that analyses not only keywords & vocabulary but also grammar and syntax.
4. Writing Style: A common tactic that is frowned upon in forums is commonly referred to as "ALL CAPS". It refers to the use of only capital letters and mimics the human behaviour of shouting, potentially irritating other forum users. Such writing styles, along with more complex ones ("Leet Speak", "Camel Writing") may infer lower Trust ratings to users.

Group C: Behaviour Analysis
5. Forum Activity: Forums typically keep track of when a user last logged in, how long his last forum session was etc. All this can be grouped together to create a profile of the user's Forum Activity. Users who log in frequently may appear more trustworthy, whereas users who log-in once a month usually lose focus of the forum's state.
6. Posting/Replying Habits: A Forum user can usually be grouped into two categories according to his posting habits. "Lurkers" are people who rarely start new threads and prefer to reply to threads created by other users. This may show a type of community fear and could affect a user's Trust rating. On the other hand, there are users who are most often the originators of discussion threads. This is an interesting variable to consider.
7. Success of Posted Threads: Another variable that can be used to infer Trust is the success of the threads started by the user. A potentially difficult variable to compute (as the success of a thread can be measured in a multitude of ways, always context-dependant), it is still a deciding factor when determining the contribution of a user to the forum. For example, a thread that spawns over 100 replies is often tagged in forums as a "hot thread" and typically contains useful information or a heated debate.
8. Significance in Forum: Concluding the behaviour analysis group, the significance metric attempts to calculate how important the user is to the community. For example, if a large percentage of the forum's threads are originated from this user, it is highly likely that he is an important figure in the community (a "contributor", usually a well known member of a forum community and typically rewarded with a special title beneath the username). This could also be computed from the contribution of the user in each specific thread (if a user contributes to a large percentage of the forum's threads).

Group D: Miscellaneous
9. Role-Based Trust: In a forum, there are distinct groups of leaders: administrators, usually in charge of both technical and social balance, and moderators, whose job is to keep the forum in a healthy state by removing irrelevant posts, banning users who cause problems etc. It is safe to assume that users who have been granted these privileges are more likely to be trusted. They are essentially the managers and staff of the forum and as human beings we've learned to assume a certain degree of trust on leadership figures (although philosophers and academics challenge this on a daily basis (O'Hara et al. 2004)). This is one of the bases of Trust and should be factored in on any potential Trust calculation. Role-based Trust is a notion that often appears in agent-to-agent interactions as well (Huynh et al. 2006).

Group E: Plug-in-dependant metrics
10. "Thanks": Over the course of Forums' history, there have been efforts to extend the basic capabilities by adding plug-ins (also called "mods"). One of the most often-used mods is the "Thank you" mod where people can say "thank you" to a poster by a single press of a button. This could be thought of as a Forum's

version of feedback, which is commonly encountered in most commercial websites at the moment. The existence of such a variable could very well be a basic indicator of trust in some forums.

One thing to note is that all of the above constitutes a framework of metrics – this is an attempt to define the variables that affect Trust in Forums and not the effect that each of them should have when calculating Trust. Computer scientists, and most importantly, psychologists & sociologists using such a system should be able to plug-in their own metrics & edit them according to their preferences and according to the current research scenario.

2.5. Implementation - TrustBB

During the course of this work, a PHP-based system was built in order to implement and test all the design guidelines proposed. Codenamed "TrustBB", it consists of a PHP-based website that acts as an interface to a phpBB forum. In addition, the website interacts with its own mySQL database to store and retrieve data. This Web application was built using common PHP editing tools and tested on a standard WAMP server (Windows Apache MySQL PHP).

TrustBB is a prototype implementation of the ideas presented thus far. Its purpose is to be a configurable rule-based system to calculate Trust in a phpBB Forum by using already existing or easily-computable data. Usually, most of the data required for the implementation of the metrics is available on the phpBB SQL database tables. A detailed demonstration of how the proposed metrics are applicable on a phpBB forum and how they were implemented in TrustBB can be found in (Pagkalos 2008).

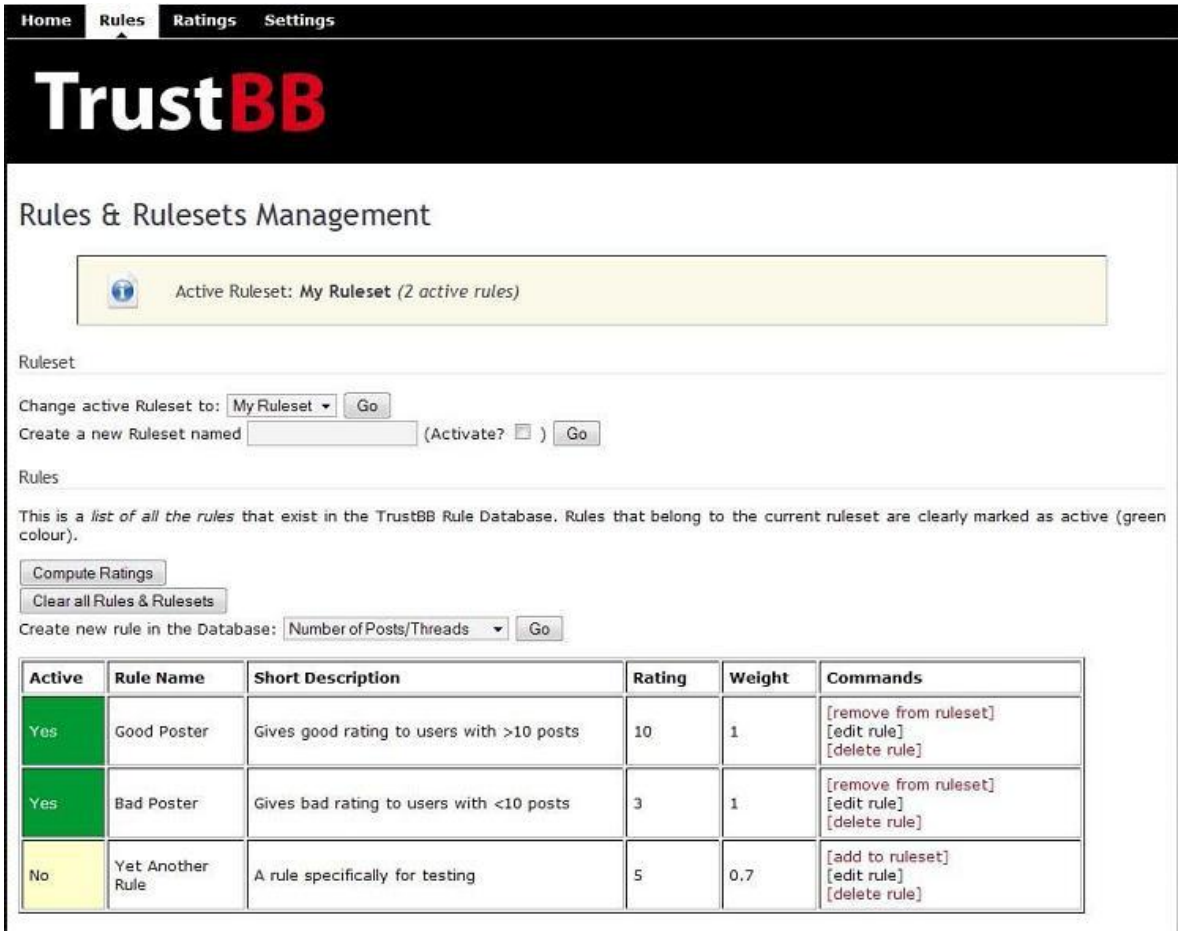

Figure 1: TrustBB Screenshot

From a TrustBB user's perspective, the process to create a simple testing environment is intentionally kept simple and straightforward. The TrustBB user (hereon called "Administrator") creates a new ruleset named appropriately for the current test scenario. He/she then proceeds to create rules that dictate how each of the metrics should affect each forum user's Trust rating. Finally, he/she adds the rules to the current active ruleset and proceeds to compute the ratings via the click of a button. The separation of rules into rulesets allows for switching between different testing scenarios on the fly and allows the potential for complex trust calculations where different rulesets are active on different timeframes.

2.6. Evaluation methodology

In order to obtain some qualitative feedback on both the Trust metrics described in this work as well as the prototype implementation, a short interview was conducted with an expert psychologist (Pagkalos 2008). The questions asked revolved mostly around the choice of metrics and their use in psychology, if any. The psychologist was then asked to evaluate the metrics independently and suggest her own. The psychologist confirmed the importance of Role-Based Trust as the most defining aspect of Trust among the list. She also identified Metric #8 (User's significance in the Forum) as a potentially basic aspect of Online Trust due to the power of leading figures in a community. It was argued that users who are significant in a community usually carry group leadership values as well, which in turn imply trust.

The users who participated in the initial questionnaire were subsequently asked to evaluate the results of a prototype TrustBB implementation running on the Forum. They were presented with a list of known users to which a Trust Rating was appended and asked to comment on the results. In addition, they were asked their opinion on whether the Trust Rating and the way it was calculated should be publicly available.

Overall, the comments on the resulting Trust Ratings were varied but close to their beliefs (though non-quantifiable). Regarding the two subsequent questions, 84% of the user base thought that the rating should be readily available but 75% agreed that the way it is calculated should be hidden.

It should be noted that this should not be treated as solid statistical results, as this was a prototype implementation. Nevertheless, it shows that the approach is promising and potentially fruitful.

3. CONCLUSIONS & FUTURE WORK

3.1. Conclusions

Trust is a major component of human interaction, be it online or not. The Web has created an environment where new ways of communication exist. In the end, these are not entirely novel as they are always based on human behaviour and are limited by the medium on which they are expressed. Nevertheless, the Forum is a solid ground for the development of fruitful online conversations and thus was an ideal candidate for such research.

Over the course of this work, various ways to describe, evaluate and compute Trust were described. Based on previous scientific work in the area and an initial survey performed on regular forum users, the range of metrics proposed creates a model that explicitly describes Trust. This initial grouping of 10 different metrics encompasses counter metrics, language/behaviour analysis as well as well-known and well-researched approaches to Trust such as Role-Based Trust. All metrics were intentionally kept abstract and generic in order to adapt to as many research and testing scenarios as possible, something which was later shown to be a correct choice by users.

The prototype tool developed (TrustBB) represents the first step in defining and implementing a model that could potentially help user communities everywhere. Interfacing over an existing phpBB Forum, TrustBB's rule-based approach to Trust aggregation allows administrators as well as researchers to easily create and maintain a Trust computing environment without changing anything on existing forums. Since this is a prototype stage, there is much work to be done, but early results show that there is also much promise.

3.1. Future Work

Research as early as Gambetta (1988) suggests that there are still many unidentified factors that play important roles in online trust. Future research needs to be conducted to explore those hidden factors and make trust models more complete. More basic metrics should be added to provide a larger, more adaptable framework. Age, for example, is one of the likely candidates to be included in the model. In addition, the value of metrics such Keywords & Language Analysis could be drastically increased by adding Natural Language Processing techniques. This could allow grammar and syntax to be taken into consideration along with keywords.

System-wise, the initial (bootstrapping) value of the system is "neutral". Researchers such as O'Hara (2004) suggest that there are multiple ways to initialise such a system and extensive testing should be performed to determine the best one for each research scenario. Completely new metric categories such as social network metrics could be added but its implications should be carefully studied and explored. Some forums already allow users to express friendship relations, and this opens up new possibilities. Metrics could also be combined to express even more complicated and varied situations.

By implementing Semantic Web technologies, the design can allow for agents (spiders) to take advantage of the results produced by TrustBB, and possible extensions include the ability for agents to request the input data only and do their own Trust calculations. Regardless of implementation, this provides the base for a Semantic Web reputation system where agents could look for a user's Trust ratings over the Web's many Forums. Most importantly, with the introduction of interdisciplinary work into such a project, the model could be used to create exciting new environments for research and experimenting, something which realises the Web's potential to its fullest.

REFERENCES

Berners-Lee, T., Hendler, J., and Lassila, O., 2001, The Semantic Web, *Scientific American*, vol. 284, pp. 28-37.

Feng, J., Lazar, J., and Preece, J., 2004, Empathic and predictable communication influences online interpersonal trust, *Behavior and Information Technology*, vol. 6, pp. 145-147.

Gambetta, D., 1998, *Trust: making and breaking cooperative relations*, B. Blackwell,

Huynh, T.D., Jennings, N.R., and Shadbolt, N.R., 2006, An integrated trust and reputation model for open multi-agent systems, *Autonomous Agents and Multi-Agent Systems*, vol. 13, pp. 119-154.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., Crawford, A, 2002, Internet Paradox Revisited, *Journal of Social Issues*, vol. 58, pp. 49-74.

O'Hara, K., Harith, A., Kalfoglou, Y., Shadbolt, N. 2004, Trust Strategies for the Semantic Web, *Proceedings of Workshop on Trust, Security, and Reputation on the Semantic Web, 3rd International Semantic Web Conference*

Pagkalos, I., 2008, Trust on the Web: A Model for Online Discussions, *MSc Thesis*, University of Southampton, UK

Ridings, C.M. and Gefen, D., 2004, Virtual Community Attraction: Why People Hang Out Online, *Journal of Computer-Mediated Communication*, vol. 10.

Silva, P.P.D., McGuinness, D.L., and Fikes, R., 2006, A proof markup language for semantic web services, *Inf. Syst.*, vol. 31, pp. 381-395.

Teacy, W.T.L., Patel, J., Jennings, N.R., Luck M., 2005, Coping with inaccurate reputation sources: experimental analysis of a probabilistic trust model, *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pp. 997-1004.

Wang, Y.D. and Emurian, H.H., 2005, An overview of online trust: Concepts, elements, and implications, *Computers in Human Behavior*, vol. 21, pp. 105-125.

Weiser, M. 2001, Whatever happened to the next-generation Internet?, *Communications of the ACM*, vol. 44, pp. 61-69.

Zacharia, G. and Maes, P., 2000, "Trust Management through Reputation Mechanisms," *Applied Artificial Intelligence*, vol. 14, pp. 881-907.