Amateur Fiction Online - The Web of Community Trust
A Case Study in Community Focused Design for the Semantic Web

K. Faith Lawrence, dr monica schraefel

Intelligence, Agents, Multimedia (IAM) Group
School of Electronics and Computer Science
University of Southampton

1 Introduction

This paper considers two of three related projects that are currently being undertaken as part of a larger Human-Computer Interaction investigation into whether the semantic web can be brought to hobbyist groups on the Internet. The user group chosen as a case study for this project was online amateur fiction community. This community was chosen because it could benefit from Semantic Services in the form of improved searching, improved meta data, automatic recommendation amalgamations, trust-webs and personalisation of such systems. Because this is a Human-Computer Interaction project evaluation will be through usability tests comparing the developed systems to those that are currently available.

This paper will first consider the work that has been done in the area of online communities and trust which relates to this project. It will then describe the way that user requirements were gathered from the user community through an online questionnaire as well as through direct interaction with volunteers from the community. Having defined the requirements of the community this paper will go on to describe two of the projects that are trying those needs. The first defines a foaf extension tailored to the needs of amateur writers and readers. This paves the way for the second project, adding a web of trust to the community and investigating whether this is a possible way of dealing with some of the community's access control needs. Finally I will consider what the next steps in this research are.

1.1 Related Work

Community Centred Design While the FicNet project touches on other domains it is at heart about human-computer interaction and accessibility. This accessibility is important both for the user-as-reader and for the user-as-creator since both are required for the successful implementation of the proposed system. Usability and interaction research is required in the following areas:

To determine the user needs and requirements for this project. The nature of the design is one of community focus and therefore the requirements of the community need to remain the primary focus [16, p.203-229, p.299-330]. This is being undertaken through continual interaction with the community and members who have expressed interest in the project and through questionnaires, and eventually user studies related to specific interaction questions.

To consider the way that the reader-facing applications present themselves and the community. Misunderstandings can cause problems between communities. If an application serves as a portal to the world of amateur writing it is important to recognise the responsibility that that entails. Part of that will involve making clear to the users what the system can and cannot do and where the points of failure may occur. Problems in this area can be seen in the systems currently in use. For example, the Google SafeSearch does not take into account any Platform for Internet Content Selection meta tags [4] that are attached to a website despite these being a World Wide Web Consortium Standard for marking Internet content since 1996 [14, 10, 7]. It could be argued that sites claiming to be child-friendly are not necessarily trustworthy in their assertions but the fact that SafeSurf ignores meta data added by sites with the express purpose of warning for adult content has caused problems between site owners and parents - the one thinking they have taken the necessary steps and the other unhappy with the site still appearing on the supposedly child safe setting. While annoying this is excusable since Google is not designing for a specific community where this is a known problem. As the designs that come out of this project are aimed at a specific user group it is good design to make such things clear so that the community being opened up does not suffer as a result.

Trust and Semantic Communities The semantic web "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries." [3] The concept behind the semantic web is to provide machine readable meta-data which can be processed by computers. It "is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." [3] Relationships between data can be explicitly stated or inferred allowing for the information to be integrated, correlated and collected. Trust exists on many levels within this environment do we trust the meta data attached to the document? Do we trust the data being returned and the mechanism that returns it? Do we trust the person requesting access to the document?

To consider these problems we have to first consider what is meant by 'trust'? In many respects that is a question for the philosophers. It is certainly one they have been discussing since the discipline first arose and reluctantly started taking notes [15, pp.30-36]. The dictionary definition is long and contains at least three totally distinct meanings depending on context. Academics who have investigated trust in social, philosophical and security domains frequently define the term to fit in with the point they are making - whether it is to do with fulfilling expectations or acting within a role [15].

Recent work has looked at trust in online social contexts. It is this domain of trust into which this research falls. The small world concept was opened up for study by Milgram [13] in the 1960s and recent work has shown that the Internet can be considered a small world system [1]. Research by Jenifer Golbeck has taken the small world concept and applied it to the electronic social networks. Golbeck created a trust ontology that could be integrated into the Friend of a Friend network [8]. This schema allowed people to assign a rating, either in general or within a specific area of expertise, to their friends. This rating equates to a value between one and nine with one being 'distrusts absolutely' and nine being 'trusts absolutely'. The trust network for the system can then

be mapped with the further possibility of calculating recommendations about the trust level between any two given people on the network. This system was tested with a mail filtering application and for film recommendation. We have extended this type of web of trust to create one which fits the user needs of the FicNet system. In this paper trust is used to refer to the expectations that arise that an individual will not act in a way that is detrimental to another individual or community.

2 Amateur Writing Online - A Case Study

The user group selected for this case study is that of online amateur writers and their readers. One of the most noticeable thing about the amateur writing community is that it is a community, or what has been defined as a series of overlapping communities [2, p.26-31,41]. This structure predates the Internet by many years, the electronic age facilitating communication between the geographically dispersed members by exchanging physical mail for its electronic counterpart. Having proposed that the amateur writing community as one which could benefit from additional semantic data and services a questionnaire was created to identify what the needs and preferences of both those within the the community were and those who were not part of it but came into contact with it. Because of the nature of the community and the broad range of people involved it was considered necessary to get the opinions of people outside the community who might have an interest in it. The most obvious group who fit this profile were the parents and guardians of the younger members of the community. The questionnaire sought to cover a number of areas from general information about the respondent to their reading habits, their social involvement within the community and their preferences and opinions of the access and distribution of different types of material within the community.

Having run a test period to confirm that the survey was in full working order it was advertised to the public starting on 7th December. This advertising took the form of requests to archives to pass the information on to their users, posts on Live Journal, e-mails to persons known to have an interest and post to the mailing lists and bulletin boards of related interest groups. Due to the number of responses the decision was made to harvest the first set of results after the survey had been running for two weeks. This first set of data comprised of 1118 responses of which two were found to be doubles. The remaining 1116 responses included 4 from prior to the 7th but otherwise spanned the 7th till noon on 21st December. Analysis of the IP address logged as part of the duplication identification process suggested that responses had come from over thirty countries. While America, Great Britain, Australia and Canada ranked the highest of the known contributing nations the international nature of the interest can be seen in the contributions from countries such as Finland, Russia, Brazil, Singapore, Japan, Estonia, Israel, India and Argentina as well as most of the nations in western Europe. Ages ranged from 4 people in the 10-12 years category to one person who was over 66 with 97% of respondents describing themselves as readers or writers or amateur fiction. Over 50% were in the 18 - 35 years range with 13.8% under 18. The information gained from the survey, which unfortunately there is no room to describe here, was used to inform the design of the ontologies and hypothesised applications described in the rest of this paper as well as the OntoMedia ontology described in [11].

3   Fan Online Persona

The Fan Online Persona (FOP) ontology is an extension of the FOAF ontology which has been trailered to the needs of online readers and writers. The extension was created for two main reason to clarify the options available and applicable to the needs of amateur writers and their related readers and to de-emphasis those which were not required. The following describes the steps that were taken to do this and why.

3.1   "I am perceived, therefore I am"

[12]

The first change was to subclass the `foaf:person` class to `fop:persona` and create the `fop:NomDe` class which allows users to designate a context with which a specific use-name should be associated. This is to create an environment totally separated from the off-line world. Studies of computer mediated environments such as UserNet or MMORPGs have shown that personae are created by users. Mackinnon [12, p.118] credits the creation of such persona on usenet to a combination of interference by the limitations of the written medium of the "user's personalities and unique qualities" and the control that the user has over the representation that they project.

The illusion of anonymity is a fundamental part of the fan fiction community and the fan fiction community makes up a large percentage of the amateur writing community. People may choose to give up that option but the option to keep "real life" and "fan life" separate is very important to those involved. Most authors write under a pseudonym, some under different names in different areas of interest. This is not just coyness, stories of people losing jobs, friends and family are common within the community. Even those who write under their own names do not tend to put their full name for this reason [p.200-202 9, 2, p.207/8]. The same applies, but even more so, to readers. Over 80% of respondents to the fan and amateur fiction survey gave 'valid e-mail address' as the most personal information that a reader or writer should be asked for even on an archive that contained adult material. That eighty percent included 15% who thought no personal information should be asked for from anyone and 28% percent who thought only authors should need to provide an email address. A significant percentage of respondents also mentioned the importance of privacy to them.

It has been suggested that "the more fundamental issue holding back widespread adoption of FOAF is privacy" [17]. The privacy issues are solely due to the open nature of the FOAF files. While users can decide how much information to reveal the possibility of other people revealing your personal details can lead to difficulties. There is no requirement for FOAF files to contain information related to a person's off-line details or even to refer to a "real" person. While the primary aim of FOAF is to create machine readable description of "people, the links between them and the things they create and do" it is clear from the available options that some combination of on and off line details is expected if not encouraged by the forms available for data entry. The creation of the fop:Persona construct is intended to reinforce the differentiation. It may seem that the low requirement for FOP file creation and the lack of strong link between off and on-line will make such files meaningless but evidence from studies of on line communities suggest that people become very attached to the persona they create and

4

do not often create new ones unless a specific reason arises: usually having gained a bad reputation. In this case a new persona will only be created if the disadvantages faced by the current identity are worse than those faced by a new one. The low level of risk that non-critical and non-commercial data are assigned makes this unlikely to be a large-scale issue. The most likely time when this would occur would be related to access control especially an underage person accessing adult material. In this case the person with the most vested interest in preventing this happening is the parent, who (having been alerted to the problem) will be able to monitor the creation of the new persona.

The expansion of `foaf:document` and `foaf:groups` allows FOP users to specify details of their own creations and make recommendations about others' work. The creation, exchange and review of works is the *raison d'etre* of the online amateur writing community. Where FOAF files consider work and school information, FOP files dismiss this information as potentially dangerous and irrelevant when compared to specifying which archives a persona's work is stored in.

The options presented in the FOP file were determined through a long term study of the meta data commonly attached to works when they were posted to mailing lists or distributed on websites backed up by the results from the survey.

3.2 Adding Trust

As described previously the trust ontology created by Jen Golbeck [8] allows one person to assign a trust rating to another. This data can used to create webs of trust that exist on top of the social networks that FOAF describes. A similar system can be seen as the basis of FilmTrust (http://trust.mindswap.org/FilmTrust/) a service created by Jen Golbeck. FilmTrust uses version of FOAF that has been extended to include a trust rating. The success of this application lays the ground work for further projects, like this one, which have a similar initial premise. The issue of trust is an important one in the amateur writing community. If we consider the relationship between the reader and the writer then we can see that a negotiation of trust occurs between the two (see Fig.1). The reader has to trust that the information that they are provided by the author or distributor falls within an acceptable level of accuracy. This problem is not unique to this domain and occurs whenever meta-data is created. In the amateur writing community it is considered more acceptable to be over-cautious than lenient. This is partly due to differing standards of what constitutes acceptable content.

When a story is over-rated then the reader may be disappointed but less trust is lost than when the story is under-rated. This can be seen in the recent discussion that occurred on the HelpingHands live journal [6]. The journal allows fan and other amateur writers to engage with parents over the technical problems and solutions related to preventing children stumbling across adult content. While it was not seen as a solution one of the steps taken to help parents was the creation of a list of sites suitable for children. The suggestion was tabled to remove one well known site from the list due to the level of distrust that existed about the self-assigned ratings attached to each story - a minority mis-labeling their work had a significant effect on the reputation of the archive. In this case the resulting compromise was that the archive remained within the list but a note was added to warn readers that meta-data attached to the stories might be misleading [5]
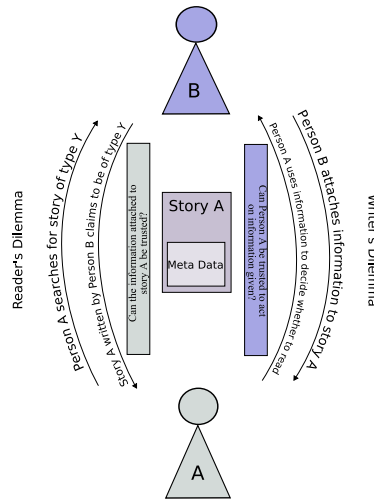
Fig. 1. The Perceived Trust Structure between Reader and Writer

thus propagating the message of distrust while allowing other readers the opportunity to judge for themselves.

The reader's dilemma, (at its heart just the standard problem of the accuracy of meta data), is less interesting in many respects than the writer's. For a long time the understanding that existed in the amateur writing community was that the writer would provide information to a known level of detail (which varied between communities), and the reader could choose whether or not to read the work. This negotiation assumes that the information provided is correct. Having made that assumption the decision to access the document is left wholly to the requester. The trust at this stage revolves around whether the provider can trust the requester to make the right decision. Since the person who has the most to lose is also the requester it would be a logical place for that decision to be made. However this has changed with the weight of perceived responsibility being moved away from the requester to the provider. The implicit assumption now being made is that the requester cannot be trusted.

To deal with this new mistrust inherent in the network the FOP files include the options necessary to model a web of trust. Rather then using the general structure seen in the Golbeck trust ontology the choice was made to create a specialised system. The largest contention faced by amateur writers is the aforementioned access of adult material by younger readers. This is a very debated issue since it is in effect preventing people doing on line what it is acceptable off-line. Other then obscene material the written word does not carry a formal rating system in most countries and children are allowed in all areas of libraries and book stores.

The initial design of the access trust web focused very heavily on age. It allowed one of two predefined assertions to be made that the persona represented someone over

the age of eighteen or that the persona represented someone under that age. It was then possibly for the persona to specify other personae that would vouch for them. It also provided a way for that persona to vouch for others. In much the same way that the general trust ontology allowed a gradient of trust and distrust, personae were allowed to qualify there support in a scale ranging from claiming off line knowledge supporting the statements through various levels of on-line knowledge to off line knowledge contradicting the statement. The webs of trust that can be constructed can then be used in a variety of ways to calculate the a trust rating for the requester which the the provider can use to decide whether to comply with the request. There were two immediate questions raised by this idea. Firstly was it possible to the add enough annoyances that creating clusters of fake persona to validate yourself was not worth the pay-off. Secondly whether people would make non-malicious statements of distrust and, if this proved to be the case, how the two distinct webs of trust and distrust could be integrated.

4 Future Work

The groundwork for FicNet has been laid insofar as the necessary ontologies have been developed to the point that they can be used as the basis for designing and building applications. While these will be refined as time and experience suggest the main focus of future work will be on designing the applications which will open the system up to users. The information already received from the user group will inform both the designs of the interfaces and the priority given to the different application. Smaller studies will be undertaken to confirm user preferences such as what types of statements should be required for the web of trust and how far the construction of these statements will effect both the perceived and actually accuracy of the statements. These trials will aim to gather the opinions of non-users such as parents of younger users as well the users themselves. The reason for this is that one of the advantages of such a system is to allow the parents and guardians of younger users more control over what they access and therefore the balance needs to be found between the needs of the users and the desires of the computer owners where these two do not coincide.

Work will also be done to integrate these systems with the OntoMedia ontology created to describe the content of works on FicNet and thus create a link between the people within the community and the work that they create.

5 Conclusion

From the survey we confirmed that the amateur writing community could benefit from semantic services and users would be interested in such a system if the benefits could be clearly seen to outweigh the disadvantages and it didn't cost anything. A wide variety of opinions were also collected which strongly suggested that personalisation and openness about the system were important requirements. Concerns about legal action and inappropriate access of adult works was balanced by concern about privacy and censorship. At this point we must conclude that the building blocks are in place but only a working model will let us see whether such a system could work in practice.

Bibliography

[1] Lada A. Adamic. The small world web. In *Proceedings of European Conference on Digital Libraries '99*, pages pp. 443–452, 1999.

[2] Camille Bacon-Smith. *Enterprising Women: Television Fandom and the Creation of Popular Myth*. University of Pennsylvania Press, Philadelphia, 1992.

[3] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific American*, May 2001.

[4] HelpingHands community members. Google's response. LiveJournal Community, April 2005.

[5] HelpingHands community members. Kid friendly fic sites. LiveJournal Community, April 2005.

[6] HelpingHands community members. A place to pitch in and help - a website creation resource and project. LiveJournal Community, 2005.

[7] Christopher Evans, Clive D.W. Feather, Alex Hopmann, Martin Presler-Marshall, and Paul Resnick. Picsrules 1.1. W3c recommendation, World Wide Web Consortium, December 1997.

[8] Jennifer Golbeck, Bijan Parsia, and James Hendler. Trust networks on the semantic web. In *Proceedings of Cooperative Intelligent Agents 2003*, 2003.

[9] Henry Jenkins. *Textual Poachers: Television Fans and Participatory Culture*. Routledge, New York and London, 1992.

[10] Tim Krauskopf, Jim Miller, Paul Resnick, and Win Treese. Pics label distribution label syntax and communication protocols. W3c recommendation, World Wide Web Consortium, October 1996.

[11] Katharine Faith Lawrence. Ontomedia - creating an ontology for marking up the contents of fiction and other media. June 2005.

[12] Richard C. Mackinnon. Searching for the leviathon in useNet. In Steven G. Jones, editor, *CyberSociety - Computer-Mediated Communication and Community*, pages 112–137, London, 1995. Sage Publications.

[13] Stanley Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[14] Jim Miller, Paul Resnick, and David Singer. Rating services and rating systems (and their machine readable descriptions). W3c recommendation, World Wide Web Consortium, October 1996.

[15] Kieron O'Hara. *Trust, from Socrates to Spin*. Icon Books, 2004.

[16] Jenny Preece. *Online Communities - Designing Usability, Supporting Sociability*. John Wiley and Sons, Ltd., 2000.

[17] Joseph Smarr. Technical and privacy challenges for integrating foaf into existing applications. Presented at 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web, September 2004.