WEB BASED SEMANTIC COMMUNITIES – WHO, HOW AND WHY WE MIGHT WANT THEM IN THE FIRST PLACE

K. Faith Lawrence
University of Southampton
Highfield, Southampton, UK
kf03r@ecs.soton.ac.uk

m. c. schraefel
University of Southampton
Highfield, Southampton, UK
mc@ecs.soton.ac.uk

ABSTRACT

This paper describes an investigation undertaken as part of the FicNet Human-Computer Interaction project into the online amateur fiction community. By working with the community to determine current practices and areas of concern we consider how future technologies such as the semantic web might be used to design applications to support the community. As a first step in this process we gathered opinions both from members of the community and from those outside the community who had come into contact with it. Taking this information we consider the community as it is and what it might become.

KEYWORDS

Community, HCI, Collaborative Standards, Semantic Web,

1. INTRODUCTION

FicNet is a human-computer interaction project focusing on identifying the needs of the amateur online writing community. Issues facing this community come both from outside the community and within. External complaints, witchhunts and legal threats can force sudden relocation or removal of content while lack of resources or loss of interest can disable sites permanently. All of which add additional complications to an already diverse system of differing standards and expectations and vocabularies.

By definition the online amateur writing community exists around the creation of media, mostly text but also illustrations and audiovisual items. Taken altogether amateur fiction probably represents one of the largest online electronic libraries currently in existence. Unfortunately, it is one which doesn't come with a catalogue and is constantly in flux. While the larger archives are fairly fixed in their position, the smaller archives and personal pages are frequently changing address, going down temporarily for maintenance or due to bandwidth limitations or just vanishing. "Can anyone tell me where to find..?" is a frequent question on many lists as is the popular "Can anyone recommend..?" or "I am trying to find a story that contains...". A large part of this ever changing nature is attributable to the subject matter and the very amateur nature of the enterprise. What keeps it all together is the community that surrounds it and the interaction within it.

In this paper we consider the user needs and requirements of the community and how semantic services could integrate and improve on the current architecture. Direct interaction with members of the community through observation, questionnaire and interviews was used to gain understanding of the specific difficulties and issues that community members face. Following initial discussions with community members as part of a preliminary assessment, a questionnaire was used to gain greater insight into the issues that had been highlighted. This method was chosen for the initial data gathering because it allowed us to involve a larger

section of the community than would have been possible through alternative methods. The results of this study are detailed below as we consider how future technology might affect the community and therefore how this and similar communities might affect future technology development.

2. THE AMATEUR WRITING COMMUNITY

The amateur writing community is made up of two mostly separate groups – media inspired or fan authors and 'original' authors. Of these two, the former are more vocal online because, through necessity, they have eschewed traditional publishing in its official form and thus rely on community published works, or zines, and individual dissemination. While paper based zines still continue to be produced as a means of distribution the ease of electronic publication has lead to a massive migration to the Internet. Equally most fan authors are hobbyist orientated, writing for fun and because of their interest in the source material. They, therefore, are more motivated to freely share their works since that in itself is their main goal. Some 'original' amateur authors have also taken advantage of this alternative means of publication with new sites such as LuLu offering "free" print-on-demand services. However for many others their presence within online writing groups is about practicing with the goal of writing professionally. For this reason the FicNet project is mostly concerned with media-inspired authors and their creations although it deliberately tries to avoid excluding original-world authors not least because of the crossover between the two (see Figure 1).

Fan and Original Author Representation
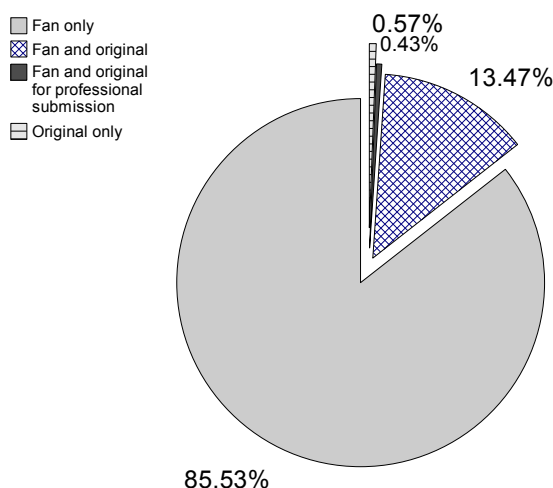(of 698 respondents who specified in Q3.10 of user survey)

Figure 1: Division Of Author Types Seen In Our Survey

To understand the media-inspired side of amateur writing, or fan fiction, it is necessary to take it in context with the rest of online fandom. It is not about lacking the originality to create new characters and settings but about exploring existing and loved ones. While the stereotypical 'fan' is often depicted as a white, male virgin this image has been frequently contested and discredited. This is particularly true for media fans [10, 2], of which fan writers are a part, and media fans online [6, P.134]. The social aspect of these communities have been documented in a number of pre- and early Internet studies for example [10, 2] and in online culture and computer-mediated communication studies looking at online communities [3, 4, 6, 7, 15].

What was clear in the studies that have been done is that online fandom, especially the fan fiction component, has a very strong female presence. Costello goes as far as saying that "the general Internet sample from the most recent GVU survey [1998] is virtually a mirror image of the cyber-fan sample" [6, P.134-5]. This is an interesting contrast to the experience with online fandom detailed by Janis Cortese in [14] where the online fan presence was strongly male dominated and unreceptive to female appreciation of the male members of the Star Trek crew. Given the other findings in this area it might be suggested that rather then demonstrating the typical and unequivocal male bias and double standards as suggested by [14] what this in fact shows was that Cortese was unlucky in her choice of discussion group. While perhaps less obvious there were many communities in existence at that point on the Internet where Cortese' comments would have been welcomed and where she would have found herself among the gender majority.

The problem here, other than the lack of civility shown by the group she joined, was that as a newcomer Cortese had few clues as to how to navigate the disparate sprawl of related groups and find ones appropriate for her specific interests. Whether this interest would have included the large amount of fan fiction undoubtedly existing and featuring those same crew members about whom she posted we have no way of knowing but the principle and the problem of finding the right group for any given user is the same.

The FOAF ontology (http://xmlns.com/foaf/0.1/) which describes people and their relationships is one of the most populated and popular ontologies [8] although at least part of this can be put down to the decision by LiveJournal to automatically generate FOAF files for its many users. While not specifically geared towards communities it does provide a simple way for people with like interests to find each other within the vastness of the Internet. Social networking systems such as Friendster and Orkut are in many ways one of the two main areas within which the semantic web can be seen as having some success (the other being academic settings, especially science and computer science). However while these interest-connected networks *may* describe a community, common definitions [1, P.10][17] require more self-identification as a group, at the least, for it to be counted as a community.

Analysis of the online amateur writing community (see below) suggests that they would benefit significantly from the advantages that computer readable metadata and the processing of such and the semantic web offers one method to provide this capability. The advantage of the semantic web as a method is that it can be argued that there are enough similarities in terms of structure and social dynamics to those groups for whom semantic tools and services have already been created to allow us to build on this earlier work. In doing so online amateur writing community provides access to a demographic that is often overlooked in computer science, the mature female user, as well as providing insight into what would be required to make an online community into a semantic one.

2.1 Gathering Community Requirements

2.1.1 The Questionaire

The decision was taken to create an online questionnaire. The reason this method of user response was chosen is because it allowed us to retain the most control of the way that the questions were presented and the ways in which they could be answered [1]. The survey was split into seven parts with some sections aimed at specific participants (see Table 1). An eighth section was added shortly after the questionnaire was made public to allow anonymous feedback and comments since while contact details were made clear they were via e-mail and therefore not as anonymous as the questionnaire. The draft version of the questionnaire was shown to the members of the Fan Fiction Ontology Yahoo Group. At that time the group had just over fifty members drawn mainly from the fan writing community. Changes were made to the questionnaire based on feedback received from those members.

Invitations to participate took the form of posts on LiveJournal, e-mails to persons known to have an interest and post to the mailing lists and bulletin boards of related interest groups. A number of archives and fan sites mostly related to Harry Potter were also contacted with a request that they publicise the survey to their users. Since those contacted directly were mostly adults or adult dominated groups the general Potter archives and websites were contacted in the hope of getting more younger fans. This last received a limited response but at least two archives did post notices. Notices were also posted in a number of places both around the university campus and at other locales where it was thought they might be seen. This was done to solicit opinions from those who were not directly involved in the amateur writing community but had some contact with it via friends or relations and therefore wanted to express their views as a outside observer.

2.1.2 The Response

The questionnaire was put online on December 3rd 2004 and a selected group of volunteers including adults and children were contacted directly to test the system. These tests were done over the following few days and when no problems were found with the technical aspects the questionnaire went officially live on December 7th. The response to the questionnaire was greater than anticipated. When it quickly became clear that over a thousand responses might be reached the decision was made to harvest the first wave of responses

after the questionnaire had been publicised for two week. This first set of data comprised of 1118 responses of which one was known to have been superseded.

The IP address of each response were logged and analysed in conjunction with the answers given and this revealed one duplicate entry beyond the one that was known. The remaining 1116 responses included 4 from prior to the 7th but otherwise spanned the 7th till noon on 21st December 2004. 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.

Table 1: Sections of the Community Requirements Questionnaire

No.	Section Description	Respondents
1	Questions related the respondent's familiarity with amateur fiction online, their age and the age of adulthood in their region	Everyone
2	Questions related to preferences for story access and display and level of community involvement	Readers
2b	Questions related to access controls	Underage Readers
3	Questions related to practice regarding story access and display and publishing methods	Writers
4	Questions related to how people outside the community came into contact with amateur writing online	Interested Others
5	Questions related to specific sub-types of amateur fiction and opinions on the access of "adult" content	Everyone
6	Questions related to technical knowledge and ability	Everyone
7	Questions relating to blocking and filtering amateur fiction and the collection of personal information	Everyone

As well as being international the age range of respondents (see Figure 2) went from 10-12 years to over 66, while over 50% fell within the 18-35 group and more than 20% were over 35. This is comparable to that age range found in online fandom in general by Costello [6] although he does not include any under 18s in his data.

2.1.3 Bias

Questionnaires are by their very nature biased because as a surveyor you only hear from those who wish to respond. Despite the range of answers we received they represent a very small fraction of the amateur writing community. As well as the self-selecting nature of the survey some of the bias inherent in this study can be seen as coming directly from the way it was marketed. While this method of viral marketing was a success in that it allowed word of the questionnaire to reach a large number of people but the method of transmission also heavily favoured people
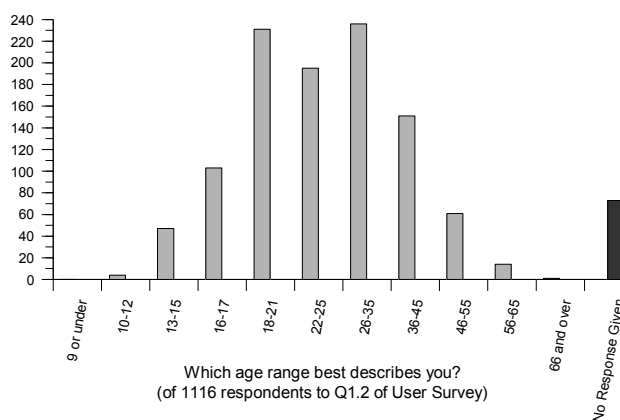
Figure 2: The Age Range of Questionnaire Respondents

who had a community involvement. Since media-inspired writers are more connected to the larger interlocking community structure than other amateur writers there is a immediate bias towards fans and away from original writing groups. Many groups are wary about catering to younger writers and therefore it has been suggested that this as well as other reasons means that they are less involved in the community aspect of online amateur writing and thus are less likely to hear about, and due to Internet safety lectures, respond to a questionnaire request.

Questions on gender were not included as part of questionnaire due to sensitivity within the community about revealing personal data. From Media Studies and Popular Culture research we know that there is a strong female bias among fan writers [10, P.48, 106-116][2]. Costello [6] showed that female fans were more likely to be involved as active social participants including information exchange and fan fiction. As we have already shown there is a large bias towards media-inspired rather then original writers it was assumed before the questionnaire was released that the majority of those who responded would be female. While there is no way to prove this supposition the tone of many of the responses suggested it was in fact correct. Not least because there was also a noticeable preference for "slash"[1]. It is not clear whether this was due to a genuine bias in the surveyed population, a reflection of the online reality or whether this group is more self-identifying since the information was given without being specifically asked for. Since slash stories are most often singled out for mention as transgressive or dangerous the size of the response from this section of the amateur writing community was not considered problematic.

2.2 Analysing the Responses

2.2.1 Requirements

It is not possible to detail all of the results of the survey here and the conclusions that were drawn from it. However I will touch on a number of points that became clear and the areas that they opened up for development. Anonymity and privacy were highlighted as areas of concern. The illusion of Anonymity is a fundamental part of the fan fiction community and as previously mentioned 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 [10, p.200-202][2, p.207/8].

The same applies, but even more so, to readers. While some details are expected from writers so that feedback can be sent to them readers see no need why their identities should be required of them. Over eighty percent 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 fifteen percent who thought no personal information should be asked for from anyone and twenty eight 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.

The problem of privacy is tightly bound with that of identification. If a person is not identifiable then how can one regulate access if required. The increasing use of LiveJournal, JournalFen and other similar electronic journaling systems with their friends and filtering option has created a new layer of options for community members. Standard practice on LiveJournal requires entries deemed 'not safe for work' to be placed behind cuts. A survey of nineteen community members carried out as part of a task analysis diary study revealed that approximately forty percent of the volunteers accessed community resources from work with reading journals being mentioned by the majority of that subgroup. However while some filtering occurs regularly on personal journals which are used for both community and non-community posts it is practically unknown on those journals, community or personal, that are reserved for community interaction. Those that do not wish to restrict adult content within journal communities may require some indication that the requester is an adult but in reality few checks are made, and the benefit of the doubt is given by default. This is one area where conflict can arise between the standards accepted by and within the community and those which external forces wish to impose.

---

1 Stories featuring a same-sex relationship often as a main aspect of the story. The majority of much stories are involve male-male pairings although female-female stories sometimes called femmeslash, femslash or saffic are becoming more popular. While not exclusively so, the vast majority of the readers and writers of such stories are female.

Recently there has been a lot of publicity given to concerns, especially parental, about the accessibility of adult content on the Internet. Nearly two thirds of those who answered the question on the content of what they had written said that they included adult content (by which we mean violence and other similar themes as well as sexual content), almost a quarter occasionally wrote stories they would rate as 'R' and over 80% read them. Given this, some addressing of the issue of adult content is necessary. While some concern is reflected from the community the unease is mostly related to younger children being exposed, with nearly a third expressing this in conjunction with the belief that it is either a positive way for teenagers to explore adult issues or unlikely to contain anything they are not already aware of. Almost another quarter expressed concern that parents allowed children who could not be trusted to respond to warnings appropriately to surf unsupervised.

This suggests that while the community believes in taking some steps with over half agreeing additional precautions should be taken with regard to access for story containing adult content they also believe that a large amount of responsibility rests with the reader, and where the reader is a minor with the reader's parents. When asked what steps would be considered reasonable precaution metadata attached to the story did not rate nearly as highly as human-readable warnings however this might be related to the generally low knowledge of them (not knowing what ICRA/PICS or equivalent tags were being the most frequently given reason for not including them) despite them being the second most commonly used method after warnings and the low incidence of filter use within the community. There was also some concern expressed about the level of detail on such metadata systems, possibility of stigma and misuse of the filtering system.

Previous initiatives to have community members add some form of increased access control to sites have almost always ended in bitter disputes. While in-community efforts have been received with more welcome than those from external sources anything that suggests possible control or censorship is met with profound suspicion. Misunderstandings at profound levels aggravated by indefinite terminology and the existing debate over the depiction of contentious issues such as underage or extreme relationships, violence, drug use, sexuality and religion have made it difficult to prevent the debate turning into the exchange of immovable extremes. Such issues would need to be addressed before any semantic web system could be given widespread acceptance by the community. However given the willingness to add human-readable metadata it does not seem beyond the bounds of possibility that machine readable data could also be added if it were presented in the right way. While such information can already be added, and in many cases is, lack of information and understanding of the technology on both sides of the debate confuses the issue.

Any applications need to support 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 [5] that are attached to a website despite these being a World Wide Web Consortium Standard for marking Internet content since 1996 [9, 12, 13]. 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.

From within the community the feeling is strongly towards that of the informed reader making a choice. However the question of what metadata should be available to readers resulted in a very mixed response with some people wanting to know everything before making a decision on whether to read and others only wanting to know the basic bibliographic details (see Figure 3). While there was a match between the metadata desired and the metadata given on the basic details the gap between the two grew on more contentious issues (see Figure 4). From this we can conclude that personalisation would be a very useful feature and, further, allowing people to tailor their options might also help with some of the problems of access noted above.

The most common reason given by authors for not wanting to provide information (other than time constraints) is that of spoiling the plot. This, unsurprisingly, is the same reason given by readers for not wanting to know. This is where the difference between human-readable and machine-readable information might be most useful to the community. One of the respondents to the community requirements survey noted

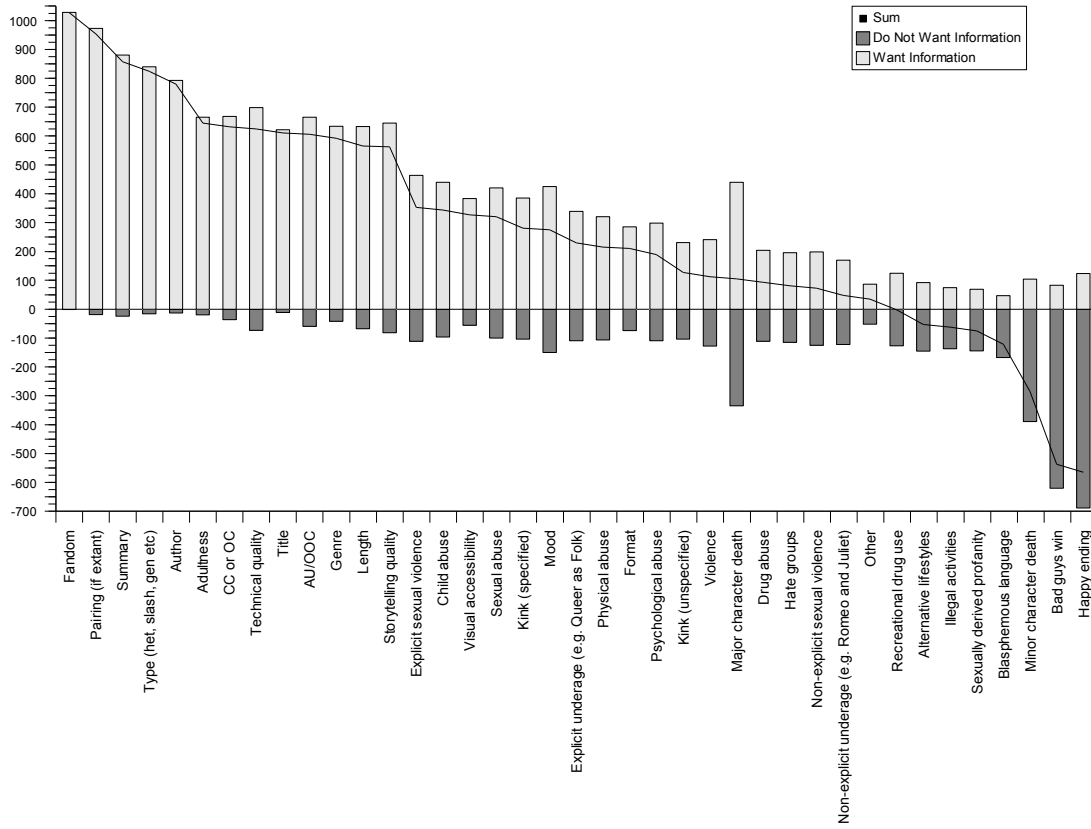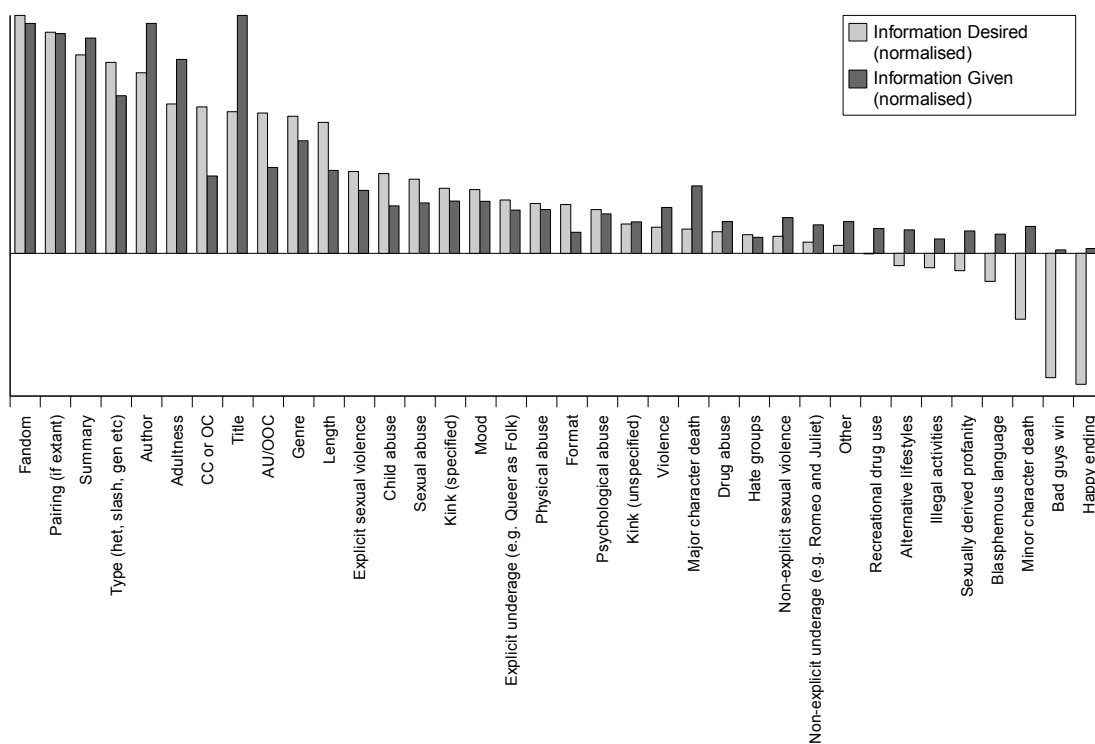*Figure 5: Preferences for Availability of Descriptive Metadata*

Figure 6: Normalised Comparison of Sum of Information Desired and Information Provided

that they wanted to be warned for subjects that they didn't like but wanted to be surprised by those that they did. By allowing for metadata to be used for searching and filtering at a machine level it is possible to provide community members with this capability. For the most part authors want their work to be consumed by the audience for whom it was intended so might be more amenable to adding additional data if they could also specify that it could be hidden. Equally for topics such as character death where there is no clear consensus individuals can decide for themselves whether that information is given to them. As more detailed information becomes the norm then members will even be able to search for or avoid the deaths of particular beloved or hated characters.

3. CONCLUSION

The semantic web offers many opportunities that are not otherwise available within this context because, while the community already generates a large amount of human-readable information, it does not also associate machine-readable metadata. This would allow better integration of the distributed systems, improved searching and filtering and more personalised services. These could benefit the experienced user by expending their options and creating new ways with which they could interact with the community as well as aiding the newcomer by easing their introduction into a community which has its own expectations, unwritten rules and obscure terminology.

In this paper we have detailed the methods through which we have extracted user requirements and the main concerns that were raised. The amateur online writing community has many issues which could benefit from the application of semantic services. It represents a large, diverse user group and, further, one which is often overlooked in computing studies. By understanding the issues that the community brings to the semantic web, as well as those that the semantic web brings to the community, we can improve the design of services and identify and deal with problems that would otherwise adversely affect use and the acceptance of this new technology. Many of the problems that online amateur writers face are those faced by all online communities. By working with the community to find possible solutions on the micro scale we take a step, however small, to solving some of these problems on the marco scale. Unfortunately many of those problems have their roots in the offline world and until those are resolved, a process that thousands of years of human history has yet to manage, the best that can be hoped for is a temporary compromise.

REFERENCES

[1] D. Andrews, B. Nonnecke, and J. Preece. Conducting research on the internet: Online survey design, development and implementation guidelines. International Journal of Human-Computer Interaction, 16(2):p. 185–210, 2003.

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

[3] Nancy K Baym. CyberSociety - Computer Mediated Communication and Community, chapter The Emergence of Community in Computer-Mediated Communication. Sage Publications Ltd., London, 1995.

[4] Nancy K Baym. Tune In, Log on: Soaps, Fandom and On-line Community. Sage Publications Ltd, 2000.

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

[6] Victor J Costello. Interactivity and the 'Cyber-Fan': An Exploration of Audience Involvement within the Electronic Fan Culture of the INternet. PhD thesis, The University of Tennessee, August 1999.

[7] Donna Cromer. Scifi tv fandom: The influence of the internet. In Leslie Fife, editor, 2000-2002 Southwest/Texas Popular Culture Association/American Culture Association Conference Proceedings, pages P. 317 – 331, 2002.

[8] Li Ding, Lina Zhou, Tim Finin, and Anupam Joshi. How the semantic web is being used:an analysis of foaf. In Proceedings of the 38th International Conference on System Sciences, January 2005.

[9] 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.

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

[11] Barbara Kantrowitz. Men, women and computers. Newsweek, page 48, 16 May 1994.

[12] 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.

[13] 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.

[14] Dale Spender. Nattering on the Net - Women, Power and Cyberspace. Spinifex Press Ltd, North Melbourne, Australia, 1995.

[15] Nessim Watson. Virtual Culture: Identity and Communication in Cybersociety, chapter Why We Argue About Virtual Community: A Case Study of the Phish.Net Fan Community. Sage Publications Ltd., London, 1997.

[16] Steve Whittaker, Ellen Isaacs and Vicky O'Day. Widening the Net: Workshop report on the theory and practice of physical and network communities. SIGCHI Bulletin, 29(3), July 1997.