

User Requirements and the Semantic Fan

K. Faith Lawrence, m.c. schraefel

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

Abstract. This paper describes an investigation undertaken as part of the FicNet Project to determine the user requirements for a semantic web system for online amateur writers and their readers. The overwhelming response from an wide-ranging demographic highlights some of the difficulties in creating a system that will be easily usable by all.

1 Introduction

1.1 FicNet

FicNet is a human-computer interaction project focusing on identifying the needs of the amateur online writing community. Taken altogether amateur fiction probably represents one of the largest online electronic libraries currently in existence. Unfortunately, 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. Complaints, witch-hunts and legal threats can force relocation while lack of resources or loss of interest can remove sites entirely. All of which add additional complications to an already diverse system of differing standards and expectations and vocabularies. 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 expanding 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 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 an preliminary assessment a questionnaire was undertaken to gain greater insight into the issues that could be addressed through technological support. This method was chosen because it allowed us to involve a larger section of the community than we would have been possible through alternative methods. It also laid the groundwork for later more focussed studies.

1.2 The Amateur Writing Online 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 used as a means of distribution the ease of electronic publication to share works 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 up this alternative means of publication but for many others their presence online and in online writing groups is about practising on the way to 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)

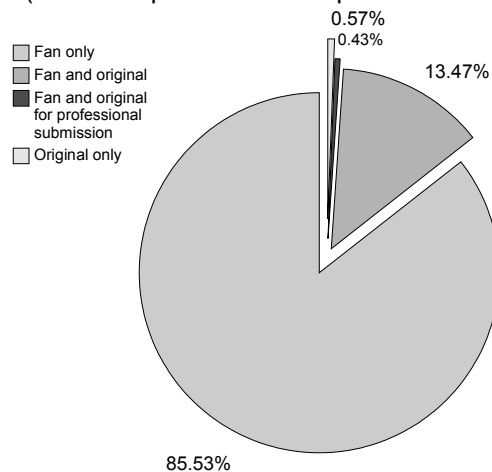


Fig. 1. Division Of Author Types Seen In Our Survey³

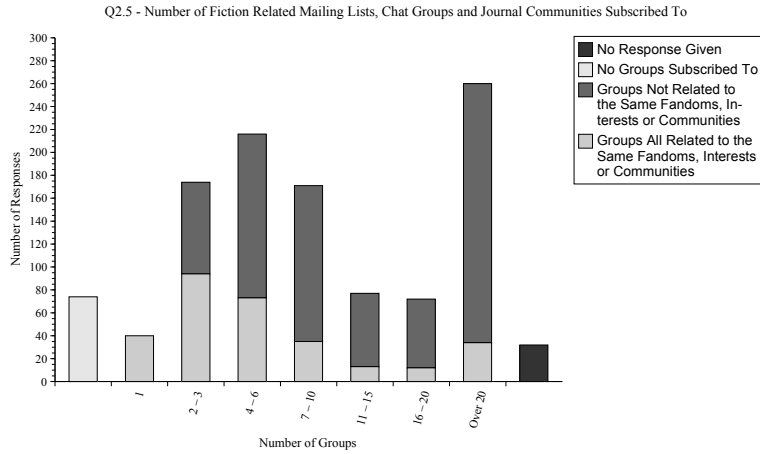


Fig. 2. Community Involvement of Respondents

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 (see Figure 3). 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 [13, 2] and media fans online [6, P.134]. The social aspect of this community has been documented in a number of pre and early Internet studies for example [13, 2] and in online culture and computer-mediated communication studies looking at online communities [7, 3, 4, 20, 6]. Beyond this, reputation is an important factor in the fan community hierarchy and is gained through interaction with other members of the community most specifically through the exchange of creations, information and opinions. In this way it is very similar to academia in that status is gained through discourse and publication.

What was clear 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 GYU 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 [15] 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 than demonstrating the typical and unequivocal male bias and double standards as suggested by [19] 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

³ see section 2.3 for discussion of bias.

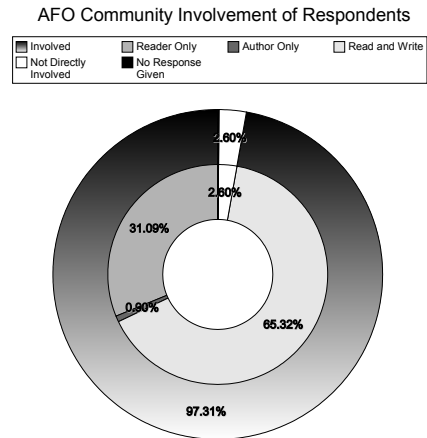


Fig. 3. The Community Involvement of Respondents

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 reason that this group was selected for consideration is because in many ways they are ideal for the purpose. The two areas in which the semantic web has had some successes is within academic settings, especially science and computer science, and outside academia with social networking systems such as Friendster and orkut. The FOAF ontology¹ 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. The online amateur writing community mirrors or includes many of the attributes and structures of both of these groups. While having enough similarities in terms of structure and social dynamics to those groups for whom semantic tools and services have already been created allowing us to build on this earlier work the online amateur writing community provides access to a demographic that is often overlooked in computer science, the mature female user, and to a widespread and populous section of the Internet community.

¹ <http://xmlns.com/foaf/0.1/>

1.3 Where The Semantic Web Could Help

The amateur writing community, both fan and original components, is very distributed and very diverse. While some central archives exist such as FanFiction.net and FictionPress.net (FanFiction.net's sister archive for original stories) many other stories exist in smaller domain specific websites, on personal homepages and in electronic journals and mailing list archives. An idea of the scope under discussion can be gathered from the numbers from six of the larger archives randomly chosen via Google search.

Table 1. Selection of Multi-Domain Amateur Writing Archives²

Archive	Authors	Stories (Poetry)
FanFiction.net		1, 087, 412
FictionPress	125, 206	214, 536 (489, 620)
AdultFanfiction.net	35, 871	41, 994
Freedom Of Speech	2, 337	4, 900
Slashfanfiction.net	1, 504	2, 652
WWOMB	1, 431	6, 717

Any design decisions should 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 [18, 16, 9]. 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.

2 Gathering User Requirements

2.1 The Questionnaire

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.

² Figures obtained from respective archives on 13/06/05

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	Under-age 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

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 Live Journal, 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.

2.2 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.

As well as being international the age range of respondents (see Figure 4) 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 [6] although he does not include any under 18s in his data.

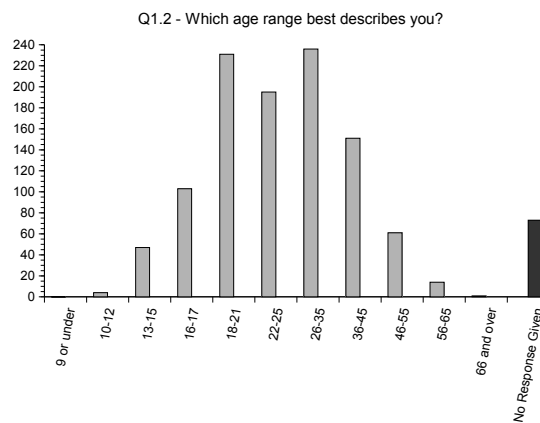


Fig. 4. The Age Range of Respondents

2.3 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 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 [13, P.48, 106-116][2]. [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 than 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”². 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.

3 Analysing the Responses

3.1 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 [p.200-202 13, 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

² 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 femslash or femslash are becoming more popular. While not exclusively so, the vast majority of the readers and writers of such stories are female.

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.

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 address 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. Such issues would need to be addressed before any semantic web system could be given widespread acceptance but 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.

The amateur writing community has the advantage that it already spends a significant amount of time associating metadata with the media items that they create. (see Figure 5). This is perhaps aided by the strong overlap between the writers who add the metadata and the readers who make use of it (see Figure 3) thus reducing the 'markup prisoners dilemma'. It is encouraging that the majority of those who responded to the question would be willing to consider spending slightly longer if it would raise the visibility of their work among its target audience with a strong interest also being shown in metadata re-usability (see Figure 6).

While Internet Explorer held a dominant position a broad range of browsers were used with Mozilla/Firefox, Netscape, Opera and Safari also being popular. Equally while Windows was the most popular operating system, Macs, Linex, handheld computers as well other operating system were also used suggesting that cross-platform and well as cross-browser compatibility would be a useful feature. Just over half of respondents gave their level of technical ability as comfortable with approximately 20%

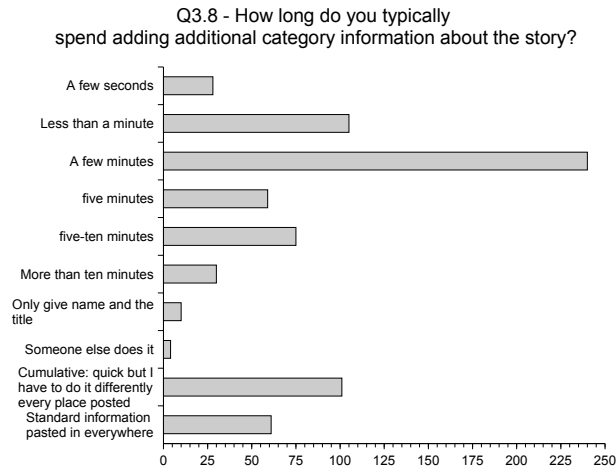


Fig. 5. Time Taken to Add Metadata

being happy making webpages although they did not regularly deal with any underlying code. Beyond that approximately 30% worked with HTML directly although they did not know how to program or use scripting languages. The remaining 15% were programmers with just over 5% web-scripters and nearly 5% programming professionally. The percentage of those people who identified themselves as writers increased from just over 50% at the lowest level of technical knowledge to nearly 80% at the 'able to create dynamic web content' level before remaining in the high 50s to 70% percent level at the higher technical ability bracket. Despite the presence of a high level of technical ability in some members of the community, writers especially, a middle-ground would need to be found between complexity and usability or the system risks losing users. However it

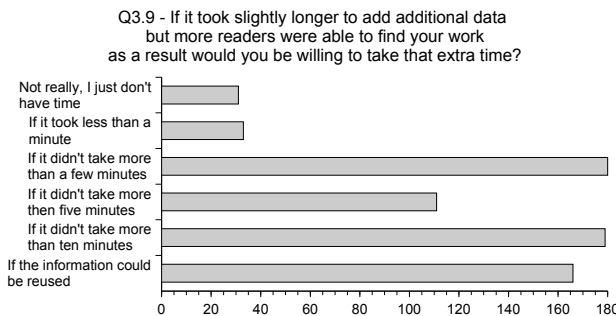


Fig. 6. Additional Time For Semantic Metadata

appears from these results that a system aimed at the current level of GUI web creation software or slightly above would probably be acceptable.

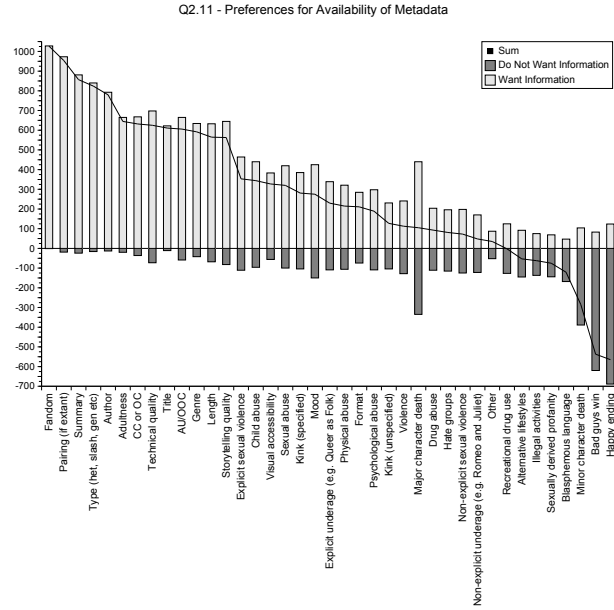


Fig. 7. Metadata Preferences

There was a strong anti-commercial sentiment. While just over 50% of respondents would consider voluntary donations to an author they liked, for upkeep costs or would pay for e-zines, approximately a third would not pay. Beyond the concern expressed by a number of respondents about what would happen to the community if commercial interest became involved approximately 70% would not pay for such a system with just over 20% wanting to test it first and less than 1% saying they would pay for such services.

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 7). 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 (Figure 8). From this we can conclude that personalisation would be a very useful feature and one which the addition of semantic data could provide.

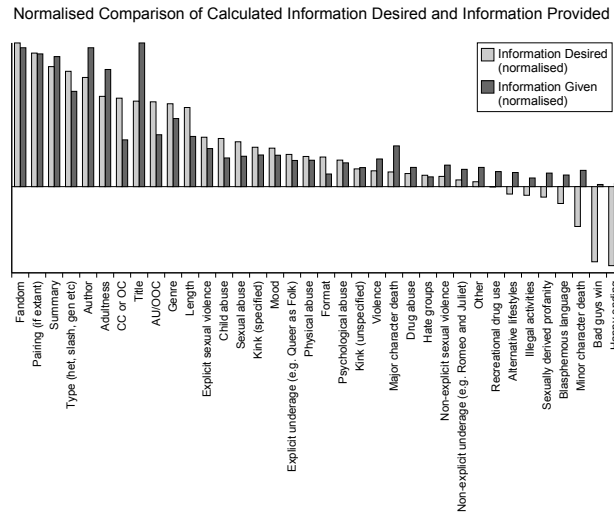


Fig. 8. Currently available Metadata

4 Future work

Work is continuing on using an extension of FOAF and Golbeck’s work of trust [12, 10, 11] to create a community based online identification system which would fulfil both the desire for privacy and mitigate external pressure for access controls.

The OntoMedia Ontology was devised with the results of this survey taken as a requirement indicator to describe the contents and narrative of a media object [17, 14]. This ontology is intended to bridge the gap between the various vocabularies already in use while also allowing the user to define their own understanding of the various terms in use. It is also designed to allow the user greater control over what information is made available to them.

5 Conclusion

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.

Bibliography

- [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. Picrules 1.1. W3c recommendation, World Wide Web Consortium, December 1997.
- [10] Jennifer Golbeck and James Hendler. Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *Proceedings of EKAW 04*, 2004.
- [11] Jennifer Golbeck and Bijan Parsia. Trusting claims from trusted sources: Trust network based filtering of aggregated claims. Submitted to ISWC’04, 2004.
- [12] Jennifer Golbeck, Bijan Parsia, and James Hendler. Trust networks on the semantic web. In *Proceedings of Cooperative Intelligent Agents 2003*, 2003.
- [13] Henry Jenkins. *Textual Poachers: Television Fans and Participatory Culture*. Routledge, New York and London, 1992.
- [14] Michael O. Jewell, K. Faith Lawrence, Mischa M. Tuffield, Adam Prugel-Bennett, David E. Millard, Mark S. Nixon, m.c. schraefel, and Nigel R. Shadbolt. Ontome-dia: An ontology for the representation of heterogeneous media. In *Proceedings of Multimedia Information Retrieval Workshop*, 2005.
- [15] Barbara Kantrowitz. Men, women and computers. *Newsweek*, page 48, 16 May 1994.
- [16] 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.

- [17] Katharine Faith Lawrence. Ontomedia - creating an ontology for marking up the contents of fiction and other media. In *Proceedings of 1st AKT Doctoral Colloquium*. AKT, June 2005.
- [18] 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.
- [19] Dale Spender. *Nattering on the Net - Women, Power and Cyberspace*. Spinifex Press Ltd, North Melbourne, Australia, 1995.
- [20] 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.