Mark Borden for Dr. Lu Xiao

Self-Guided Research: Final Report

Introduction

The Wikipedia Article for Deletions (AfD) forum allows online users to argue for, or against, the continuing existence of a Wikipedia article that is deemed controversial. This form of collective decision making is interesting and rather unique to online open source communities (Schneider, Samp, Passant and Decker, 2013), to which Wikipedia belongs. Articles on Wikipedia are jointly-created and edited by a number of otherwise unrelated collaborators online so it actually makes sense that the curation of that content would be done collectively as well. Min (2007) notes that any discourse involving online communication, via the internet, is underpinned by elements of the democratic process by the very nature of the medium itself. It, again, is not surprising that the AfD forum is both transparent and open to any user which wishes to contribute.

Literature Review

For this literature review I will examine the current academic research on collaborative online environments, more specifically, Wikis. Wikipedia is a major source of information on the internet and is a model that is also within brick and mortar institutions to create shared resources so it is worthy of serious consideration. The uniqueness of the collaborative nature of Wiki's, which is to say user created and curated, is perhaps their most interesting and defining character

and for that reason will be the focus of this review. Understanding user interactions, behaviours and motivations will provide valuable insight to how Wikis develop and are managed.

User Interactions

Min (2007) suggests that, while the benefits of face-to-face deliberations have been empirical proven to be beneficial on improving civic engagement, studies about online deliberations remain scarce. Given the growing prevalence of online communications and, for our purposes the deliberative nature of Wikipedia's Articles for Deletion forum, understanding if the same were true for computer facilitated deliberations would be interesting when considering out results. . Min (2007) designed a study to test just that which aimed to measure participant's opinion and knowledge of particular issues before, and after, deliberating with others online about it. Min (2007) cites former studies that suggest deliberation can change people's opinions quite dramatically on issues but, in his study, this did not hold true. Participants showed little opinion change before and after the deliberation occurred. Interestingly, however, a handful of participants claimed that their position or opinion on an issue actually strengthened after the deliberation occurred. While this is only one study it may be that users, in an online setting, are less willing to budge on certain issues or just less likely to consider alternative viewpoints.

Min's (2007) results are interesting in light of research conducted by Schneider et al. specifically on Wikipedia's Article for Deletion process. Schneider et al. note that online communities like Wikis, open source communities and some standardization bodies use online deliberations between users as their primary way of decision making. If the above discussed research held true one may expect to see these communities bogged down by discussions which

do not go anywhere because users are unable to budge on their opinions in online deliberations. Schneider et al. (2013) note that this, of course, is not the case at all and were interested in understanding the role user experience played in having others accept arguments, if any, and which argumentation schemes are accepted. Schneider et al. (2013) suggest that their findings point to experienced users also being more familiar with Wikipedia's notability policy, the main reason for up to 28% of deletions, and thus are able to argument for effectively using it. The other conclusion drawn by Schneider et al. (2013) is that experienced users do not appeal to emotion as much as inexperienced users or use personal pronouns to suggest courses of action.

User Motivations

Arazy et al. (2010) conducted an interesting research studying examining the motivations behind why users contribute to Wikis. It should be noted that, in their study, Arazy et al. (2010) mostly considered how users contribute to internal wikis. These are wikis located in a corporate or government intranet, generally not accessible outside of the workplace, which employees would be able to contribute to. An example may be employee generated content around a particular software package they may use at work which could be used as a training guide or to trouble-shoot. Arazy et al. (2010) also did consider the site Wikipedia and the growing trend for corporations and governments to help build their articles there and how their research could be applied in this direction. Arazy et al. (2010) consider, for our purposes, two major points in their study. The first being the shift that is occurring with Wiki's away from strict peer-based governance, to a more traditional top-down approach, with the focus here on highlighting editors' recognition to improve article content quality. The idea behind this change is heavily influenced by Wiki's adoption taking place in business and government settings wherein

stakeholders perceive a necessity for attributing contributions to specific editors. The idea is users may be driven to contribute, in these settings, by motivations like advancing their career so it is paramount that work can be attributed to them. The second point which is considered is how this approach changes the perceptions and behaviour of the Wiki editor. If an editor knows that contributions can, and are, ascribable to them does that necessarily lead to improved content in the long run..

The take away from Arazy et al. (2010) is that, although there is merit in the idea according to their empirical findings, context and implementation are major stumbling blocks. Arazy et al. (2010) note that this approach may not be appropriate for an online setting which traditionally values anonymity and collective collaboration rather than a formal institution like the workplace. The other speed bump, and perhaps the harder to overcome, regards implementation. In this paper Arazy et al. chiefly examined whether they could produce a "score" representing the quality of a user's contributions algorithmically. The algorithm attempted to capture the quality of a user's contributions by measuring things like quantity of contributions, how often their material was edited by others, how often their work was cited in other articles, and how frequently articles they worked on were visited as a few examples. The algorithm would then give the users a triple digit score with hope being that this would entice users to want to improve their contributions. Results were mixed and lead the researchers to the conclusion that, although this method showed some promise in improving article quality, more work needs to be done.

Yuan, Cosley, Welser, Xia and Gay (2009) take a different approach and hypothesize that individuals whom work on these collective undertakings, which they refer to as "adopters" in their study, do so for a multitude of reasons for which recognition does not belong. Yuan et al.

(2009) argue, based on their results, that "adopters" involved in contributing to Wikipedia did so because of interpersonal elements including homophily, like-minded people being drawn together by mutual interest, and individual attributes. In other words I believe that their results suggest that there may be something specific or unique about the type of person that wishes to work in this environment which explains at least, partially why this is more of a niche activity. As Yuan et al. (2009) point out that more research exists on adopters of Wikipedia, or other collaborative online environments, then does on people whom never do so. These personality traits or tendencies may factor into some users decisions to, as Stvillia, Twidale, Smith and Gasser (2008) suggest, self-select into certain topics. Stvillia et al. (2008) note that this process can be a double edged sword as, on the one hand, it allows for users work on topics that matches their personal interests and background. For stable topics this dynamic seems to work quite well at producing quality articles (Stvillia et al, 2008). On the other hand it could lead to users self-selecting into controversial topics for any number of reasons and can produce editorial groups which are dysfunctional. Stvillia et al. (99, 2008) present a scenario where dueling editors can sabotage and delete one other's contributions out of malice.

How Users Mange Conflict and Dissent

Arazy et al. (2013) followed up with another study again looking at Wikis but this time attempting to measure the effects conflicts and disagreements have on the group performance. Arazy et al. (2013) conducted an empirical study on 96 Wikipedia Articles looking for signs of group, defined as contributors to the article, conflict with group performance defined by the quality of the article. The major findings of this study is that when conflict was centred around procedural differences or devolved into personal attacks the quality of the articles suffered

significantly. When differences or conflicts were predicated solely on task results were generally indistinguishably from articles without conflict. In general, however, the majority of user posts in this study were task-oriented providing evidence that contributor's preference appears to be towards problem-solving rather than simply disagreeing.

The most surprisingly or unexpected result discovered by Arazy et al. (2013) was the effect group size had on article quality. Large group size had a significant positive effect on the outcome variable which may seem counterintuitive on the surface. One may assume that conflicts would be more likely to occur when increasing the amount of people involved but this result is contrarian to that. While Arazy et al. (2013) did not postulate on why this may be the case insight may be found with one of the conclusions drawn by Black, Welser , Cosley and DeGroot (2011) on similar research related to deliberation in online groups. In a study of 282 posts and threads on Wikipedia regarding the no personal attacks policy and found that nearly half of all posts looked at included a solution to the problem and, 63% of the time, the participant displayed consideration of others' views in a positive manner. Only a minority of posts (145) were deemed 14% by Black et al. (2011) to indicate lack of respect and, of those, only 10% contained explicit evidence of disrespect. The take away from this may be that, in large groups, these minority of disruptive or abusive posters cannot get the same foothold in the conversation that they do in smaller groups. This is speculative but, perhaps, rude or sarcastic comments are buried or not as noticeable as they would be in a discussion involving a handful of contributors.

Cross-Cultural Differences

Hara, Shachaf and Hew (2010) take a similar approach to measuring politeness among user comments and posts to Black et al. (2011) but with two significant departures. The first being that instead Hara et al. (2010) examined talk, user talk, and Wikipedia talk spaces in comparison with Black et al. whom only considered posts and threads regarding the Wikipedia no personal attacks policy. The key difference however is that Hara et al. (2010) examined non-English Wikipedias which, they note, accounts for 75% of Wikipedias on the internet. I think this article is important because there may be something about unique about the English Wikipedia, or its users, which could potentially be influencing research results. Having this study by Hara et al. (2010) on the English, Hebrew, Japanese, and Malay talk, user talk, and Wikipedia talk spaces could provide unique insight. Thirty pages were randomly selected from Wikipedia in each of the four different languages and analyzed. Groups were created (English/Hebrew as Western, Japanese/Malay as Eastern) and user comments were coded. Even though the studies are not directly comparable the results produced by Hara et al. (2010) were similar, or at the very least consistent, to those found by Arazy et al. (2013) and Black et al. (2011) Most user comments were task-oriented postings, as seen in the above mentioned, and community well-being postings, such as courtesy, were the norm which is very similar to the conclusions found specifically in Black et al. (2011) Overall the big take away from Hara et al. (2010) is that they were unable to find definitive cross-cultural differences and found that users behaved similarly among these four groups.

Research Background

Dr. Xiao has previously collected data studying the content of Wikipedia's Articles for Deletion (AfD) forum. This data was made available to me as I was interested in completing a

guided research project, with Dr. Xiao acting as my supervisor, in the Master of Information and Library Science program at Western. During the course of this guided research project I performed qualitative coding on that data with my results presented below.

Research Question

I want to know more about how users employ directive speech, formed as imperatives, when engaging in online deliberations. More specifically I want to understand how users in a collaborative online environment, the AfD, utilize directive speech in their arguments/counter-arguments during deliberations. I think this is an important question to answer because is Wikipedia a major information source now and the article curation process, and specifically how group consensus is reached, deserve academic attention. I also think that crowd sourcing type activities become more common understanding user interactions is important for future research.

Methodology

Using software created by Dr. Xiao instances of users employing direct speech, formed as imperatives, was extracted from 84 separate AfD discussions. The 84 files were then run through an additional piece of software to generate them into the more readable .txt file format. By opening each of the files and counting the number of sentences present within I determined the "Total Sentences" for each. Using a template I created each imperative sentence was copied from the original document and into the corresponding category based on the definitions below:

(1) **Actual Article Suggestion** – A course of action is suggested of how to treat or manage the content of the article being discussed. These imperatives are specific to the article being considered and cannot be generalized to others.

(2) **Wikipedia Technical** –Imperative sentences recommending or instructing others to engage in a technical action within the Wikipedia framework like "merging" or "moving" the article. These are more general as they have to do with the support structures around the article and not the content within it.

(3) **Norm or Cultural Practice** – Imperative sentences centred on instructing or articulating user community behavioural norms and not necessarily the article itself. These imperative sentences can be broader in that they do not have to refer directly to the article at hand.

(4) **Thinking or Prompting** – A user directs others to "think", or more semantically "to consider" or "reconsider", or themselves present additional evidence and/or an alternative viewpoint. There may, or may not, refer to a specific source and can include requesting other users to weigh an abstract concept.

(5) **Reference** – User references material within Wikipedia (ie. Other Articles)

(6) **Reference** – User references material outside of Wikipedia (ie. Other websites)

(7) **Reference** – User references Wikipedia criteria (ie. Deletion Policy)

Derailing Imperatives:

Separate list will be kept of items that are technically imperatives but are not constructive to the process. Examples would be insults, sarcastic or incendiary remarks and other similar sentences that were made to, seemingly, derail the proceedings or conversation.

After coding was completed the total number of imperatives was manually counted and noted at the top of each of the files. The number of imperatives present in each of the categories was then counted and manually added to a spreadsheet in a row alongside the name of the corresponding document. This spreadsheet was later used to sort the data and perform analysis. A list of "hedging words" used within imperatives was kept at the bottom of each of the 84 documents and then totaled in a separate file. These hedging words were drawn from a list provided by Dr. Xiao and was referred to by myself during the initial qualitative coding process

to find matches. Once completed I ran through each of the documents again using the CRTL + F feature to specifically look for hedging words from the list. Lastly imperatives containing internet acronyms like “lol” were spelled out in their entirety. For more obscure or lesser known acronyms urbandictionary.com was consulted and a Google search was used to verify the information.

Special Consideration

Category four, “Thinking or Prompting”, is perhaps overrepresented as identical imperatives suited for that designation would appear multiple times in some of the datasets. For example imperatives asking users to reconsider an argument, categorical a four by the definition in the coding schema, would sometimes appear multiple times in a given set. Arguably these duplicate sentences could be attributed to a single user spamming or repeating themselves, a bot, or some unknown technical issue with the software. I chose to leave those sentences in the final tabulation because, not having access to the original AfD, I am unable to say for certain that it was not multiple users repeating an imperative to drive a point home.

Results

Code	Total	Percentage	Examples
4	448	32%	Let's not forget that newspapers are a for-profit enterprise, as are the vast majority of news sources. Remember, we both agreed that after cutting needless plot summary, the article would only amount to a stub. (Note that there is a topic Topological quantum computer, referred to in the article, which is quite different and does appear to have a substantial independent research literature and research community.) Let's not pull the trigger so fast.

			Please provide evidence that this is the case.
1	283	20.8%	Go ahead and bundle that article in if you'd like, and see "Income in the United States" template at the top right-hand corner of that article for more. In cases where this may be necessary, (e.g. Nationwide opinion polling for the United States presidential election, 2012), consider using tables to enhance the readability of lengthy data lists. Consider using tables to enhance the readability of lengthy data lists. Where it is not necessary, as in the main article United States presidential election, 2012, omit excess statistics altogether and summarize any necessary data concisely. Remove the subsections for each edition, remove line breaks after each sentence, create proper paragraphs and you 're left with content taking up much less space.
5	166	12.2%	Check the Dutch or French versions of the page and you will find them. See "WP:NRVE" for additional information. See, for example, Rabindranath Tagore and Rabindra Sangeet. See, for example, Kazi Nazrul Islam and Nazrul Sangeet. See also Wikipedia:Articles for deletion\/Alpha (Magic:The Gathering) (group nom of 16 articles).
3	124	9.1%	Incidentally, at the risk of being a stickler for language please note that this is not a vote but a discussion. Please, only explain with arguments your point of view. Please allow admins as much time as possible to review Take it to "WP:ANI" If you have a problem with my editing patterns, there are more appropriate forums than this. Please keep it civil.
7	111	8.2%	Please refer to Wikipedia general notability guidelines on significant coverage (http :\ \/ \ \/ en.wikipedia.org \ \/ wiki\/Wikipedia:Notability #General _ notability_guideline).

			But as Garamond has now started dismissing valid sources as invalid, let me reproduce here the Wikipedia policy regarding use of self-published material. Please consider studying Wikipedia:Arguments to avoid in deletion discussions, especially its "Denying the antecedent" section. Please see Wikipedia:Notability _ (organizations_and_companies) #Primary _ criteria. Please see "WP:SPS" Please note the text of the referenced patent, which cites the product's legal novelty and distinctions from products in the kitchen appliance space.
2	106	7.8%	Summarize, merge, and redirect to 2012 Burgas bus bombing. Try clicking ` Refresh ' after you get that error. Merge with Ken Wilber. Either Keep or merge to Ultra Monsters. Either merge with audiobooks or rename to List of BBC audiobooks. Merge and redirect to Subhash Chandra Bose.
DR	90	6.61%	Use some damn common sense. If you like Wikinews so much, go write there instead. Please explain ... You explain yourself, you miscreant !! He's notable, do a little research before I set you up with Manti's girlfriend! If the shoe fits, fine, but if it doesn't, stop maiming your feet. read the damn intro, Hammer.
6	65	4.8%	Have a look here on WorldCat. Look at Blingo for instance. See p. 343 ("Romantic Attack" is followed by "Romi Opening") or p. 472 (the opening index, where no variations of the Marshall Attack are listed.). Check this link:http :\ \/ \ \/ www.travelchannel.com Click on the book, which takes you to google books where you can see that it is mentioned one time and only in passing (on page 313).

*Percentages rounded

In total 2734 sentences were captured from the 84 datasets by the software and, of those, 1361, or 49.78%, were imperatives. Users directing or prompting others to reconsider an argument or issue, or undertake some action to do so, was the most common use of directive speech in this study. Imperatives in this category account for 448, or 32%, of the total. 283, or 20.79% of the total, were users directing others on matters specifically related to the article in question. In light of Schneider et al. (2013) discovery that Wikipedia's notability policy is the main reason for up to 28% of article deletions it is interesting to see these results. In this study users referred to other articles or sections, like user pages, 166 times whereas imperatives referring others to Wikipedia criteria accounted for 111 of the total imperatives. Users using directive speech to correct or inform others on cultural or normative practices were tallied 124, or 9.11% of the total. 111, or 8.15%, of the imperatives were categorized as derailing which was difficult to pin down exactly, given that context was missing, so benefit of the doubt was given to imperatives on the fence with the most egregious and obvious being placed here. Lastly users referring to sources outside of Wikipedia accounted for 65, or 4.77%, of the total.

Discussion

Arguably the biggest take away from this study is the breakdown of how users engage with directive speech, via imperatives, to interact with one another of the Wikipedia AfD forum. More specifically the results suggest that, far more often than not, users remain on task during the deliberation process. Imperatives where users asked others to reconsider an idea or argument, or prompt them for more information, and those related directly to the article appeared most frequently. These findings are consistent with those of Arazy et al. (2013) as they also concluded that most user interactions were task-oriented and it was only a minority of the time that this was not the case. This remains true when it comes to users citing or referring to materials external to

the article in question as those three categories combined accounted for 25%, or ¼, of the imperatives studied for this project. As mentioned above this 25% figure is in line with Schneider et al. (2013) discover that they found the notability criteria was used to justify 28% of article deletion in their study. I believe that users citing policy, or referring to other materials, are examples of task-oriented behaviour and, again, are quite consistent with other research discussed earlier in this paper.

Interestingly the findings in this study on users engaging in derailing comments, which are imperatives that are insulting, sarcastic or otherwise incendiary are similar those uncovered by Black et al in their research. Black et al. found that 10% of user comments contained explicit evidence of disrespect and, in this study, 8.15% of all imperatives were coded as derailing. It should be noted that Black et al. looked at other discussion pages on Wikipedia, whereas this study focused on the AfD forum. In addition Black et al. considered any and all user comments and, for our purposes, we looked specifically at imperatives. With that being said, although these results may not be directly comparable because of these differences, they are in line with one another and are potentially telling as to how users conduct themselves online. It would be interesting to consider these results in light of the research conducted by Arazy et al. on the effect attributing work to a specific editor has on user motivation and conduct. If there were appreciable repercussions for rude or otherwise socially unacceptable behaviour on Wikis, as there would be in a workplace setting, it would be interesting to see how many of these types of comments persist. It may also yield dividends to compare how many rude or sarcastic comments are made in face-to-face deliberations compared to what we see here.

It would seem that AfD users discuss their ideas rather matter-of-factly and/or use language that is decisive as evidenced by the low occurrence of hedging words present in our

findings. This is consistent with Schneider et al which suggests users, especially experienced editors, do not shy away from using strong language to articulate themselves during these discussions. As discussed above Min (2007) found that, although users were likely to change their opinion on an issue after having deliberations face-to-face, online deliberations did not have the same effect. Perhaps there is something about how users interact with each other online, or the anonymous or detached nature of electronic communications, that prompts people into entrenching themselves when challenged about an issue.

Conclusion and Possible Future Research

By and large the conclusions reached in this study are consistent with previous literature on the subject. The majority of users engaging in online deliberations during the AfD process are using directive speech towards one another in a task-oriented way. For future study it would be interested to expand on Min's (2007) findings to see if there is anything specific to online deliberations, compared to face-to-face, which causes users to be less receptive to changing their opinions on issues. Another possible avenue for future research may be looking at how users interact with one another in other online collaborative communities like multiplayer videogames which require teamwork to complete a task. It would be interesting to see if, in other online communities, if users remain as task-oriented when confronted with differing opinions as the evidence suggests participants of Wiki's AfD are.

References

Arazy, O., Stroulia, E., Ruecker, S., Arias, C., Fiorentino, C., Ganev, V., & Yau, T. (2010). Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. *Journal of the American Society for Information Science and Technology, 61*(6), 1166-1179.

Arazy, O., Yeo, L., & Nov, O. (2013). Stay on the Wikipedia task: When task-related disagreements slip into personal and procedural conflicts. *Journal of the American Society for Information Science and Technology, 64*(8), 1634-1648.

Black, L., Welser, H., Cosley, D., & DeGroot, J. (2011). Self-governance through group discussion in Wikipedia: Measuring deliberation in online groups. *Small Group Research, 42*(5), 595-634.

Hara, N., Shachaf, P., & Hew, K. F. (2010). Cross-cultural analysis of the wikipedia community. *Journal of the American Society for Information Science and Technology, 61*(10), 2097-2108.

Jodi Schneider, Krystian Samp, Alexandre Passant, and Stefan Decker. 2013. Arguments about deletion: how experience improves the acceptability of arguments in ad-hoc online task groups. In *Proceedings of the 2013 conference on Computer supported cooperative work* (CSCW '13). ACM, New York, NY, USA, 1069-1080.

Min, S. (2007). Online vs. face-to-face deliberation: Effects on civic engagement. *Journal of Computer-Mediated Communication, 12*(4), 1369-1387.

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology, 59*(6), 983-1001

Yuan, Y. C., Cosley, D., Welser, H. T., Xia, L., & Gay, G. (2009). The diffusion of a task recommendation system to facilitate contributions to an online community. *Journal of Computer-Mediated Communication, 15*(1), 32-59.