Title	A Content Analysis: How Wikipedia Talk Pages Are Used
Author(s)	Schneider, Jodi; Passant, Alexandre; Breslin, John G.
Publication Date	2010
Publication Information	Jodi Schneider, Alexandre Passant, John G. Breslin, (2010) 'A Content Analysis: How Wikipedia Talk Pages Are Used.' In WebSci2010, Web Science Conference. Raleigh, NC
Item record	http://hdl.handle.net/10379/2501

A Content Analysis: How Wikipedia Talk Pages Are Used

Jodi Schneider, Alexandre Passant
Digital Enterprise Research Institute
National University of Ireland, Galway
Galway, Ireland
firstname.lastname@deri.org

John G. Breslin
School of Engineering and Informatics
National University of Ireland, Galway
Galway, Ireland
john.breslin@nuigalway.ie

1. INTRODUCTION

Coordination and decision-making in Wikipedia happens, in part, at the article level, through discussions on an article's Talk page. In recent years, Talk pages have been added more quickly than articles, growing at a rate of 11x, compared to a 9x growth rate for articles [8]. Over a 2.5 year period, [7] found that edits to Talk pages nearly doubled, from 11% to 19% of all page edits, while article edits nearly halved from 53% to 28% of all page edits.

While other studies have examined Talk pages, their scope has been limited, either in the manner of selection (e.g. 25 "purposefully chosen" articles from [8]), the sample size (e.g. [3] coded Talk pages for 9 articles, compared by discipline and rough page age), or the categories of articles chosen (e.g. [7] studied information quality discussions in featured and random articles). Purposefully chosen articles may obscure similarities and differences between categories of Talk pages; furthermore, small sample sizes may hide effects that a larger sample could reveal.

We report on a larger, in-progress study of Talk pages, consisting, in part, of a manual content analysis of 100 Talk pages. Analysis has been completed to date on three categories (comprising 60 articles, and 58 unique articles), making our pilot already comparable in size to Stvilia's [7] analysis. Further, we use this analysis to discuss how structured and meaningful annotations, based on dedicated ontologies and Semantic Web technologies and added to Talk pages with a lightweight annotation process, could help to better classify the type of edits that happen in these pages. Consequently, decision-making and page management based on Talk page edits could potentially be streamlined.

The structure of the paper is as follows. Following the introduction, we review the state of the art of research on Wikipedia Talk pages. In Section 3 we explain our selection criteria and detail the five categories of articles to be analyzed. Section 4 discusses preliminary analysis of 58 articles in three of these categories. Section 5 discusses a lightweight annotation system for Talk pages based on Semantic Web technologies. We conclude with a discussion of future research.

2. RELATED WORK

In [7], the authors examined 60 discussion pages in their analysis of information quality. They tallied the types of info quality problems discussed in Talk pages, hand-coding the problems into non-mutually-exclusive categories based on the following terms: accessibility, accuracy, authority, completeness, complexity, consistency, informativeness, relevance, verifiability, volatility.

[8] studies Talk pages quantitatively, revealing that articles with Talk pages are more highly edited, and have more editors than articles without Talk pages. In particular, "94% of the pages with more than 100 edits have related Talk pages". Further, they manually classify 25 Talk pages, chosen by hand to ensure a wide variety of topics. The dimensions used in their classification are further discussed in Section 4.1.

[3] studied various aspects of 9 articles and their discussion pages. Like [8], they found that requests/suggestions for editing coordination were the most common contribution to Talk pages. However, the remaining categories differ significantly in their prevalence between the two analyses, pointing for further needs for standardization of the categories, as well as a better understanding of other types of contributions to Talk pages. They detail the interaction between Talk page discussions and the article evolution, pointing out how Talk page comments from one editor may support actions of a subsequent editor, how editors use Talk pages to request assistance, and how bots provide justification for some of their edits on Talk pages. Timescales are of particular note; in the examples cited, discussions and relevant edits may be separated by "three hours" or by "six days" or more. The article also provides an example of how vandalism to articles may coincide with vandalism on Talk pages, and how Talk pages may draw attention to controversial sections of an article. However, during the observation period, "several of the issues raised in the Talk page discussions did not directly translate into article editing activity" [3].

Improvements to Talk pages in MediaWiki have been suggested since at least 2006, when a MediaWiki extension, LiquidThreads[1] was developed as part of Google Summer of Code 2006, to add threaded comments, permalinks, archiving, and subscription features to while maintaining wiki syntax and formatting. A number of extensions allow MediaWiki's discussion pages to function as issue trackers, bulletein boards, and guestbooks, as well as to make discussion pages more visible[2].

Finally, there is a growing body of work about coordination in peer production, and in particular in Wikipedia. Much of this work throws light on the larger phenomena, and research on Talk pages often takes place in the context of coordination (e.g. [6]).

3. METHODOLOGY

We are currently undertaking a manual content analysis of Talk pages from 100 articles and a quantitative analysis of Talk pages from 5000 articles, with one-fifth of the sample from each of the following categories:

---

[1] http://www.mediawiki.org/wiki/Extension:LiquidThreads
[2] http://www.mediawiki.org/wiki/Category:Discussion_and_forum_extensions

1. Articles with the most contributors in November 2009.[3] We examined Talk pages for the articles with the most (372 to 77) contributors[4], to see whether coordination techniques changed as the number of editors grew.

2. Most-viewed articles of 2009[5]. The most-viewed articles for a year[6] are an indication of the topics attracting sustained attention; articles which are heavily viewed may also be heavily edited, such as "Deaths in 2009", which appeared both in this category and the previous one.

3. Controversial articles[7]. Controversial articles require significant coordination in order to come to consensus, and we expect increased discussion length and volume. Using a dedicated script, we randomly selected articles,[8] using the categories given at the list of controversial articles[9]. One topic randomly selected, "Berlin Wall", also appeared in the "most contributors" group.

4. Featured articles. Articles must go through a review in order to attain Featured Article status[10], requiring nomination and subsequent coordination, some of which may occur on the Talk page. We expect discussion before editing to be the norm for these articles, which are selected in part based on their stability.

5. Random sample. We expect Talk pages to differ according to the age of the article, number of participants, frequency of editing, and type of article. A random sample may draw attention to other factors requiring further consideration, and will validate or call into question discoveries made on the other categories of articles.

---

[3]The 20 articles selected by this criterion were: "Fort Hood shooting", "Deaths in 2009", "2012 (film)", "Call of Duty: Modern Warfare 2", "The Fame Monster", "Climatic Research Unit e-mail hacking incident", "The X Factor (UK series 6)", "Assassin's Creed II", "Bad Romance", "Rated R (Rihanna album)", "Berlin Wall", "2010 Formula One season", "Herman Van Rompuy", "John Allen Muhammad", "Lady Gaga", "2009 World Series, November 2009 Great Britain and Ireland floods", "Michael Jackson's This Is It", "New Moon (2009 film)", and "France vs Republic of Ireland (2010 FIFA World Cup Play-Off)". The remaining numbers 21-25 have approximately the same number of editors as #20, with 72 contributors for the 25th article.

[4]http://stats.wikimedia.org/EN/TablesWikipediaEN.htm

[5]The 20 articles selected by this criterion were: "Wiki", "The Beatles", "Michael Jackson", "YouTube", "Wikipedia", "Barack Obama", "Deaths in 2009", "United States", "Facebook", "Swine influenza", "Eminem", "Lost (TV series)", "Watchmen", "World War II", "Twitter", "Transformers: Revenge of the Fallen", "Slumdog Millionaire", "Lil Wayne", "Adolf Hitler", and "India".

[6]http://wikistics.falsikon.de/2009/wikipedia/en/

[7]The 20 articles selected by this criterion were: "UEFA Cup finals", "Plato", "Hans Asperger", "Occitan Language", "Disk jockeys", "Death Note", "Environmental vegetarianism", "BitTorrent(protocol)", "Sex offender", "Orgasm", "Separation of Church and State", "Christophobia", "Berlin Wall", "Antisemitism", "Poppers", "Mental disorder", "American Jews", "American Hunters and Shooters Association", "Nancy Pelosi", and "Stephen Colbert".

[8]1 from each of sports, philosophy, psychiatry, linguistics, entertainment, media/culture, environment, and technology. 2 from each of sexuality, religion, history, science/biology/health, politics/economics, and people.

[9]http://en.wikipedia.org/wiki/List_of_controversial_articles

[10]http://en.wikipedia.org/wiki/Wikipedia:Featured_article_candidates

This sample is chosen to counter two limitations of existing research: small sample size and limited understanding of the significant variance between pages.

First, the sample size is larger than previous research. Previous studies (e.g. Bencherki and Uwatowenimana [1]) have reported on as few six talk pages ("which in fact totalled more than a hundred pages when printed out"). Our pilot study of 58 articles is comparable in sample size to the largest published study, which used 60 featured and random articles [7].

Second, we use existing groups of articles to investigate the categorical similarities and differences between Talk pages. Although previous studies have noted significant variance between articles (e.g. see Table 7, the Distribution of postings on talk pages, in [8]), little study has focused on the differences between Talk pages, outside of Featured Articles [7]. While [3] categorized Talk pages by discipline and page age, her study mainly piloted methodologies, and assembled further evidence of this variance. We hope that by taking a significant (20 article) sample from each of these 5 categories, we might identify factors in the variance. Follow-up with quantitative analysis, which has been used to a greater or lesser extent in most previous research, should also be helpful. We are particularly interested in see how editing patterns correlate (or not) between an article and its Talk page.

Next we discuss the content analysis of the first three categories.

4. CONTENT ANALYSIS

Our content analysis used 15 non-mutually-exclusive classifications. We provide an example of each type below. In addition to existing classifications, we have added 4 new classifications. This adds a new dimension to the analysis because some of these classifications are widely used. However, since comments may be assigned several types (aside from the generic "Other" classification), our results can be compared to previous research.

4.1 Classifications from Viégas

First, we used the 11 classifications defined by [8], as represented in Table 1. The examples in that table all come from Wikipedia Talk pages that we analyzed.

Figure 1: Comment which proposes a new infobox, from the Swine influenza Talk page

4.2 Additional Classifications

Our classification began organically from the items in Talk pages we reviewed. These coalesced into a set of classifications, which we then compared with classifications used in [8] and [7]. While Stvilia's classifications were narrowly focused around information quality, Viégas' classifications were very similar to our own. Since they had already been used for at least two studies, and matched our own needs, we decided to adopt Viégas' classifications. To

Requests/suggestions for editing coordination	Ideas, comments, or suggestions involving editing the article.
(example)	"Currently some of the refs are YYYY-MM-DD format and some are Month DD, YYYY. Which format do we want to standardize to? Staxringold talkcontribs 01:38, 3 November 2009 (UTC)"
Requests for information	Questions asked by someone who doesn't intend to edit the page.
(example)	"Where is Ligurian spoken in the Var ? Ericd 00:19, 26 Jun 2005 (UTC)"
References to vandalism	Mentions of vandalism.
(example)	"I've semi-protected the article for another week, the signal-to-noise ratio of the IP edits seemed too low. Tim Vickers (talk) 18:06, 4 May 2009 (UTC)"
References to Wikipedia guidelines and policies	References to guidelines and/or policies of Wikipedia.
(example)	"The section I removed had no sources / references - if you have sources they're no good being kept a secret ;) WP:VERIFY, WP:CITE. Thanks/wangi 08:45, 5 July 2006 (UTC)"
References to internal Wikipedia resources	References to internal Wikipedia resources such as diffs, Talk page discussions, old version of a page.
(example)	"Would it be a good thing to re-add the links that were taken off in August? Somebody made them into a template that was subsequently deleted. The edit to recover the old links is here: [6] Star Garnet (talk) 19:51, 24 October 2009 (UTC)"
Off-topic remarks	Remarks not relating to editing the article.
(example)	"PLATO IS THE BEST MAN ALIVE! LONG LIVE PLATO Preceding unsigned comment added by 69.6.101.167 (talk) 02:19, 8 December 2009 (UTC)"
Polls	Formal proposals followed by statements such as Support and Oppose, with justifications.
(example)	"A month should be deleted from the "Deaths in [CURRENT YEAR]" page ONE WEEK after the month ends..."
Requests for peer review	Requests for peer review.
(example)	"Users hoping to elevate articles to featured status may solicit a peer review." [8]
Information boxes	Special boxes with information, usually found at the top of a Talk page. Occasionally new info boxes for the article are proposed and may be discussed on the Talk page, for example, the Swine influenza page discusses such a template.
(example)	See Fig. 1
Images	Images posted on the Talk page
(example)	See Fig. 2
Other	Describes items that don't fit elsewhere. Unlike the other categories, this one is exclusive.
(example)	"This review is transcluded from Talk:Wiki/GA1. The edit link for this section can be used to add comments to the review."

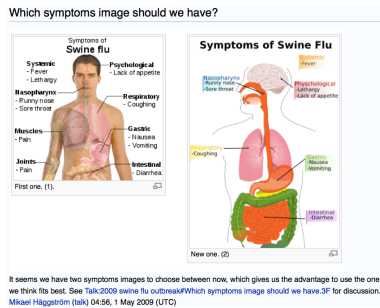Table 1: Classifying comments in Talk pages

Figure 2: Comment which contains images, from the Swine influenza Talk page

capture other features we were interested in, we added 4 new, non-mutually-exclusive classifications as represented in Table 2.

These 4 new classifications were designed to draw attention to other discussions we expected to take place. First, how often were sources explicitly discussed? Second, how often did editors come to discuss or contest reverts, removed material, or other controversial new edits? Third, how often did an editor use the discussion space to notify others about a new edit. Fourth, how often did editors use a Talk page to draw attention to another project?

We added these types because:

- Sources are heavily discussed in Talk pages, and some comments seem to be made solely to deposit a source. While many sources are on the open Web (and can be detected as external links), print resources, inexact references, and deep web resources may also be provided.

- Disagreements about article content often take place in the context of reverts to the page. Discussions about removing content or editing controversial material may also take place on the Talk page before the article is edited.

- The Talk page may be used to notify other editors about a recent edit, perhaps to provide further description, anticipate questions, or clarify that a suggestion has been implemented. Editors may also explain their own edits in discussions of reverts and edit wars.

- The Talk page is often seen as a site for communication with editors who have interest in or knowledge about a given topic. Requests for help, like Requests for information, draw on that perceived expertise.

4.3 Results of the Content Analysis

For the articles with the most contributors (Fig. 3(a)), the most common elements of Talk page postings are coordination, references to guidelines, indication of sources, discussions of reverts and contentious edits, and off-topic remarks; the remaining classifications apply to fewer than 7% of postings on average. Four classifications—references to vandalism, posting of images, recruiting help, and peer review requests—each averaged less than 2% of postings in this category.

For the most-viewed articles (Fig. 3(b)), the most common elements of Talk page postings are coordination, references to guidelines, info boxes, discussions of reverts and contentious edits, notifications of edits made, and polls; the remaining classifications apply to fewer than 7% of postings on average. Two classifications—recruiting help, and peer review requests—each averaged less than

2% of postings in this category. Infoboxes are particularly significant; one indication of their length is the "skip to table of contents" infobox at the top of some Talk pages! On [11] discussion is archived quickly, leaving mostly infoboxes on the Talk page. Seven of the 20 most-viewed articles were featured Articles[12]; roughly one-tenth of a percent of Wikipedia articles are featured[13]. Almost all the articles had restricted editing[14][15] Surprisingly, that contributes to the relative frequency of images: unverified users may request particular edits, and subsequent notification of completion often contains a checkmark (or X for non-completion). On a few pages[17], revision, selection, or addition of images was a point of discussion. Discussions of vandalism sometimes occurred in the context of protecting or unprotecting a page; for instance a user might comment that IP users made helpful edits, or that it had been awhile since unprotecting the page had been tried.

Surprisingly, controversial articles did not have significantly more discussion overall[18]. However, some discussions became repetitive, with participants cycling through the same arguments without convincing each other. Wording and naming proved important in many cases: phrasings were more carefully discussed and new contentious material was often removed to the Talk page. Furthermore, many of these articles had been renamed or recently had a move proposed, perhaps because in some controversies naming indicates differences between perspectives. Sources were often discussed in general terms, for instance in the context of claims of systematic bias; individual sources were less frequently posted or discussed on the Talk page.

Completion of the analysis will prove instructive, for considering whether controversial articles do have more discussion than random pages. Other groupings of controversial articles could be considered, for instance by limiting to pages with significant current editing of the article and/or Talk page, or by examining articles undergoing neutral point of view disputes[19], which an editor has identified as possibly biased.

Our analysis proceeds with Featured Articles (over ten times larger than other Talk page according to [7]) and random articles. Based on our preliminary research, we notice that throughout Wikipedia, different standards are applied by the individual editors who gather on Talk pages. The number of repeat editors and the extent to which individual editors are mutually engaged in conversation varies widely. So does the timescale of comments: If many editors watch a page, responses could come in minutes, but for other Talk pages, months or years might pass between comments. Comments may serve not only to advance particular arguments, but also to improve the discussion, and some otherwise off-topic comments may serve to lighten the mood (i.e. to change the topic after a difficult discus-

---

[11] http://en.wikipedia.org/w/index.php?title=Talk:Michael_Jackson&oldid=341362256

[12] By contrast, there were no featured articles in the most contributors and controversial selection; one featured list was in the most contributors list

[13] http://en.wikipedia.org/wiki/Wikipedia:Featured_article_statistics

[14] "Watchmen", "Slumdog Millionaire", and "Deaths in 2009" were not semi-protected when reviewed

[15] One article, Barack Obama, was further on article probation,[16] which provides consequences for unhelpful editing.

[17] Swine Influenza, 2009 Great Britain and Ireland floods, World War II.

[18] In fact, one selected Talk page, for Sex offender, had no archives and no info boxes, indicating a very limited amount of past discussion and coordination.

[19] http://en.wikipedia.org/wiki/Category:NPOV_disputes

References to sources outside Wikipedia	References to sources outside Wikipedia, including print and dark web resources.
(example)	"Exclusive! Mighty Stef records football protest song"Hot Press. Not sure where to put it but I'll leave it here as somebody might find it useful... –candlewicke 20:53, 27 November 2009 (UTC)"
References to reverts, removed material, or controversial edits	Discussions of reverts, removing material, or controversial edits.
(example)	"I noticed some people edit the page into what it will be in 10 minutes but someone is reverting it...just let it be. The Fear (talk) 02:13, 11 November 2009 (UTC)"
Reference to edits the discussant made	Applied when an editor discusses his/her own article edits on the Talk page.
(example)	"Added the About.com review since the review was part of the reception section.–Sevilledade (talk) 04:29, 3 December 2009 (UTC)"
Requests for help with another article, portal, etc.	Solicitations for assistance elsewhere, or recruiting editorial help in the Talk page for another article.
(example)	"This is just to invite attention to the page Facebook statistics just created; of all interested editors. I have just placed a merge to tag in it. Thanks. Arjun#talk 17:53, 27 December 2009 (UTC)"

Table 2: Additional classifications for comments in Talk pages

Figure 4: *HyperNews, an early implementation of hypertext news, asks users to indicate the "kind of message" they are posting*

sion) or as social grooming. The ongoing modification of Talk pages causes some challenges for analysis. For instance, refactoring or removal of unwanted comments is rarely obvious without digging through page history, though it sometimes comes up in discussion, e.g.[20].

5. OPPORTUNITIES FOR SEMANTIC WEB

Structured and meaningful annotations, based on dedicated ontologies and Semantic Web technologies [2] and added to Talk pages with a lightweight annotation process, could help to better classify the type of edits that happen in these pages. Annotations could enable new ways to browse Talk pages based on the comment types, for instance to instantaneously list all requests for help on a particular set of pages, aiming to recruit volunteers for editing. Consequently, decision-making and page management based on Talk page edits could potentially be streamlined.

Regarding the annotation process, the user would be asked to add the comment type, chosen from a short list, when submitting a Talk page comment. Such annotations have already been used successfully in some systems, for instance the HyperNews webboards[21]

(see Fig. 4).

The user's annotation could be leveraged along with a JavaScript plugin when browsing pages, for instance to highlight only certain types of Talk page comments (e.g. show me all References to sources outside Wikipedia). With more sophisticated techniques, only new comments of certain types (e.g. comments made since the user's last visit which are References to sources outside Wikipedia) could be shown. Customizable Watchlists would also be possible (e.g. only alert me to Talk page changes when removed material is being discussed).

Since users, especially novice users, may be unsure of where to comment on particular topics, transclusion is another exciting possibility. For instance, SPARQL[22] could help automatically collate relevant comments, for instance transcluding Requests for information into a more appropriate spot, such as the Reference Desk for that topic, thus enabling new ways to automatically gather particular kind of comments, and facilitating coordination.

We are currently developing a system to enable such annotations and browsing of semantically-enhanced Talk pages. In addition, in order to represent these categories in RDF(S)/OWL[23], we have provided a lightweight ontology available at `http://rdfs.org/sioc/wikitalk` that can be used with the aforementioned annotations and querying process.

6. FUTURE WORK

Our immediate plans include completing the content analysis for featured and random articles and completing a quantitative analysis with particular focus on the relationship between Talk page editing and article editing. By marking up a testbed of Talk pages, and creating a JavaScript plugin, we will also be able to conduct user tests to evaluate the potential both for uptake and for use of the resulting data.

While we have started by analyzing the comments added to Talk pages, another approach would be to start from modeling user tasks, to determine what information might be needed to support those

---

[20]`http://en.wikipedia.org/w/index.php?title=Talk:Poppers&oldid=339247624#Not_a_forum`
[21]`http://www.hypernews.org/HyperNews/SECURED/edit-response.pl/www/collab/conferencing.`

`html`
[22]`http://www.w3.org/TR/rdf-sparql-query/`
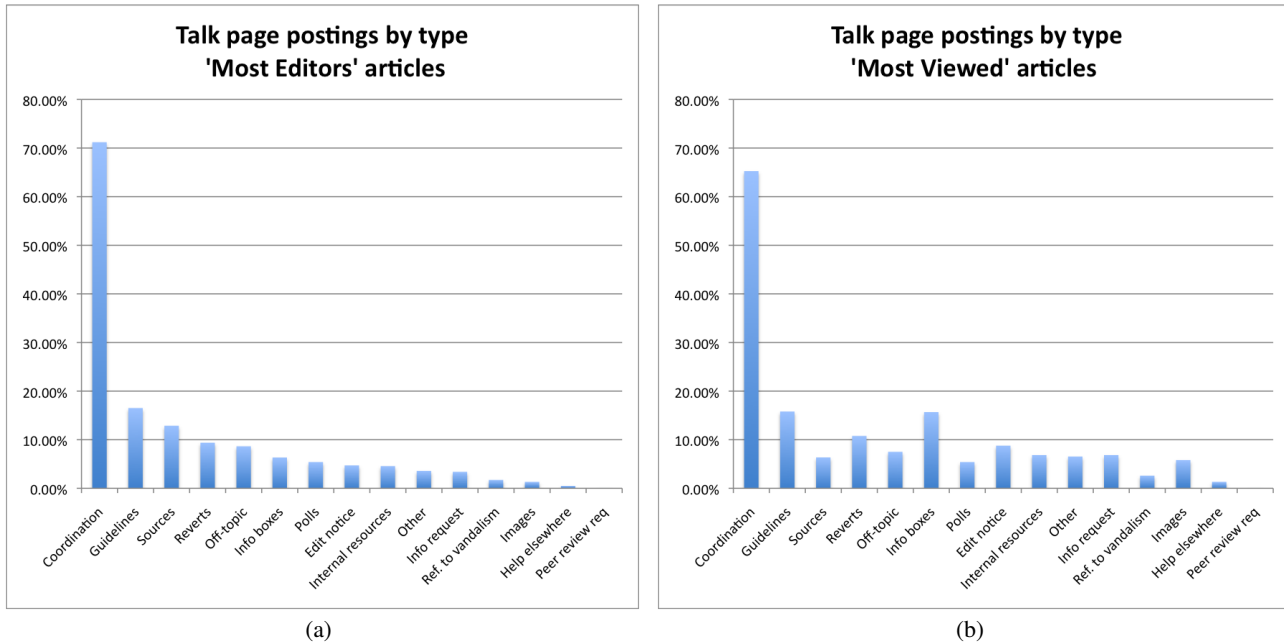[23]`http://www.w3.org/TR/2004/REC-rdf-primer-20040210/`

Figure 3: Distribution of classification in Talk pages for: (a) Pages with most editors; (b) Most-viewed pages.

tasks. Coordination is consistently the most common aspect of Talk page comments, and categorizing coordination according to what the writer is trying to accomplish (e.g. add a section, remove uncited material, reword the lede) could be fruitful. Further exploration and refinement of some existing categories could also be helpful; for instance, edit notifications sometimes indicate completion of a task suggested on the Talk page, but in other cases they open a debate or generate further discussion. These differences might also help suggest alternative rationales and mechanisms for archiving Talk page discussions, beyond current approaches, which are mainly based on time elapsed or number of discussions, as with existing bots[24].

Some recent research has also examined Wikia and the differences in coordination [5]. While Wikipedia is considered a single community, Wikia is construed as a set of over 6000 wikis, with less reliance on guidelines and policies, and smaller, topic-centered communities. This suggests further comparison of Wikia and Wikipedia Talk pages, especially to improve our understanding of the "state-change"-like differences that appear in Talk pages, depending on factors such as the number of users editing the article and the Talk page, whether the article is protected, and so forth. Aniket and Kittur, for instance, suggest community size and type as relevant factors [5]. Further "state change" factors within Wikipedia may be suggested by the quantitative analysis we are also pursuing.

Other research could explore how to reduce the overhead of discussion. Talk pages use varies by type of editor, and may place a particular burden on subject experts; [9] says "when substantive experts contribute to content pages, their contributions are likely to be more costly (take more time and thought, and are more likely to require explanation, justification and discussion on the content Talk pages)." The unlimited nature of debate adds to the burden. As Wikipedia user Hans Adler explains[25]: "That's the problem with

endless discussions: They get even longer because you have to repeat every argument each time someone asks who (understandably) didn't read everything." Naturally, users tire of repeated discussion, and newly-revived conversations are sometimes cut off quickly by long-standing page editors. Unfortunately, avoiding such repetition by claiming consensus is disingenous: new users may have valid counterarguments. Summarizing long-standing debates with argumentation visualization, which is both easier to parse and more rhetorically effective than prose [4] seems particularly promising. Dialogue maps may be a useful technique for sharing a visual overview of these arguments, especially if they can be semi-automatically generated.

7. CONCLUSIONS

We have presented preliminary results from our content analysis of 58 Talk pages in three categories: articles with the most editors, most views, and controversial articles. To structure this analysis we augmented Viégas's classification with 4 new classifications of Talk page comments. The most common types of discussions vary by article type, but include Requests/suggestions for editing coordination, References to Wikipedia guidelines and policies, References to sources outside Wikipedia, and References to reverts, removed material, or controversial edits. The most helpful or most common discussion types could form the basis of a lightweight ontology. When entering a comment, Wikipedia editors could indicate the type of comment from a short list presented in a simple interface such as a radio button or checkmarks. As a result, the system could display comments in a more appropriate spot, for instance to instantaneously list all requests for help on a particular set of pages, aiming to recruit volunteers for editing.

While our study is in progress, one aspect not discussed in previous analyses is worth particular mention. We are surprised that many controversial articles do not have significant discussion, at least compared to articles with the most editors and most views.

---

[24]http://en.wikipedia.org/wiki/Category:
Wikipedia_archive_bots

[25]http://en.wikipedia.org/w/index.php?title=

Talk:Antisemitism&oldid=347719211

6

Comparison with random articles will be instructive, since the most views category includes many Featured Articles, which tend to have highly developed Talk pages.

Further, controversial articles point to another factor: there seems to be significant variance between discussion threads on their Talk pages. While numerous discussions might consist of a single comment, or a small number of comments, other discussions are long and complex, sometimes stimulating further discussion on additional topics (e.g. see [26], which has six discussion threads but yields 19 printed pages, yet almost all the comments are in two threads, both on Madoff. These threads were created by the same user, one month apart.) Our continuing analysis should yield more information.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Nicolas Bencherki and Jeanne d'Arc Uwatowenimana. Writing a wikipedia article: Data mining and organizational communication to explain the practices by which contributors maintain the article's coherence. In *annual meeting of the International Communication Association*, Montreal, Quebec, Canada, May 2008.

[2] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):34–43, 2001.

[3] Katherine Ehmann, Andrew Large, and Jamshid Beheshti. Collaboration in context: Comparing article evolution among subject disciplines in wikipedia. *First Monday*, 13(10), October 2008.

[4] Paul A. Kirschner, Simon J. Buckingham-Shum, Chad S. Carr, and Tim van Gelder. Enhancing deliberation through computer supported argument visualization. In *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, pages 97–115. Springer, 1 edition, 2003.

[5] Aniket Kittur and Robert E. Kraut. Beyond wikipedia: Coordination and conflict in online production groups. In *CSCW 2010*, Savannah, Georgia, February 2010. ACM.

[6] Aniket Kittur, Bongwon Suh, Bryan A. Pendleton, and Ed H. Chi. He says, she says: Conflict and coordination in wikipedia. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 453–462, San Jose, California, USA, 2007. ACM.

[7] Besiki Stvilia, Michael B. Twidale, Linda C. Smith, and Les Gasser. Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001, 2008.

[8] Fernanda B. Viégas, Martin Wattenberg, Jesse Kriss, and Frank van Ham. Talk before you type: Coordination in wikipedia. In *40th Annual Hawaii International Conference on System Sciences*, pages 78–87, 2007.

[9] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith. Finding social roles in wikipedia. In *Proceedings of the American Sociological Association 2008*, Boston, MA, USA, 2008.

---

[26]http://en.wikipedia.org/w/index.php?title=Talk:American_Jews&oldid=344759427