What's in the Content of Wikipedia's Article for Deletion Discussions?

Towards a Visual Analytic Approach

Ali Javanmardi
Department of Computer Science
University of Western Ontario,
London, ON, Canada,
sunny_javanmardi@yahoo.com

Lu Xiao†
School of Information Studies
Syracuse University, Syracuse, NY,
U.S.A., lxiao04@syr.edu

Keywords

Wikipedia discussions; Wikipedia Article for Deletion dataset; Visual analytics; Natural language processing

ACM Reference format:

Ali Javanmardi and Lu Xiao. 2019. What's in the Content of Wikipedia's Article for Deletion Discussions?: Towards a Visual Analytic Approach. In *Proceedings of WWW '19: The Web Conference (WWW '19), May 13, 2019, San Francisco, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3308560.3316750

1 Introduction

As a successful decentralized online peer production system, Wikipedia has a large number of users participating in self-assigned tasks based on their interest. Participants often coordinate on the tasks about editing an article through the article's talk page discussions. Researchers analyze Wikipedia talk page discussions for various aspects such as coordination [13], conflict management [1], emotion [8], leadership [12], the relationship between Wikipedia article's talk page discussion and the article editing activity [5], etc.

Besides talk page discussions, Wikipedia's deletion discussions have also caught researchers' attentions. In Wikipedia, there are four mechanisms with respect to deleting an article from the web site: speedy deletion, proposed deletion, BLP Deletion, and deletion discussion. Deletion discussion mechanism is designed for controversial articles that the deletion needs to be discussed to reach a conclusion. Also, if a proposed article cannot be tagged with one of the first three mechanisms, it will be tagged for deletion discussion.

Researchers have explored various factors of the deletion discussion process and outcome. For example, [19] examined voting behavior in deletion discussions, with a specific emphasis on how early opinions in the debate affect later voices. [6] discussed the importance of being familiar with the Wikipedia community's values, norms, and practices in deciding whether the discussed article should be deleted. [9] considered the impact of Wikipedia's "Ignore All Rules" policy on decision-making in deletion debates. [21] found that the outcome of an article's deletion discussion correlates with the type of votes in the discussion and the category of the article's topic (e.g., art, location, sport, etc.).

Through open and iterative coding processes, researchers analyzed the deletion discussion content. They found that these discussions are logical and follow Wikipedia's policies in general, and notability is the most commonly used rationale for keeping or deleting an article [14, 21]. The discussion participants also referred to various Wikipedia policies and guidelines in justifying their positions regarding the articles [21, 22]. Nonetheless, these studies analyzed small samples, which made it challenging to generalize their findings. Analyzing a large amount of the discussion content is desired to help better understand and leverage this discussion space, but requires computational content analysis. On the other hand, the deletion discussions are unstructured text data. Significant efforts are needed to clean raw discussion content for computational analysis. For example, a user may forget to add a closing tag or his/her signature at the end of the comment, which makes it difficult to detect the end of a user's comment.

To address this challenge, we retrieved about 40,000 deletion discussion content from Wikipedia web pages, cleaned and stored the content of 39, 177 discussions into a structured discussion database. With this cleaned and structured dataset, the automatic processing and analysis of the discussion content becomes more manageable to the researchers. With this database, we developed interactive visualizations that offer insights on how the outcomes of the articles are related to different aspects of the discussions, including the types of votes, the mentioned policies, and the categories of the articles (e.g., art, people, sport, etc.). This database is available at https://drive.google.com/open?id=1n3GJ6h5bSykrC7cgE2_oz4N6LSM3UnaV.

The rest of the paper is organized as follows. We first review studies related to the factors of the deletion discussions and visualizations. We then describe in details the open access structured database including how we cleaned and filtered the noise, and annotated and categorized the data. We then present our interactive visualizations and discuss their implications. We

† Corresponding author.

WWW '19, May 13, 2019, San Francisco, USA

ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316750

conclude with new research ideas that leverage the open access database, shedding light on the future research direction.

2 Related Work

2.1 The Decision Factors in Wikipedia Deletion Discussions

The deletion discussions in Wikipedia are often referred to as Article for Deletion (AfD) discussions. An AfD discussion starts when a user nominates the article to be discussed for deletion and offers his/her justification for the nomination. During the discussion time, any Internet user, often a Wikipedian, can participate in the discussion and deliberate on what to do with this Wikipedia article - to delete it from the Wikipedia, to keep it, or to make another strategic move such as merging it to another article. The participants are required to provide their opinions as well as the rationales that justify their opinions. As a main quality control mechanism, the users' opinions and rationales are expected to follow Wikipedia's policies and requirements about the articles. When a Wikipedia administrator or a general user makes the decision about the article, the AfD decision should be based on the rationales provided by the participants, instead of the majority vote rule of the participants' opinions. Also, often the user who makes the decision and closes the discussion is someone who did not participate in the discussion process.

AfD discussions are mainly con-ducted by a small number of long-established users [6, 18]. The early opinions in a discussion are found to affect the following opinions, and the users exhibit different voting patterns in the process at subgroup level [19]. The naturally formed groups make better decisions than those that recruit participants, and a group with a moderate diversity of newcomer and expert participants concludes better decision [11]. Additionally, non-unanimous situations tend to have more participants, which implies that participants are more inclined to voice their different opinions than agreeing to others 14, 21].

[21] examined whether and how different types of votes affects the outcome of an AfD discussion. They found that the existence of other vote (i.e., a vote other than keep or delete) has a statistically significant impact on the outcome of the article – an article that has the other votes in its deletion discussions is more likely to be suggested for actions other than delete. The authors also found that certain categories of an article correlates with the likelihood of the article to be deleted. For example, articles about people, for-profit organizations, or definitions are slightly more likely to be deleted than expected, while articles about locations or events are more likely to be kept than expected, and articles about nonprofit organizations and media are more likely to be suggested for other options (e.g., merge, redirect, etc.) than expected.

"On Wikipedia, notability is a test used by editors to decide whether a given topic warrants its own article" (https://en.wikipedia.org/wiki/Wikipedia:Notability). Prior studies found that notability policy indeed plays a critical role in the AfD discussions. Taking the assumption that an article's level of popularity correlates well with its level of notability, [10] measured an article's notability by measuring its readership (i.e., the number of visits to the article) and its rank in the Google search. The authors found that the most common reason for deleting an article is its lack of notability. A later content analysis study (N = 229 AfD discussions) also found that notability is the most commonly used rationale for keeping or deleting an article [21].

While these studies have shown interesting results, their data samples are limited, especially in those that focused on the content analysis, e.g., [14, 21].

2.2 Visualizations with Wikipedia Data

Wikipedia data have been used in various visualization projects. For example, 509 administrators' activities in Wikipedia's article editing process were visualized into a colorful tabular format (Chromogram) which showed that the administrators switch between multiple tasks [20]. [2] implemented a radial layout visualization to illustrate the relationships among the categories of Wikipedia. They also visualized data streams of recent changes in Wikipedia by mapping time and activity levels into a two-dimensional display of x and y respectively, and encoding the different types of activities to different shapes [3]. For more examples of visualizations based on Wikipedia content, please refer to the web site: http://infodisiac.com/Wikimedia/Visualizations/.

[17] and [19] leveraged visualizations to investigate Wikipedia's AfD content. The authors chose 100 longest AfDs that had the deletion outcome and visualized them based on the type of the votes (keep or delete) and the sequence of the votes. The researchers also chose 100 longest AfDs that had the keep outcome and visualized them using the same strategy. In addition, they visualized the frequency distribution of the AfDs regarding the length of the discussions. They showed that a visualization can elaborate more levels of details of an online deliberation to the Wikipedia community.

3 An Open SQL Database for AFD Study

AfD discussions are informal and unstructured just as the typical online discussions. It is a technical challenge to apply computational techniques to analyze this type of texts – there are human errors (e.g., typos and poor grammar) and noises (e.g., mis-format, Internet slang). We detail how we obtained, parsed, and organized a subset of Wikipedia AfD data into a relational database that is ready for computational processing or analysis. It is available at https://drive.google.com/open?id=1n3GJ6h5bSykrC7cgE2_oz4N6LSM3UnaV.

Wikipedia manages the deletion discussions based on the date they are proposed for Deletion Discussion. The content of each proposed date is publicly accessible through URLs. An AfD discussion consists of four parts: article title, nomination reason – why the article is proposed for deletion, participants' comments – the participant's opinion and rationale to justify the opinion, and discussion outcome – the decision regarding the article and the rationale of this decision. With the retrieved data,

we observed over thirty types of opinions, e.g., weak delete, delete, withdraw, userfy, speedy decline, relisted, move, etc. And, there are five possible outcomes: Delete, Keep, Redirected, Merged, or None-consensus. All these parts except the first one end with user signatures. A user signature consists of the user's username and timestamp (date and time). Our script visited the web sites through the URLs of the proposed dates from May 15, 2013, to May 15, 2015 (i.e., 720 dates) and extracted these parts from the raw HTML content. To prepare the dataset for further computational analysis techniques, we removed 5% of the dataset that caused difficulties when parsing the HTML content to extract the parts. These difficulties come from the following noises:

Fatal Errors Fatal errors cause a complete shutdown in the parsing process. When such errors occur, we deleted the whole AfD discussion. On average, every 275 AfDs generates 3 to 4 such fatal errors, that is, 1.1% to 1.5% of the error rate.

The Error of Closing and Opening Tags Occasionally, AfD participants did not close HTML tags in their comments. Thus, there are opening tags without the closing tags. These HTML tags are often nested making it more difficult to detect the missing spot automatically. About 0.7% of the AfDs had this problem.

Noises over Very Long Discussions The longer an AfD discussion content is, the more complex it is to handle the noise. In some cases, the first author manually went through the AfDs to tackle the noise. The error rate for this kind is 1%.

Noises over User Signatures Whenever a user participates in an AfD discussion, his signature is automatically generated by the system. Each signature consists of the username and the timestamp. Technically speaking, a user's signature can used to separate subparts of an AfD from each other. However, the user signature has noises. For example, a timestamp might be completely missing from the end of a signature. The error rate is between 0.8% and 1.4% for the noise due to the user signature.

Through this cleaning process, we stored 39, 177 discussions in the SQL database from the raw about 40, 000 discussions. Approximately 80% AfD discussions had between 4 and 12 comments. Additionally, we stored different aspects of the discussion information into different tables and fields. Besides the aforementioned four parts, we also calculated and recorded other information, e.g., the total number of comments in an AfD discussion as well as the total number of keep/delete opinions it contains. Additionally, we identified and stored the category of a Wikipedia article and the Wikipedia policies mentioned in the discussion, as well as the hierarchical relationships among different categories and among different policies. We detail these processes below.

The Category of a Wikipedia Article Wikipedia offers a semi-hierarchical structure scheme for article categories such that one category may have multiple parents besides multiple children. In other words, Wikipedia has a category graph not a tree. Converting this graph into a tree is not a trivial task. To tackle this problem, [3] first constructed the graph, and then identified the parent category for each category by measuring the category similarity based on the co-occurrences of the articles in the two categories and the depth of the current category in the category graph. [4] developed a technique to address the noise in the hierarchy level of the Wikipedia article categories. In our SQL database building process, the first author compared all the parent categories of a category and chose one to be the parent category based on his judgment on the similarity between the parent category and the article being labelled.

The Policy Structure in the Database A Wikipedia policy can have a parent policy and one or more children policies. For example, the notability in a politician is the sub-set of the notability policy. The retrieved HTML data from AfD discussion pages do not show the parent-child relationships of the policies. Additionally, the participants often used shortcuts of a specific policy in the discussion. The mapping between a policy and its shortcuts is usually one-to-many and such knowledge is nested in Wikipedia pages, i.e., there isn't one place where such mapping is presented comprehensively for a policy. For example, Wikipedia:Deletion policy has three shortcuts WP:DP, WP:DEL, and WP:DELETE listed in the directory. But when one goes to the actual policy page, there are more than 20 additional shortcuts for different parts of the policy description. To tackle this challenge, the first author browsed all the Wikipedia policy pages to identify these inner-relationships and the use of multiple terms and shortcuts, and entered the data to the policy table manually. It is worth noting that while Wikipedia site distinguishes a policy from its guideline essays, many guideline essays are referred to and regarded as the policies in the community. Our database therefore did not distinguish the guideline essays from their policies.

Through this process of labelling categories and extracting policy information from participants' comments, the database has about 86% of the comments referred to Wikipedia policies. The total number of unique categories and policies are 321 and 3,583 respectively.

4 Factors of an AfD Discussion Outcome

Taking a visualization approach to explore the factors of an AfD discussion outcome, we developed interactive visualizations to obtain an overall understanding of how the outcomes of the articles correlate with different aspects of the discussions. In our visualizations, we coded a participant's opinion as follows. Code *delete* includes opinions like strongly delete, speedy delete, delete, and weak delete. Code *keep* includes the opposite opinions or final decisions (e.g., weak keep, strong keep, etc.). All the other types of opinions are coded as *other*. In addition, we only considered the discussions that had one of the two discussion outcomes: *keep* (the article) vs. *delete* (the article).

4.1 The Outcomes vs the Number of Keep/Delete Opinions

According to Wikipedia's AfD policy, the decision regarding a proposed article should be based on the participants' rationales in their votes, as opposed to simply following the majority vote based on the total number of different opinions in the discussion.

Curious to find out whether the actual decision-making of AfD discussions follows this policy, we visualized the AfD discussion outcomes and the number of keep/delete opinions in the discussions through an interactive heat map. In the remaining session, we call these opinions votes for the ease of comparison with the idea of majority vote. This interactive visualization is accessible at http://www.mandanemedia.com/afd/view/diagram1.php. Figure 1 shows a screenshot.

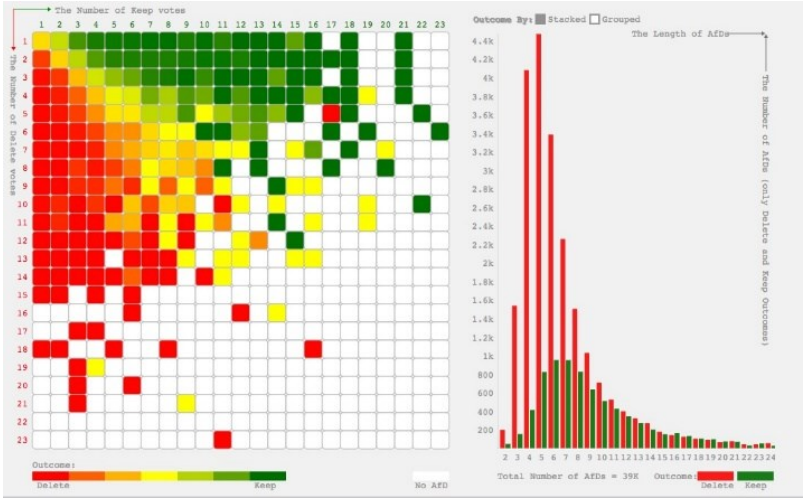


Figure 1: The number of keep and delete votes in AfD discussions and the discussions' outcome

As shown in Figure 1, the number of keep and delete votes in an AfD discussion in our dataset ranges from 1 to 23, represented by the column number c and row number r respectively. A cell represents all the AfD discussions that had c keep votes and r delete votes. The cell color reflects the aggregated outcome for that specific cell: the greener/redder the cell is, the more AfD discussions with keep/delete outcome the cell contains. The yellow cell means that it contains equal or close to equal number of keep outcome discussions and delete outcome discussions. Take the green cell in the first row and fifth column as the example. The cell represents this type of AfD discussions: the discussion that has one participant who suggests to delete the article and five participants who suggests to keep the article. The dark green color suggests that most of such discussions kept the article in the end. If one hovers over the cell, the exact number and percentages of keep-outcome and delete-outcome discussions will be displayed on the interface.

Figure 1 shows that a cell's color moves from red to green when we look from the bottom left-hand corner to the top right-hand corner. This indicates that the outcome of an AfD decision is in general consistent with the majority vote rule: the more keep votes than delete votes a discussion has, the more likely its outcome is to keep the article. Nevertheless, it is evident that the difference exists. An election dataset that follows the majority vote rule will have a green upper triangle and a red lower triangle and the diagonal that separates the two will be yellow indicating equal number of opposite votes. This diagram, in contrast, does not follow this pattern. Specifically, the darkness of the color varies in the upper triangle. Similarly, there are orange (lighter red) cells besides red in the lower triangle. In addition, in both triangles there are yellow cells. A striking example is the cell at the fifth row and seventeenth column. It is red, which means that among the AfD discussions that had 17 keep votes and only 5 delete votes, majority of the final decisions were still to delete the article. This illustrates that the decision is not made simply by majority vote in the discussion.

This visualization also offers a bar chart (see Figure 1) that uses the total number of an AfD's votes as the X-axis and the number of discussions as the Y-axis. The red bar represents the AfDs with delete outcomes, and the green ones keep outcomes. It shows that majority of the discussions in this database had five votes in total, and around 80% of these AfDs had delete outcome.

To better understand the distribution of the AfD discussions according to the number of keep and delete votes, we developed another interactive heat map that paints the color of the cell according to the total number of the discussions in the cell (http://www.mandanemedia.com/afd/view/diagram2.php). On the heat map, the lighter the cell's color, the less number of AfD discussions contained in the cell. The number of AfD discussions contained in a cell ranged from 0 to 7019. One can hover over a cell to view the exact amount of AfDs contained in it. The cells in the first column up to row 5 contains about 56% of the AfD. The cells within columns 1-5 and rows 1-6 contain 91% of the AfDs. The cells within the columns 1-10 and the rows 1-10 contain 98% of the AfDs. As the pattern suggests, most of the AfDs had less than 5 keep votes and 5 delete votes.

4.2 The Policies Mentioned in the Participants' Comments

Helping new editors get familiar with the Wikipedia policies in AfD discussions is important [15]. We developed an interactive visualization to offer an overall understanding about the policies

that the participants mentioned during the AfD discussion (http://www.mandanemedia.com/afd/view/diagram3.php). Specifically, we used a sunburst diagram to represent the policies mentioned in the discussion content and their relative frequencies. A sunburst diagram is similar to a tree diagram, except that it uses a radial layout. The root of the tree is in the center of the sunburst diagram, with the leaves on the circumference around it. Our sunburst diagram consists of three tiers: the first tier, which is the innermost circle of the diagram, represents the highest level of the policies: notability vs non-notability policies. The second tier, which is the middle circle of the diagram, represents the middle of the hierarchy; and the third tier, which is the outermost circle of the diagram, represents the bottom of the hierarchy. The length of the arc of each policy corresponds to the percentage of this policy in the participants' comments in the database.

The interaction and concept of this visualization were inspired from the Visual Information Seeking Mantra: overview first, zoom and filter, then details-on-demand [16]. When one hovers over a particular policy on the diagram, the details about the policy's percentage is displayed and the remainder of the diagram fades. The percentage displayed in the center of the sunburst diagram becomes the percentage of this policy mentioned in the comments. If this policy has child policies, this percentage considered the child policies as well. In addition, when one hovers over a policy in the sunburst diagram, the right-hand side of the diagram displays the policy title, the percentage of the AfD discussions that mentioned the policy, and the percentages of the delete, keep, and other comments in the comments that mentioned the policy respectively. In addition to the policy itself, such information of its parent's policy is displayed as well. Figure 2 shows an example of this situation when one hovers above the policy "Wikipedia: CRYSTAL" (in purple).

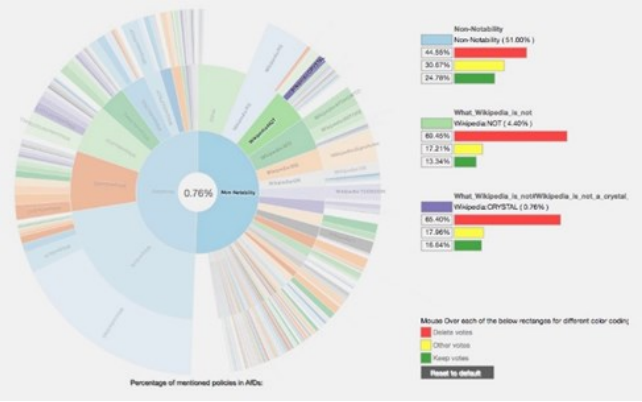


Figure 2: The policies in the AfD votes; the figure also shows the output when one hovering above the particular policy Wikipedia:CRYSTAL

With this visualization, we learned that of the mentioned policies in the discussion data, 49% are notability policies. Also, 59% of the comments that mentioned these policies are for deleting the articles. Among the comments that mentioned non-notability policies, on the other hand, 55% of them are for keeping the article or suggestions other than deleting it. This observation indicates that when the purpose of mentioning notability policies in the comment tends to be to show that the article violates the notability requirement, and the purpose tends to be to "save" the article when non-notability policies are mentioned.

This visualization offers another interaction feature. There are three colored squares on the right-hand lower corner of the diagram representing keep, delete, and other votes. When one clicks on one of them, the diagram visualizes the density of this type of votes in the AfD discussions that mentioned a policy with the darkness of the color. The darker the shade, the higher density of the vote in the discussions contained in the shade. Figure 3 displays the diagram that shows the density of the delete votes in the discussions.

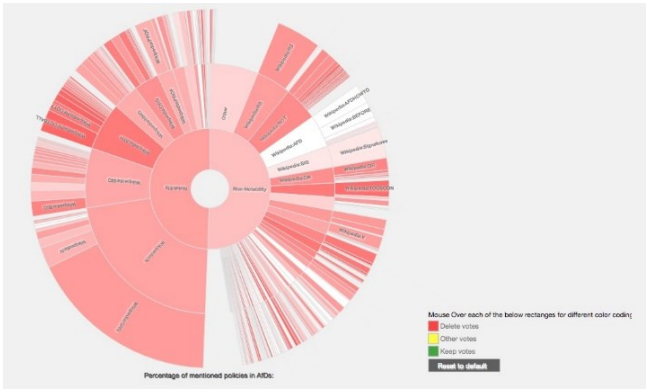
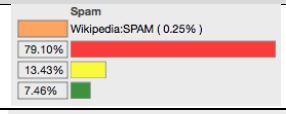
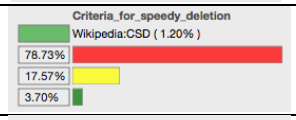
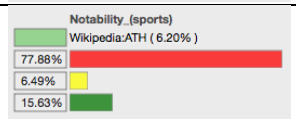
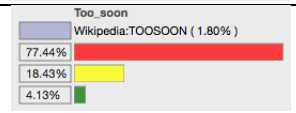


Figure 3: The density of the AfD delete votes that contained the policies

Figure 3 shows that at the second level of the policy structure (i.e., the second tier of the diagram) four policy categories had the darkest color for this type of comment. For each of the four policy categories, the percentage of its delete comments was over 75%. These policies are Wikipedia:SPAM, Wikipedia:CSD, Wikipedia:ATH, and Wikipedia:TOOSOON. Table 1 shows the percentages of the different type of comments (i.e., delete, keep, or other comments) and the percentage of the comments that mentioned one of these four policies in the discussion data. Wikipedia:SPAM and Wikipedia:CSD provide guidelines on what considered to be a spam in Wikipedia and on the criteria of speed deleting an article. The high percentage of delete votes in the comments that mentioned these policies suggest that when these policies were mentioned in the votes, often the participants felt that the articles were spams or should go through the speedy deletion mechanism instead of AfD. Wikipedia:ATH offers criteria on whether an article about an athletic person or a sports player can be included in Wikipedia. The high percentage of delete votes in the comments that mentioned this policy means that when participants mentioned this policy in the deletion discussion often they suggested delete the article. This

implies that when such articles were proposed for deletion, it is often because their authors were unfamiliar with this policy.

Table 1: Percentages of different types of AfD votes that mentioned this policy category

The Policy Category	Percentage of the comments that mentioned this policy category in the comments that mentioned policies	Percentages of keep/delete/other comments in the comments that mentioned this policy category
Wikipedia:SPAM	0.25%	
Wikipedia:CSD	1.2%	
Wikipedia: ATH	6.2%	
Wikipedia:TOOSOON	1.8%	

Sometimes a topic might be notable, but there is no independently reliable source to show it, e.g., an event that just happened and its notability and/or impact have yet to be established. Wikipedia:TOOSOON offers guidelines on these types of situations. The table shows that when this policy was mentioned in one's comment, one's suggestion is often to delete the article. This also implies the author(s)' lack of familiarity with this policy when writing an article that fits into this situation.

At the second level of the policies, three policy categories had over 70% of keep comments in the comments that mentioned them. They are Wikipedia:POINT, Wikipedia:ATATP, and Wikipedia:System bias. They were also rarely mentioned in the AfD discussions, as the percentages of the comments that mentioned them were very low (0.27%, 0.2%, and 0.18% respectively). Interestingly, all three policies are about the discussion norms and/or the community culture/bias. For example, Wikipedia:POINT offers an English Wikipedia behavioral guideline that describes the standard that are accepted by the community and that the editors should attempt to follow. The pages show various example situations as well as what an editor should and should not do. This suggests that when such policies were mentioned in the discussions, the participants' intention was to help "improve and/or save" the article by "citing" these guidelines.

At the third level of the policies, Wikipedia:BEFORE had the highest percentage of keep comments in the comments that mentioned it. Also, about 1.4% of the comments that mentioned a policy mentioned Wikipedia:BEFORE. In these comments, only 3.51% were for delete, 73.54% for keep, and 22.95% for other. This policy describes what one needs to be considered before nominating an article to be deleted (https://en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion#Before_nominating:_checks_and_alternatives). The high percentage of keep comments implies that in a lot of situations when this policy is mentioned it's because the participant believes the nominator made a mistake in nominating the article. The low percentage of mentioning this policy in the dataset implies that such mistake seldom happens.

Lastly, when one clicks on the title of a policy on the right-hand site information panel, the Wikipedia page for that policy opens in a separate window or tab. This feature is designed to support the scenario when a user wants to check the details about the policy while browsing this visualization.

4.3 The Categories of the AfDs

An earlier study with a small sample suggests that the article's category and the decision of its AfD discussion is correlated [21]. Interested in the distribution of various categories in the proposed articles, we developed another sunburst diagram that shows the categories of the articles proposed for delete (http://www.mandanemedia.com/afd/view/diagram4.php). It has the similar kind of interaction mechanisms as the policy sunburst diagram. For example, one can quickly identify which category of the articles receives the highest/lowest percentages of the keep or delete comments. This tree has four levels with seven categories at the highest level, i.e., the innermost level: Culture and the Arts, General reference, Geography and places, Health and Fitness, Religion and belief systems, Society and Social sciences, and Technology and Sciences.

From this visualization, we learned that three categories have the highest percentages of the delete comments, and they are Martial Art, Football, and Sportspeople. Specifically, 0.98% of the AfDs in our dataset are about Martial Art, and in these discussions, 84.65% of the comments were for delete. Football articles take about 1.5% of the AfDs, and 77.34% of their comments were for delete. About 3.0% of our AfDs were

sportspeople articles, and 77.47% of their comments were for delete. This suggests that when articles of these categories are proposed for Deletion Discussion, likely they will receive a lot of delete comments.

It is also observed that three categories tend to have higher percentage of keep comments: History, Crime, and Architectures. The articles in these categories appear rarely in our AfD dataset, as their percentages in the dataset are 0.57%, 0.57%, and 0.48% respectively. While their keep comments do not take the majority of their votes (37.45%, 35.20%, and 33.48% respectively), they are still higher than other categories. Also, in these three categories, the other comments percentages are moderate (23.27%, 21.66%, 20.68%). When we combine the percentages of keep and other comments together, they are the majority in the comments of these categories. This implies that articles of these categories tend not to be proposed for deletion, and if they are, there are preferences not to delete them.

According to our coding scheme of a participant's comment (see the first paragraph of section 4), a comment may suggest to keep, delete, or offer other suggestions. A common *other* suggestions is to redirect the page to an existing article of the same topic or merge the two articles. Interestingly, the categories that had the highest percentages of other comments all belong to the main category Culture and the Arts. This suggests that articles about Culture and Arts may be more likely to have the situation that the same topic was written twice or more and they had to be merged or redirected, than the other articles.

5 The Mapping of Article Categories and Mentioned Policies

We developed an interactive visualization to offer an overall mapping between the articles' categories and the policies mentioned in their AfD discussions (http://www.mandanemedia.com/afd/view/diagram5.php). This mapping between the categories and the policies was done at the second layer of policy and category hierarchy. Both layers provide an overview of the policy and category respectively, yet enough details are shown as well. When one hovers over the name of a category on the visualization, the policies mentioned in the AfD discussions within this category are highlighted on the right side, as well as the frequency distribution of these policies. For example, Figure 4 shows that 6% our dataset's AfDs are music articles. In these AfD discussions, some comments mentioned policies. Specifically, 8% of the mentioned policies are Notability (people), 31% are Notability, 43% are Notability (music), 10% are Identifying reliable sources, and 8% are Deletion_process. One can open the Wikipedia page about a specific policy by clicking on the name of the policy from this visualization.

6 Discussion

Wikipedia's Article for Deletion (AfD) discussions are critical for maintaining and monitoring the quality of Wikipedia articles. A dataset suitable for computational analysis of the AfD content is much desired given the large amount of unstructured AfD discussion content available on the web. To respond to this research need, we developed an open database that parsed and stored AfDs' discussion content in a relational database for further indexing and processing. Many important but unexplored AfD content analysis can be studied with this open access dataset. For example, [21] suggested that the presence of other votes has influence on the act of saving the articles. But the authors only analyzed 229 AfD discussions. With this dataset, one can explore this relationship at a much larger scale.

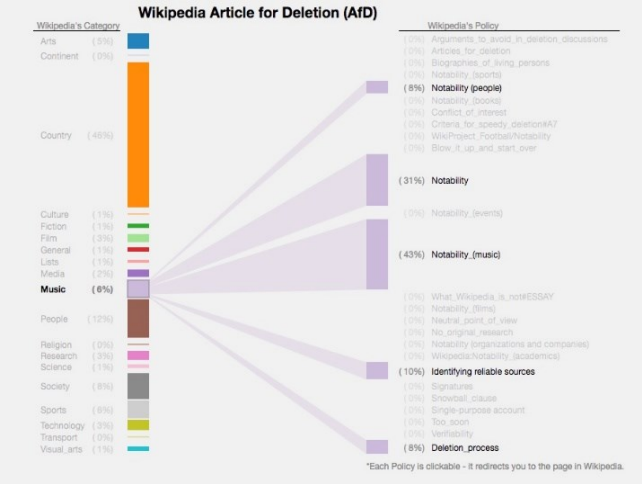


Figure 4: The mapping between the AfDs' categories and the policies mentioned in their discussions; the screenshot shows that music category is hovered above

A Wikipedia article could be nominated multiple times for AfD discussions, and the outcomes can be different from these discussions. In the dataset, about 1000 articles were nominated more than once and went through the AfD discussion process. This presents another interesting aspect to explore - to investigate these articles to understand the rea-sons behind multiple nominations and to compare the different discussions of the same article. We also found that there are AfDs that had delete out-comes but the articles still exist. It is possible that these articles were created again after they were deleted, and they were not proposed for deletion again or not yet. With this dataset, one can explore whether these articles were edited by the same person/people, the reasons of the delete outcomes, and the motivations of writing the same article again after it was deleted by the community.

With this dataset, we developed several interactive visualizations to explore the relationships between the outcomes of the AfD discussions and the number of keep/delete comments in the discussions and to obtain an overview of the policies mentioned in the discussions and the categories of the articles that were proposed for AfD. While some of these aspects were studied in the previous studies (e.g., [21]), we used a much larger dataset (about 40, 000 discussions). The benefits of our interactive visualizations are three folds. First, they present a way to visualize the Wikipedia's community memory in deletion

discussions. Second, these visualizations offer interesting insights on AfD discussions as a research exploration tool for Wikipedia AfD researchers. For example, in the first visualization, the cell in the fifth row and seventeenth column is red, which means that among the AfD discussions that had 17 keep votes and only 5 delete votes, majority of the final decisions were still to delete the article. To find out what happened in these discussions, one can go back to the dataset for further analysis. Third, there is a potential that these visualizations benefit Wikipedia community, especially to the new editors. Previous studies suggest that new editors need tremendous help in realizing the relevant policies and guidelines when editing an article or participating in an article's deletion discussion [6, 14, 15]. In Wikipedia discussions, participants can refer to policies to justify their viewpoints, because policies are supposed to be the shared understandings of how things should be done in the community. The confusions occur among the participants especially the new editors because such common ground is not established with them. By giving an overview and pointers of the mentioned policies in the discussions, new editors can be better aware of the existence of these shared understandings and get acquainted with them more quickly. For example, with the mapping visualization, new author(s) get an understanding of the Wikipedia policies and guidelines they should be familiar with given the type of the current article they are editing. If they need more details about the policies, they can navigate to those Wikipedia pages conveniently from the visualization.

One visualization analysis had an implicit assumption that the policies have stayed the same in the two years of discussions. Wikipedia's policies are the essential parts of AfDs, and the community changes, revises or creates policies in an ongoing basis. So when and how the policies were mentioned in the discussions might be affected by these changes. One future task is to visualize the changes of the policy articles and examine the correlation relationship between the use of the policies and the versions of the policies.

While our AfD dataset offers close to 40, 000 organized AfD discussions, the size is still very small given the population of these discussions. With the database being open access, we encourage researchers to put more AfD discussions in the current dataset and share them in the community.

7 Conclusion

Wikipedia's article for deletion (AfD) discussions offer a communication space for the community members to deliberate about whether or not an article should be deleted from the site. The study of these discussions offers insights on the quality control mechanisms and processes in online peer production systems. To encourage the use of natural language processing and other computational techniques to study the discussion content, we developed an open SQL database with 39,177 discussions. In this paper, we introduced this dataset including the raw data source, the challenges we faced in automatic processing and organizing the data and our strategies to address them. The purpose is to give a comprehensive view of this dataset for those researchers interested in using it in their computational analyses. We call for research studies about the AfD discussions that leverage this open database and suggested several research ideas to explore. As an example, we developed interactive visualizations to gain an overall understanding on the relationship between the deletion outcome and the type of discussion comments, on the Wikipedia policies mentioned in the discussion, and on the categories of articles this dataset include. With these interactive visualizations being open access as well, we hope they become useful resources for Wikipedia community members and for interested researchers.

8 ACKNOWLEDGEMENTS

We gratefully acknowledge the support by the second author's Discovery grant from Natural Science of The Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[1] Arazy, O, Yeo, L., Nov, O.: Stay on the Wikipedia Task: When task - related Disagreements Slip into Personal and Procedural Conflicts. Journal of the Association for Information Science and Technology, 64(8), 1634-1648 (2013)
[2] Biuk-Aghai, R. P., & Cheang, F. H. H. 2011. Wikipedia category visualization using radial layout. In Proc. of the 7th International Symposium on Wikis and Open Collaboration, 193-194
[3] Biuk-Aghai, R. P., Chan, R. C. K., Si, Y. W., & Fong, S. 2013. Visualizing recent changes in Wikipedia. Sci. Ch. Inf. Sci., 56(5), 1-15. 2013
[4] Boldi, P., &Monti, C. 2016. Cleansing Wikipedia Categories using Centrality. In Proc. of the 25th International Conference Companion on World Wide (WWW '16 Companion), 969-974.
[5] Ehmann, K., Large, A., Beheshti, J.: Collaboration in context: Comparing article evolution among subject disciplines in Wikipedia. First Monday, 13(10) (2008)
[6] Geiger, R. S., & Ford, H. 2011. Participation in Wikipedia's article deletion processes. In Proc. of the 7th International Symposium on Wikis and Open Collaboration, 201-202
[7] Iandoli L., Quinto I., Liddo, A. D., & Shum, S. B. 2014. Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. Int. J. Hum. Comput. Stud., 72 (3), 298-319.
[8] Iosub, D., Laniado, D., Castillo, C., Morell, M. F., Kaltenbruner, A.: Emotions under discussion: Gender, status and communication in online collaboration. PloS one, 9(8), e104880 (2014).
[9] Joyce, E., Pike, J., Butler, B.: Keeping eyes on the prize: Officially sanctioned rule breaking in mass collaboration systems. In Int. Proc. of Conference on Computer Supported Cooperative Work (CSCW). 1081–1092, ACM, San Antonio, USA (2013).
[10] Lam, S. T. K., & Riedl, J. 2009. Is Wikipedia growing a longer tail? In Proc. of the ACM 2009 International Conference on Supporting group work, 105-114
[11] Lam, S. K., Karim, J., & Riedl, J. 2010. The effects of group composition on decision quality in a social production community. In Proc. of the 16th ACM International Conference on Supporting Group Work, 55-64
[12] Reagle Jr, J. M.: Do as I do:: authorial leadership in Wikipedia. In: Int. Proc. of Symposium on Wikis, 143-156, ACM, Montreal, Canada (2007)
[13] Rezgui, A., Crowston, K.: Stigmergic coordination in Wikipedia. In: Int. Proc. of OpenSym, (August 22–24, 2018, Paris, France), Retrieved from http://www.opensym.org/wp-content/uploads/2018/07/OpenSym2018_paper_34.pdf (2018)
[14] Schneider, J., Passant, A., & Decker, S. 2012. Deletion discussions in Wikipedia: Decision factors and outcomes. In Proc. of the Eighth Annual International Symposium on Wikis and Open Collaboration, DOI=http://dx.doi.org/10.1145/2462932.2462955
[15] Schneider, J., Samp, K., Passant, A., & Decker, S. 2013. Arguments about deletion: How experience improves the accept-ability of arguments in ad-hoc online task groups. In Proc. of the 2013 conference on Computer supported cooperative work, 1069-1080
[16] Shneiderman, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. IEEE Symposium on Visual Languages, 336-343
[17] Stefaner, M., Taraborelli, D., Ciampaglia, G. L. 2011. Notabilia - Visualizing Deletion Discussions on Wikipedia. http://notabilia.net
[18] Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. 2008. Information quality work organization in Wikipedia. J Am Soc Inf Sci Technol, 59(6), 983-1001.

[19] Taraborelli, D., & Ciampaglia, G. L. 2010. Beyond notability. Collective deliberation on content inclusion in Wikipedia. In Proc. of the 4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO), 122-125

[20] Wattenberg, M., Viégas, F. B., & Hollenbach, K. (2007). Visualizing activity on Wikipedia with Chromograms. In Proc. of the 11th IFIP TC 13 international conference on Human-computer interaction (INTERACT'07), 272-287

[21] Xiao, L., & Askin, N. 2014. What influences online deliberation? A Wikipedia study. J Am Soc Inf Sci Technol, 65(5), 898-910.

[22] Xiao, L., & Nickerson, J. 2019. Imperatives In Past Online Discussions: Another Helpful Source For Community Newcomers? Proceedings of the 52nd Annual Hawaii International Conference on System Sciences (Jan. 8 – 11, Grand Wailea, Maui, USA), IEEE