

Endurance against oblivion: The case of the Articles for Deletion with gender perspective in Wikipedia

David Ramírez-Ordóñez
Universitat Oberta de Catalunya

Núria Ferran-Ferrer
Universitat de Barcelona

Abstract

This document presents an analysis with a gender perspective in the extension of the debates of the Articles for Deletion category in the English Wikipedia. The results show that biographies of women require much more deliberation time, comments and votes than biographies of men. This can be explained by the lack of clarity in how the concept of notability is applied to women's biographies, taking into account the difficulty of finding sources of information that talk about women. Wikipedia is just the mirror of the public sphere.

Keywords: Gender gap, Gender bias, Articles for Deletion, Biographies by gender, Notability

Introduction

The Articles for Deletion category on the English Wikipedia allows articles to be evaluated in a Wikipedian discussion for at least seven days (Cat, 2019). The decision to keep or delete an article is made based on a consensus among the votes, which can generally be in favor of deleting (vote: delete) or in favor of keeping the article (vote: keep). In this paper we carry out an analysis with a gender perspective of the sequences of the votes and the length of the debates.

Methods

We use the content analysis technique (Krippendorff, 2004) to obtain the sequences of the votes in the debates, that is, the number of participants, votes and arguments are collected and analyzed with a qualitative and quantitative approach. In this paper we present the results of quantitative analysis through sequence analysis using the TraMineR sequence mining and visualization package (Gabadinho et al., 2011). Also, we use the quantitative method to compare the extent of the debates on the deletion of 115 biographies collected during the year 2020 before the declaration of the COVID-19 pandemic by the World Health Organization in March 11th (Organización Mundial de la Salud, 2020), of men and women of the field "scientist".

During the period studied, 45863 new articles were created on the English Wikipedia. Of these new articles and some already created, 3977 were nominated for the Deletion Consultation. Of these 3977, 1507 are biographies. Of the 1507 biographies, only 115 were of scientists. These 115 biographies of scientific people are the ones we will use in this study.

Votes can have multiple values such as keep, delete, no consensus, redirect, and merge. There can also be variations like speedy delete, weak delete or soft delete. We unified all these variations, so weak delete, soft delete, and speedy delete are all considered as delete. And since for practical purposes no consensus, merge and redirect don't remove content, we made them equivalent to keep.

The approach of our study is binary, since in the selected sample only two non-binary people appeared, but not from the field "scientist" and therefore they had to be excluded.

Results

Following the two mentioned methods, we obtained two types of results: sequences and numbers of votes.

Sequences

In this paper we understand the sequences as the series of votes ordered in time. That is, in a debate in the Articles for Deletion category with 8 votes, the vote made at the beginning of the discussion will be the vote in the first position. The last vote in the discussion will be the one in position 8. Since the discussions have different lengths, in this figure 1 you can see the distribution of the sequences.

Among the essays explaining the policies and guidelines of the English Wikipedia is the Snowball Clause (Wik, 2021). This essay is cited by the editors who participate in the discussions of our study. It is used when the sequences of the votes do not have many variations and the administrator who closes the debate interprets it as a consensus. The following state frequency diagram shows the distributions in percentages of votes (Y axis) according to their location in the sequence of votes (X axis). See figure 2.

If we analyze the sequences of the deliberations according to the gender of the biographed person, we find that in the case of women, position 15 is maintained as

the place of unification of the votes. In the case of men, it is in a much earlier position, number 8. See figure 3.

In both genres, the more the deliberation is extended or lengthened, that is, the more comments and/or more votes are generated, it implies that the result of the deliberation will generate unanimity in favor of keeping the biographies debated in the Articles for Deletion deliberation. On the contrary, if the extension of the deliberation is short, it indicates an imminent deletion.

Figure 4 shows the probabilities that the votes will be maintained or change status, according to the previous vote. This implies that the votes are much more likely to stick than to have a change in sequence.

Number of votes in the debate

The averages according to the number of votes or comments in a debate show that, on average, there are more votes in the debates of women's biographies (6.971) than in the debates of men's biographies (4.630). There is a statistical significant difference (Welch $p < 0.032$).

Discussion/Conclusions

Discussion

When analyzing the data from this research on nominations to the Articles for Deletion category, we obtain the same results obtained by Francesca Tripodi, which suggest that the number of nominations for men is lower than the number of nominations for women. Also, that the biographies miscategorized as non-notable and deleted are mostly those of women (Tripodi, 2021).

This research explored the vote sequence analysis conducted in 2010 (Taraborelli and Ciampaglia, 2010). We disagree that the nomination should be counted as the first vote in the sequence, since all nominations always start with a delete vote. However, we agree that the herd effect holds: the probabilities show us that changing a sequence of votes from keep to delete or vice versa is much less likely than keeping the same voting value.

Conclusions

- To save the biographies of both men and women, it seems that it is necessary to extend the debates, which tends to increase the probability of having votes in favor of maintaining the biographies.
- Extending the debates requires resources, especially temporary ones, that women editors do not have, due to the small number of editors with free time (Ferran-Ferrer et al., 2021)
- The title of the Articles for Deletion category already implicitly suggests a voting result. To encourage discussions to be neutral, the title should be as well.

References

- [Cat2019] 2019. Category:Articles for deletion. *Wikipedia*, August.
- [Ferran-Ferrer et al.2021] Núria Ferran-Ferrer, Patricia Castellanos-Pineda, Julià Minguillón, and Julio Mene-ses. 2021. The gender gap on the Spanish Wikipedia: Listening to the voices of women editors. *Profesional de la Información*, 30(5), October.
- [Gabadinho et al.2011] Alexis Gabadinho, Gilbert Ritschard, Nicolas Séverin Mueller, and Matthias Studer. 2011. Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software*, 40(4):1.
- [Krippendorff2004] Klaus H. Krippendorff. 2004. *Content Analysis: An Introduction to Its Methodology*. Sage Publications, London, second edition.
- [Organización Mundial de la Salud2020] Organización Mundial de la Salud. 2020. Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19 celebrada el 11 de marzo de 2020. <https://www.who.int/es/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>.
- [Taraborelli and Ciampaglia2010] D. Taraborelli and G.L. Ciampaglia. 2010. Beyond notability. Collective deliberation on content inclusion in Wikipedia. In *Proceedings - 2010 4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshop, SASOW 2010*, pages 122–125.
- [Tripodi2021] Francesca Tripodi. 2021. Ms. Categorized: Gender, notability, and inequality on Wikipedia. *New Media & Society*, page 146144482110237, June.
- [Wik2021] 2021. Wikipedia:Snowball clause. *Wikipedia*, September.

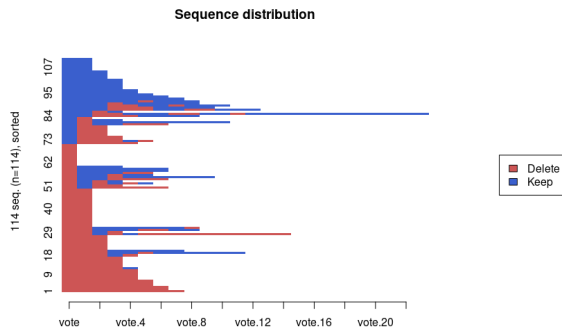


Figure 1: Sequence distribution. Own work.

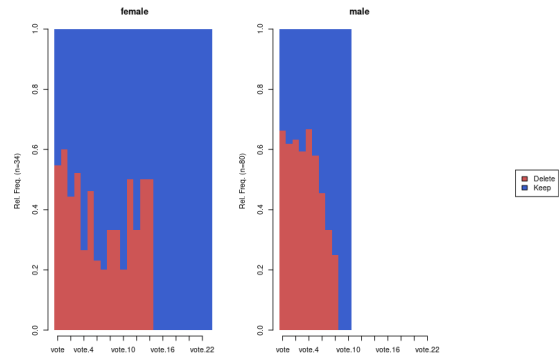


Figure 3: State frequencies by gender. Own work.

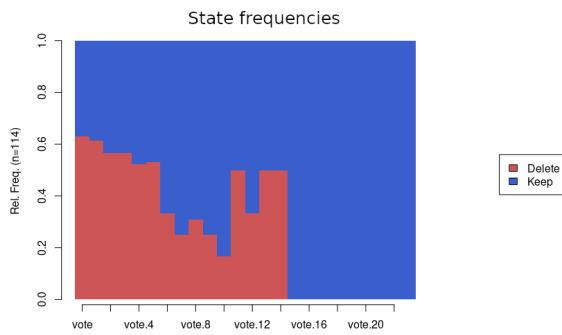


Figure 2: State frequencies. Own work.

	[-> Delete]	[-> Keep]
[Delete ->]	0.79	0.21
[Keep ->]	0.22	0.78

Figure 4: Probability of the next vote due the previous vote. Own work.