Cover Women

Núria Ferran-Ferrer
Faculty of Information and Communication
Universitat de Barcelona (UB)

Miquel Centelles
Faculty of Information and Communication
Universitat de Barcelona (UB)

Abstract

This project will look into the most popular Wikipedia page, which got 46.8 billion visits last November (WikiStats, 2023), on the English edition. This page, known as the main page, or front page from a communication perspective, will be analysed across the seven longest-standing Wikipedia editions: English, German, Catalan, French, Portuguese, Italian, and Spanish.

Grounded in a gender and intersectional perspective, this study will delve into the daily content, newsroom guidelines (principles and standards that guide the dissemination of information), and volunteer community insights.

The examination will employ communication theories like gatekeeping (Barzilai-Nahon, 2009) and agenda-setting (McCombs and Shaw, 1972).

Beyond academic research, our goal is to actively contribute to editing communities by addressing the daily challenges and needs in crafting front-page content.

Introduction

Despite Wikipedia being a key player in the public sphere and having a transformative impact on information dissemination, Wikipedia grapples with persistent gender bias in both editing and content (Antin et al., 2011; Bear & Collier, 2016; Wagner et al., 2016; Hinnosaar, 2019; Minguillon, et al., 2021) alongside additional prejudices (Redi et al., 2020, Beytía et al., 2022).

Scholars highlight the need for a comprehensive understanding of Wikipedia's knowledge production culture to address these biases and make Wikipedia more robust, reliable, and transparent (Menking & Erikson, 2015).

Reducing the gender and other intersectional gaps necessitates more than acknowledging Wikipedia as a mirror of societal biases—it involves addressing the platform's deeper logic embedded in its techno-scientific project (Ford & Wajcman, 2017).

Related Work

We have already tested the proposal by analysing the gender identity of individuals on the front pages of English and Spanish Wikipedia in the past decade. Only 28% were women in the English edition, and it was even lower in the Spanish edition at 18%, with other marginalised gender identities below 1%. And the trend is not encouraging (Ferran-Ferrer, et al., 2024).

Moreover, bias in contributions perpetuates imbalances in content coverage and discourages diversity, which further exacerbates the issue (Worku et al., 2020).

Methods

We will employ a triangulation methodology.

1. We will scrutinize the daily content selection on the front-page using OpenRefine, enhancing data with specific Wikidata properties like P21 (sex or gender), P106 (occupation), P172 (ethnic group), and others.
2. We will apply content analysis to approximately front-page editorial guidelines (newsroom guides) for each language edition, and we will explore what shapes the decision-making of the gatekeepers who determine story prominence.
3. We will conduct personal in-depth interviews with voluntary editors of the front page from all seven Wikipedia editions to ascertain decision-making processes, biases, and strategies that influence content visibility related to gender and other intersectionalities.

This research is necessary to draw further attention to the need for systemic change within the platform's newsroom/editorial practices to address disparities in gender representation in online knowledge and foster a more inclusive and diverse digital information landscape.

Date: June 1, 2024 / June 30, 2025.

Expected output

We will provide outstanding research papers and communications at conferences.

Risks

We validated the methodology with two Wikipedia editions and will expand to five more. Our proficiency in diverse data archiving and multilingual interviews is established from prior work.

Community impact plan

The project aligns with the Wikimedia Movement's 2030 strategy by focusing on delivering knowledge as a service and addressing equity in knowledge and communities overlooked by structures of power and privilege.

We will provide:

- Decade-long insights into gender and intersectionality representation on Wikipedia.
- Beyond descriptive stats, we'll reveal bias trends.
- Proposing editorial strategies for gatekeeping and agenda setting.
- Provide guidelines for ethical content selection using AI and bots.
- Provide technical guidance to enhance data archives on main pages.
- Collaborative work with volunteers ensures inclusivity, integrating advocate perspectives for a consensus-driven approach.

Built on in-depth interviews and collaboration with Wikipedia chapters and user groups, this work addresses gender identity under-representation, as reflected in the Cover Women team's table.

Evaluation

As university professors, we are assessed by traditional impact such as metrics for academic journals. But we will consider alternative metrics too, sharing out information through guidelines and recommendations reported to Wikimedia volunteers and boards, measuring impact through the production and acceptation of them.

Budget

- PhD student for 1 year 25,000
- Equipment: server for big data 6 months: 1,500
- Conference travel, registration, and accommodation: 2,000
- Open Access Publishing Costs (APCs): 3,500
- Institutional overhead: 15%

Prior contributions

Núria Ferran-Ferrer is an accomplished associate professor at Universitat de Barcelona (UB), where she directs the PhD program on Information and Communication. She is deputy of the rector for the direction of the EDI office and a renowned Wikipedia researcher in gender. Núria's expertise in the field of information, communication and gender has earned her a position on the Amical Wikimedia Board. She is also a valued contributor for Wikimedia Foundation Community Grants in North and West Europe.

Miquel Centelles is a distinguished lecturer at Universitat de Barcelona (UB) where he directs the Master's Degree in Digital Humanities. He is a research collaborator at Women and Wikipedia project, and his contributions have been instrumental in advancing the cause of women in the field of digital humanities. Miquel's expertise in the field of semantic web is highly regarded and has earned him a reputation as a thought leader in the discipline.

Recent publications of our team on gender and Wikipedia are:

Centelles, M., & Ferran-Ferrer, Núria. (2024). Assessing Knowledge Organization Systems from a gender perspective: Wikipedia Taxonomy and Wikidata Ontologies. [Forthcoming].

Ferran-Ferrer, N., Castellanos-Pineda, P., Minguillón, J., & Meneses, J. (2021). The gender gap on the Spanish Wikipedia: listening to the voices of women editors. *El Profesional de la Informacion*, 30(5), e300516.

Ferran-Ferrer, N., Boté-Vericad, J.-J., & Minguillón, J. (2024). Tracing the gender gap on Wikipedia: a scoping review of scholarly research from 2005-2022. *El Profesional de la Información*, 32(6).

Minguillón, J., Meneses, J., Aibar, E., Ferran-Ferrer, N., & Fàbregues, S. (2021). Exploring the gender gap in the Spanish Wikipedia: differences in engagement and editing practices. *PLOS ONE*, 16(2), e0246702.

References

Barzilai-Nahon, K. (2009). Gatekeeping: a critical review. *Annual Review of Information Science and Technology*, 43(1), 1–79. https://doi.org/10.1002/aris.2009.1440430117

Bear, J. B., & Collier, B. (2016). Where are the Women in Wikipedia? Understanding the Different Psychological Experiences of Men and Women in Wikipedia. Sex Roles, 74(5), 254–265. https://doi.org/10.1007/s11199-015-0573-y

Beytía, P., Agarwal, P., Redi, M., & Singh, V. K. (2022). Visual gender biases in wikipedia: a systematic evaluation across the ten most spoken languages. *Proceedings of the International AAAI Conference on Web and Social Media, 16*, 43–54. https://ojs.aaai.org/index.php/ICWSM/article/view/19271

Ferran-Ferrer, N., Fernández, L., & Centelles, M. (2024). Wikipedia's Front page 10 years evolution: analysis of gender and intersectionalities on content, newsroom

guidelines and volunteer-driven insights in the Spanish and English editions. [Forthcoming].

Ford, H., & Wajcman, J. (2017). Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. Social Studies of Science, 47(4), 511–527. https://doi.org/10.1177/0306312717692172

Hinnosaar, M. (2019). Gender inequality in new media: Evidence from Wikipedia. Journal of Economic Behavior & Organization, 163, 262–276. https://doi.org/10.1016/j.jebo.2019.04.020

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, 36(2), 176–187.

Menking, A., & Rosenberg, J. (2021). WP:NOT, WP:NPOV, and Other Stories Wikipedia Tells Us: A Feminist Critique of Wikipedia's Epistemology. Science Technology and Human Values, 46(3), 455–479. Scopus. https://doi.org/10.1177/0162243920924783

Minguillón, J., Meneses, J., Aibar, E., Ferran-Ferrer, N., & Fàbregues, S. (2021). Exploring the gender gap in the Spanish Wikipedia: differences in engagement and editing practices. PLOS ONE, 16(2), e0246702. https://doi.org/10.1371/journal.pone.0246702

Redi, M., Gerlach, M., Johnson, I., Morgan, J., & Zia, L. (2021). *A taxonomy of knowledge gaps for Wikimedia projects (Second Draft)* (arXiv:2008.12314). arXiv. http://arxiv.org/abs/2008.12314

Wagner, C., Graells-Garrido, E., Garcia, D., & Menczer, F. (2016). Women through the glass ceiling: Gender asymmetries in Wikipedia. EPJ Data Science, 5(1), Article 1. https://doi.org/10.1140/epjds/s13688-016-0066-4

Wikimedia. (2023). *Wikimedia statistics: all wikis*. https://stats.wikimedia.org/#/all-projects

Worku, Z., Bipat, T., McDonald, D. W., & Zachry, M. (2020). Exploring Systematic Bias through Article Deletions on Wikipedia from a Behavioral Perspective. ACM International Conference Proceeding Series. Scopus. https://doi.org/10.1145/3412569.3412573