

Local Content Matters:

Insights on Wikipedia Editor and Reader Engagement¹

Marc Miquel-Ribé
Tecnocampus - Universitat
Pompeu Fabra
Barcelona, Catalonia
mmiquelr@tecnocampus.cat

David Laniado
Eurecat, Centre Tecnològic de
Catalunya
Barcelona, Catalonia
david.laniado@eurecat.org

Andreas Kaltenbrunner
ISI Foundation; Universitat
Pompeu Fabra.
Barcelona, Catalonia
kaltenbrunner@gmail.com

ABSTRACT

About a quarter of each Wikipedia language edition is dedicated to representing “local content”, i.e. the corresponding cultural context (geographical places, historical events, political figures, among others). To investigate the relevance of such content for users and communities, we present an analysis of reader and editor engagement in terms of pageviews and edits. The results, consistent across 15 diverse language editions, show that these articles are more engaging for readers, and especially for editors. Regarding anonymous editors, we see a higher proportion of their edits devoted to it. These findings indicate the relevance of this kind of content both for fulfilling reader’ information needs and for stimulating the dynamics of the editing community.

CCS CONCEPTS

• Human-centered computing → Empirical studies in collaborative and social computing → Human-Computer Interaction (HCI) → Collaborative content creation

KEYWORDS

Wikipedia, online communities, user engagement, cultural context, readership, online collaboration

1 INTRODUCTION

Even though Wikipedia is not a newspaper, editors aim at covering the readers’ evolving informational needs. Wikipedia’s coverage of news and current events drives editor activity and reader attention any given week [2]. Collaborations to create these articles involve more editors and

happen at a higher speed than any other type of content. However, the most read articles do not necessarily correspond to those frequently edited, suggesting some degree of non-alignment between user reading preferences and author editing preferences [3]. Editor topical preferences depend on factors such as their domain of expertise [1], political identity [7], among others.

Wikipedia also reflects the context in which editors live and edit. To assess the extent of content representing the languages’ geographical and cultural context in each Wikipedia, often described as “local content” by editors, a method was proposed in [4] to collect all the articles that relate to the language, people and territories where a specific language is spoken, dubbing it Cultural Context Content (CCC) [5,6]. Characterizing how CCC is created and consumed can help to understand both the dynamics in the Wikipedia communities and the role of Wikipedia in society. Therefore, we want to answer the research question: Does cultural context content reflect a higher level of editor and reader engagement? To this aim, we analyze all the edits through which editors created the articles, and the pageviews they received during a period of six months.

2 DATASETS

For the identification of the content that can be considered “local” to each language, we relied on the dataset from a previous study, based on 15 language editions [4]. In line with related studies [5,6], in our dataset the percentage of CCC for the top forty languages in the number of articles is 23.1%, while for the entire set of 304 languages it is 12.5%.

Editing history and pageviews. We retrieved the number of pageviews articles received from January to June 2016 and the edit history available in the Wikimedia dumps² accounting all the edits in articles until June 2016.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
The Web Conference, April, 2021, Slovenia

¹ This paper is an **extended abstract** of the following paper:

Miquel-Ribé, M., Laniado, D., & Kaltenbrunner, A. (2021). The Role of Local Content in Wikipedia: A Study on Reader and Editor Engagement. *Área abierta*, (21).

² <https://dumps.wikimedia.org/>

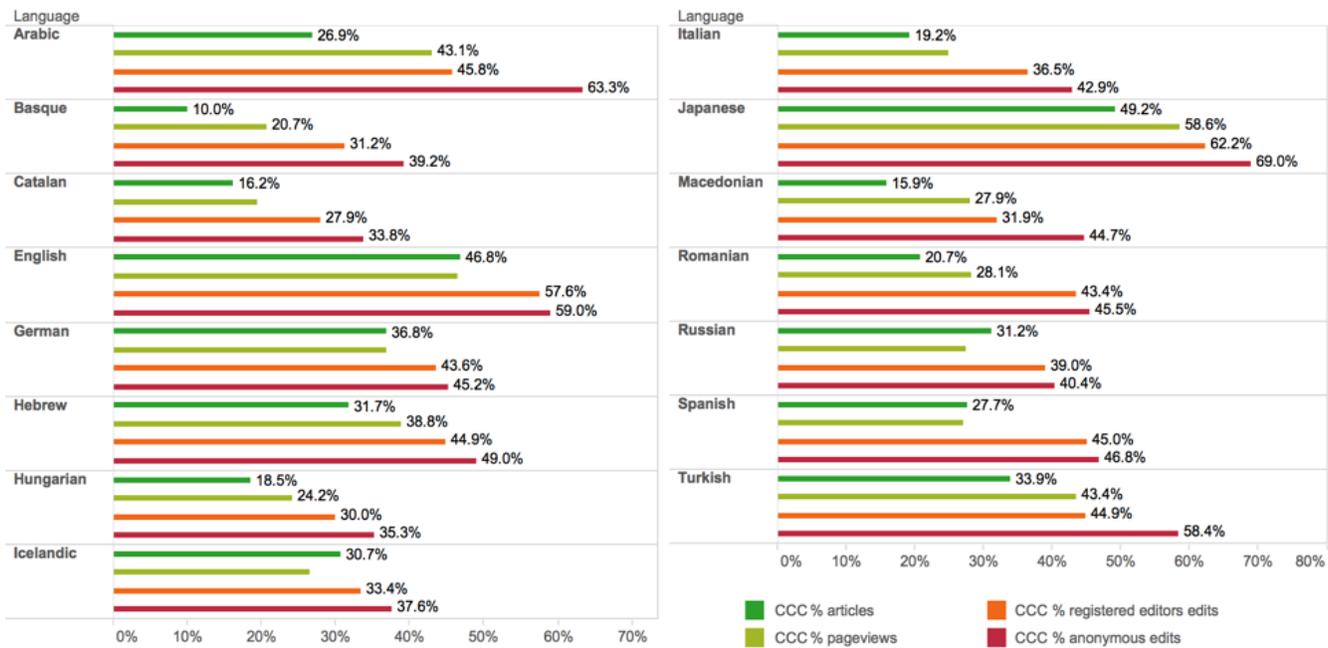


Figure 1: The relative weight of CCC in number of articles, number of pageviews, and number of registered editor edits and anonymous edits. Local content receives a greater attention than the rest of content by readers, and even more so by editors.

3 FINDINGS

We quantified the engagement around CCC in terms of number of pageviews and edits. As commonly done in previous research, we excluded bot edits. Figure 1 shows the relative weight of CCC in each Wikipedia language edition in terms of the number of articles, pageviews, registered editors' edits and anonymous edits. One can notice a crescendo according to this order in almost all cases, indicating that CCC as a whole receives overall a greater attention than the rest of the content by readers, and even more so by editors. It is important to remark that anonymous editors devote almost half of their attention to CCC (mean 47.3%, median 45.2%, and standard deviation 10.5). To verify these results at the article level, we compared CCC to the rest of Wikipedia articles, both in terms of edits (anonymous and registered editors' edits) and pageviews. We performed a Mann-Whitney test to validate the hypothesis that in general there would be more edits and pageviews in CCC articles than in the rest of Wikipedia articles. The hypotheses are confirmed (with p-values < 0.005) for both edits and pageviews and for all languages.

4 CONCLUSIONS

The higher engagement of editors with CCC points out a preference towards representing their context that seems to go

beyond a possible response to readers' information demand. This preference is particularly high for anonymous editors, which we assume to be mainly spontaneous editors, and tend to engage more in editing topics related to their nearby environment. We hypothesize that these editors' interest in the Wikipedia project is more likely to be ignited by CCC content. Based on these findings, we conclude that cultural contextualization is not a marginal or anecdotal phenomenon. Rather than being an undesired bias, it should be seen as relevant for both expanding the diversity of content and attracting new editors.

REFERENCES

- [1] Halatchliyski, I., Moskaliuk, J., Kimmerle, J., and Cress, U. Who integrates the networks of knowledge in Wikipedia? *Int. Sym. Wikis*, (2010), 1.
- [2] Keegan, B., Gergle, D., and Contractor, N. Hot Off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events. *American Behavioral Scientist* 57, 5 (2013), 595–622.
- [3] Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., and Kaltenbrunner, A. Reader preferences and behavior on Wikipedia. *ACM Press* (2014), 88–97.
- [4] Miquel-Ribé, M. and Laniado, D. Cultural Identities in Wikipedias. *ACM* (2016), 24–10.
- [5] Miquel-Ribé, M. and Laniado, D. Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions. *Frontiers in Physics* 6, (2018), 234.
- [6] Miquel-Ribé, M. and Laniado, D. Wikipedia Cultural Diversity Dataset - A Complete Cartography for 300 Language Editions. *ICWSM*, (2019).
- [7] Neff, J.J., Laniado, D., Kappler, K.E., Volkovich, Y., Aragón, P., and Kaltenbrunner, A. Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia. *PLoS one* 8, 4 (2013), 60584