

A Platform for Visually Exploring the Development of Wikipedia Articles

Erik Borra^{*}, David Laniado[‡], Esther Weltevrede^{*}, Michele Mauri[§], Giovanni Magni[§], Tommaso Venturini[¶], Paolo Ciuccarelli[§], Richard Rogers^{*} and Andreas Kaltenbrunner[‡]
^{*} University of Amsterdam, [‡] Barcelona Media, [§] Politecnico di Milano, [¶] Sciences Po

Abstract

When looking for information on Wikipedia, Internet users generally just read the latest version of an article. However, in its back-end there is much more: associated to each article are the edit history and talk pages, which together entail its full evolution. These spaces can typically reach thousands of contributions, and it is not trivial to make sense of them by manual inspection. This issue also affects Wikipedians, especially the less experienced ones, and constitutes a barrier for new editor engagement and retention. To address these limitations, Contropedia offers its users unprecedented access to the development of an article, using wiki links as focal points.

Introduction

Wikipedia articles are not static documents, as an inattentive reader might think, but living artifacts that are liable to be modified over and over again. They often become the battleground for competing points of view that need to be turned into a *neutral*¹ description gathering community consensus.

In line with its policy of transparency, Wikipedia makes the complete edit history and discussion pages of every article publicly available. However, these logs can be huge, displaying only the meta information of thousands of revisions in chronological order, and are disconnected from the discussions on the corresponding talk pages. Identifying disputes and negotiations that led to the current version of an article may thus require a significant effort, especially for big or controversial articles.

This is an issue both for readers who want to know more about an article’s development, and for editors who have to manually look through the edit history and talk pages to find out when and why something was changed. Inexperienced editors, or editors who are not familiar with an article, are especially affected as awareness of the previous history of an article, and of the negotiations that led to its current state, is paramount to successful contributions (Halfaker et al. 2012).

Contropedia² makes such valuable information easily accessible through a visual interface that allows one to identify the elements that aroused more dispute and activity, and to explore the development of a topic within an article.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

¹See <http://enwp.org/Wikipedia:NPOV>

²Available at: <http://contropedia.net>

Detecting controversies

Wikipedia guidelines³ indicate that all key concepts which can help the reader understand the article more fully should be linked to their own article. Repurposing this requirement, we use wiki links as focal points around which content and activity can be aggregated. Our approach is language independent as it relies only on parsing wiki markup and aggregating editor activity. Each edit is associated to the wiki links that appear in modified sentences; if more than one link is involved in the change, the contribution of the edit is split among all the wiki links involved. To focus on *substantial* and *disagreeing* edits, we exclude vandalism edits and contributions that only insert content. More details about the algorithm can be found in (Borra et al. 2015).

The visual interface

Two views allow for the exploration of an article. We will illustrate them taking the article “Abortion” as an example.

In order to easily locate controversial elements visually, the *layer view*, shown in Figure 1, adds a visual layer on top of an article; the hotter the color, the more controversial the corresponding elements. The view also contains a mini-fied version of the entire article, as a scroll-bar, to provide a quick overview of the sections which are the loci of most negotiations.

³<http://enwp.org/Wikipedia:Linking>

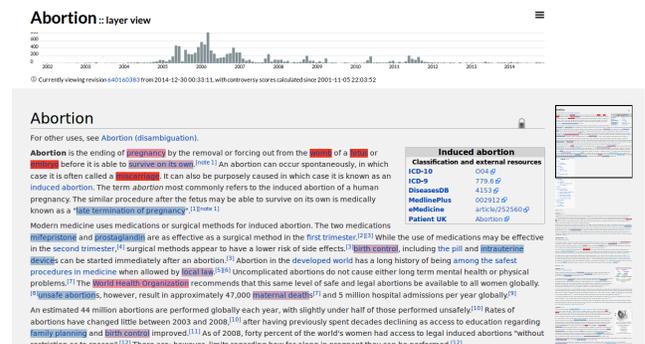


Figure 1: *Layer view* of the article “Abortion”.

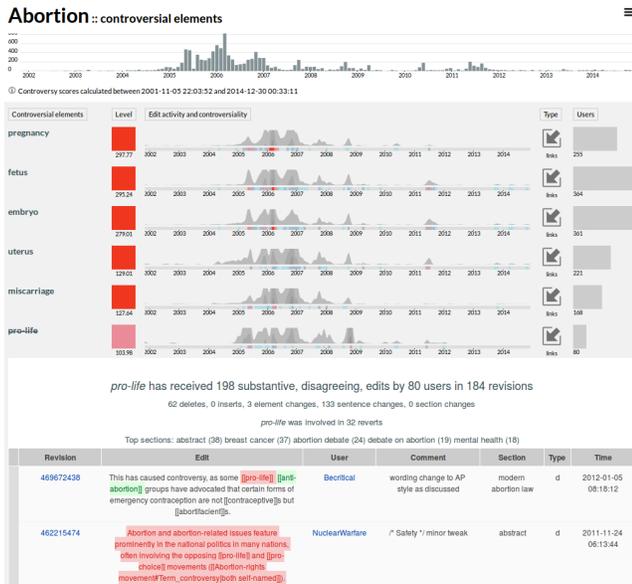


Figure 2: *Dashboard* view of the most disputed elements in the “Abortion” article, with a partial edit table of “pro-life”.

The *dashboard* view, shown in Figure 2, presents a ranking of the most controversial elements in an article. The gray bar next to each element displays edit activity over time. The thin colored bar below conveys when it was most disputed.

While in the layer view it is only possible to see the wiki links available in the latest version of the article, the dashboard view also displays those that have been deleted. Such elements are struck through, e.g. “pro-life” in Figure 2.

In both views, the timeline at the top of the page indicates the number of edits per month for the full edit history; via a menu it is possible to restrict the analysis to specific time periods. Both views also allow for a more fine grained analysis of the edits related to, as well as comments from the talk page mentioning, a specific element. Clicking an element displays the *edit table* which allows one to inspect all the edited sentences containing, and comments mentioning, the element. This way, edit activity and discussions related to a specific element can be analyzed together, even though they appear in two separate pages on Wikipedia. In the edit table, text is highlighted in green when it was inserted and in red if it was deleted in a specific edit. E.g., in Figure 2, the edit table for “pro-life” is shown; the first entry displays that the term “pro-life” was substituted by “anti-abortion”.

In the dashboard view, when one clicks on the number of editors that are involved in editing a specific element, a *network of disagreeing editors* is shown (Figure 3). When an editor makes an edit to a sentence which contains the element under analysis, an edge connects the editor to the previous editor of the sentence. The size of each node is proportional to the number of connections. One can thus see which editors are central to the development of the topic. Colors are assigned to convey camps of opposing editors, as described in (Jacomy 2014): contrary to usual community detection, but in line with our controversy measure, the algorithm as-

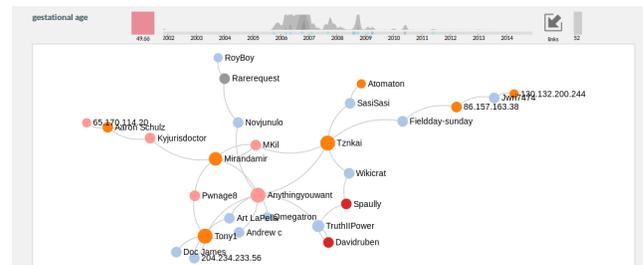


Figure 3: Network of users editing sentences that contain the link “gestational age”.

sumes that interactions represent disagreement and assigns the same color to editors who do not interact much with each other, but interact with the same other editors.

Discussion

Similar to our layer view, WikiTrust (Adler et al. 2008) offers an overlay on Wikipedia articles. WikiTrust, however, conveys the expected trustworthiness of content, while our tool highlights which topics were the locus of most dispute.

Although platforms for comparing different linguistic versions of Wikipedia exist (Massa and Scrinzi 2012; Bao et al. 2012), they only consider the current version of an article and do not dive into the edit history and talk pages in order to disclose the social processes behind article development. As our approach is language independent, our tool can be used for the exploration of articles in multiple languages.

Previous work on Wikipedia controversies has mainly focused on detecting which articles are controversial. Contropedia’s display of what is controversial *within* an article as well as when it was controversial, combined with the possibility to easily inspect the full edit and discussion activity around a specific topic, makes Contropedia useful for editors, interested readers, as well as researchers, and can lower the barriers to collaborate on Wikipedia.

References

- Adler, B. T.; Chatterjee, K.; De Alfaro, L.; Faella, M.; Pye, I.; and Raman, V. 2008. Assigning trust to Wikipedia content. In *Proc. WikiSym 2008*.
- Bao, P.; Hecht, B.; Carton, S.; Quaderi, M.; Horn, M.; and Gergle, D. 2012. Omnipedia: bridging the wikipedia language gap. In *Proc. CHI 2012*.
- Borra, E.; Weltevrede, E.; Ciuccarelli, P.; Kaltenbrunner, A.; Laniado, D.; Magni, G.; Mauri, M.; Rogers, R.; and Venturini, T. 2015. Societal Controversies in Wikipedia Articles. In *Proc. CHI 2012*.
- Halfaker, A.; Geiger, R. S.; Morgan, J. T.; and Riedl, J. 2012. The rise and decline of an open collaboration system: How Wikipedia’s reaction to popularity is causing its decline. *Am. Behav. Scientist*.
- Jacomy, A. 2014. Antagonism-based community detection. <http://github.com/jacomyal/ABCD-Gephi-Plugin>.
- Massa, P., and Scrinzi, F. 2012. Manypedia: Comparing language points of view of Wikipedia communities. In *Proc. WikiSym 2012*.

This research was supported by EU FP7 EINS grant #288021.