The bot that makes history

Journal Title
XX(X):1–8

DOI: 10.1177/ToBeAssigned
www.sagepub.com/

SAGE

Michael Falk*[1], Heather Ford[1], Tamson Pietsch[2], Nathaniel Tkacz[3]

Keywords
Wikipedia; bots; temporal regime; time; historicity; trace ethnography

Introduction: What is Wikipedia time?

Theoretical Overview: Wikipedia and the primacy of the past

How does Wikipedia portray the past? Scholars typically give three answers. Some argue that Wikipedia produces *history*: it represents the past in literary form. Wikipedia history may be more "colorful," "anecdotal" and "factualist" than "professional history," observes Roy Rosenzweig, but history it most certainly is (Rosenzweig, 2006, p. 142). Others argue that Wikipedia articles comprise *collective memories* that evoke shared experiences. From this perspective, Wikipedia's Talk pages are more important than the articles themselves, and its editors are more important than its readers. As Christian Pentzold argues, Wikipedia's Talk pages are non-physical "memory places," where editors meet to "negotiat[e]" the "memorable elements" of their experiences (Pentzold, 2009, p. 264). Numerous scholars have followed in Pentzold's wake to examine how editors "build" or "form" collective memories in Wikipedia (Ferron and Massa, 2011b,a; Porter et al., 2020). A third group of scholars argue that Wikipedia is a repository of *facts*. Wikipedia may well publish works of history and store collective memories, but its main role is to produce atomistic facts that are propagated through knowledge graphs (Ford, 2020, 2022). Wikipedia may be *memory* to thousands of editors. It may be *history* to millions of readers. But it is mere *fact* to billions of search requests and API calls.

[1]Digital and Social Media, University of Technology Sydney, Sydney, Australia
[2]Australian Centre for Public History, University of Technology Sydney, Sydney, Australia
[3]Centre for Interdisciplinary Methodologies, University of Warwick, Conventry, United Kingdom

Corresponding author:
Michael Falk, Digital and Social Media, University of Technology Sydney
Email: michael.falk@uts.edu.au

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

These approaches are not mutually exclusive. Search engines, readers and editors all produce and consume Wikipedia in different ways, and a complete account of the encyclopedia must include them all. In which case, we must ask: how are the historical, memorial and factual aspects of Wikipedia related?

One way to approach this question is to focus precisely on the *pastness* of history, memory and fact. Pastness is central to Wikipedia's self-definition. "Wikipedia is not a crystal ball," reads a famous policy, wherein we also read that Wikipedia is "not a newspaper." It is the pastness of Wikipedia that allows it to function simultaneously as history, memory and fact. Pastness is obviously a feature of both history and memory: I cannot remember an event nor write its history until it has happened. The pastness of *fact* is less obvious. Wikipedia contains facts about fictional spacecraft, embroidery techniques and the heat death of the universe. In what sense can such facts be said to be "past"?

Wikipedia itself provides an answer in two of its foundational policies. According to 'No Original Research', no new facts are to be admitted to the encyclopaedia. The only allowable facts are—the *old*. According to 'Neutral Point of View', no controversial facts are to be admitted to the encyclopaedia. The only allowable facts are—the *settled*. Facts are geological. Only time can grind down the seashells of evidence and bring forth the limestone of objectivity. Editors who wish to include new or unsettled facts in the encyclopaedia are advised that 'There is no deadline'. *Þæs oferēode; þisses swa mæg.* That passed; so may this. Eventually everything is past.

Despite its supposed pastness, Wikipedia is well-known as a source of information on current events. It is 'An Encylopedia with Breaking News' (Keegan, 2019). Current events dominate Wikipedia, accounting for the lion's share of user contributions and page views at any given time (Keegan et al., 2011). Scholars have analysed Wikipedia's coverage of current events in detail. We now know how Wikipedia's editors clash over the nature and definition of current events (Ford, 2022; Pentzold, 2009), how they link current events into larger thematic structures (Twyman et al., 2017), how they adopt newsroom practices to co-ordinate their efforts (Avieson, 2019), how they revisit old articles to commemorate traumatic events (Ferron and Massa, 2014), and how they shape the interpretation of events using images (Porter et al., 2020). One thing we *don't* know is how Wikipedia's editors decide what is 'current'. How does Wikipedia distinguish the past from the present at the very threshold of time? How does it resolve the contradiction between the pastness of the encyclopaedia and the presentness of the current?

Most readers of Wikipedia will have seen what editors do when an article trespasses on the present: mark it with one of the available Current Event Templates. The main template is Template:Current, which at the time of writing is available on 115 language editions of Wikipedia. When the template is added to an article, a familiar banner appears at the top of the page (Figure 1), and the article is automatically added to Category:Current Events or a related category. Each language edition has its own distinct version of Template:Current, and may also sport a range of related Templates. French Wikipedia, for instance, distinguishes 'Événements en cours' [ongoing events] from 'Événements récents' [recent events] in its main template, and provides several related templates such as Modèle:Bataille en cours [Template:Ongoing battle] and Modèle:Mort récente

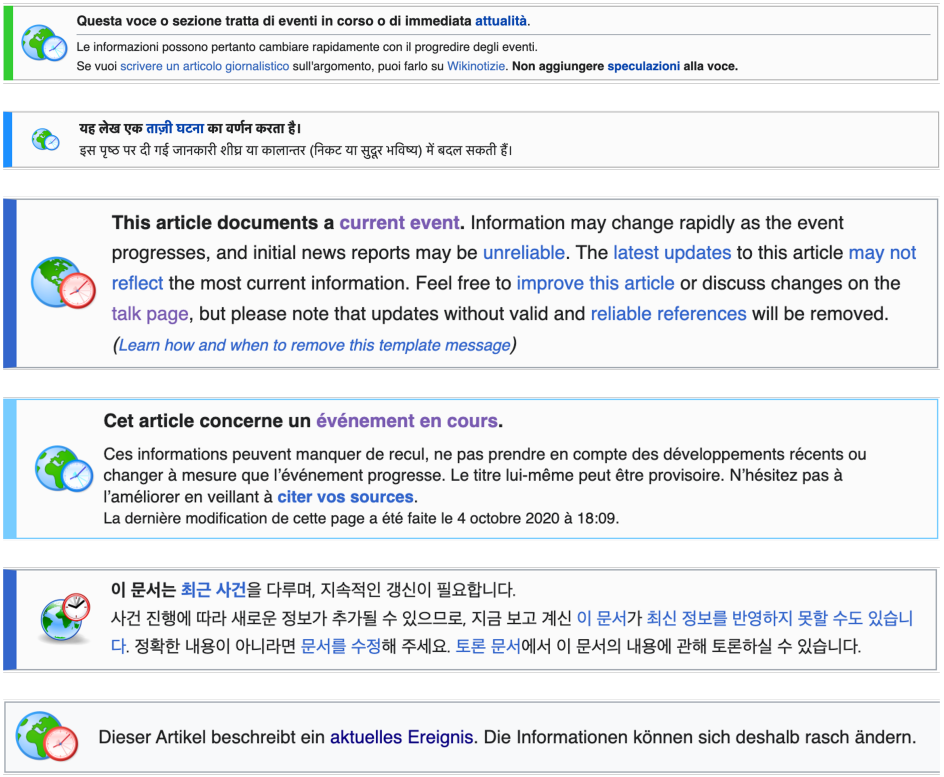


Figure 1. Template:Current in Italian, Hindi, English, French, Korean and German (as of 14 March 2023)

[Template:Recent death]. German Wikipedia, by contrast, has only a single Current Events template, but it is customisable, so that editors can replace the phrase "aktuelles Ereignis" [current event] in the banner with a more specific description such as "die derzeitige Sportveranstaltung" [the ongoing sporting event].

On the surface, Template:Current might seem like a simple phenomenon. Editors mark an article when it is "current," then remove the template when its currency is ended. Avieson (2019) likens Template:Current to the "live" icon on a newspaper blog or television news. While Template:Current is present, the article functions as live coverage; when the template is removed, the article gradually becomes encyclopaedic.

But the use of Template:Current is not simple, and has vexed Wikipedia's editors for years. Avieson (2019) herself grapples with the problem. Although she argues for a distinction between "news" and "encyclopaedias," she also observes that Wikipedia's coverage of current events "blurs the boundaries of both news and temporality." These blurred boundaries are a problem for Wikipedia's editors, and editors in different langauges have clarified the distinction between past, present and future in different ways.

In this context, English Wikipedia is extremist. Unlike other language editions, English Wikipedia strictly polices the use of Template:Current with a bot: Yapperbot/uncurrenter. Yapperbot/uncurrenter scans English Wikipedia hourly, examining every article that includes Template:Current and deleting the template if the article has not been edited in the last five hours. English Wikipedia is also one of the few larger language editions without a Template:Future to mark events that have yet to occur. English Wikipedia deleted Template:Future in 2009 after an official process, and several attempts to resurrect the template have foundered. Meanwhile French, Italian, Bengali, Chinese and 51 other-language Wikipedias maintain a Template:Future.

Why does practice vary across the different language editions? What led to the extremely strict approach of English Wikipedia, in which Template:Current is ruthlessly policed by an artificial agent, and the future is not explicitly marked? What can this tell us about Wikipedia's "temporal regime" (Assmann, 2020)?

We try to answer these questions by focussing on Yapperbot/uncurrenter. We describe the history of Template:Current and recount the debates that led to the bot's creation. We then examine Yapperbot/uncurrenter's contributions to English Wikipedia, comparing its practice with the practice of human editors on English Wikipedia and other-language Wikipedias. As many scholars have observed, bots are powerful actors in Wikipedia's "sociotechnical system", and account for a large share of contributions (Niederer and van Dijck, 2010; Dijck, 2013, pp. 137-140; Geiger and Ribes, 2010; Geiger and Halfaker, 2013; Geiger and Halfaker, 2017; Halfaker and Riedl, 2012; Livingstone, 2016). Bots are also culturally significant. In a series of pathbreaking papers, Stuart Geiger has demonstrated how bots enact or incarnate Wikipedia's culture (Geiger, 2009; Geiger, 2011; Geiger, 2013; Geiger, 2017; see also Kennedy, 2010). As he explains, it is not sufficient to read a bot's source code, although there may well be important policies, procedures or ideals "encoded" in the source (Geiger, 2017, p. 9). To understand bots, it is essential to observe how they act in the wild, and to observe how human users and other bots interact with *them* (Geiger, 2011, 2017). In that spirit, we pursue Yapperbot/uncurrenter through Wikipedia, to see how and when it consigns articles to history. Does it solve the problems identified by the editors who summoned it into existence? And what were those problems anyway?

Methods

```
## Rows: 1648 Columns: 10
## -- Column specification --------------------------------------
## Delimiter: ","
## chr  (3): user, title, comment
## dbl  (5): userid, pageid, revid, parentid, ns
## lgl  (1): texthidden
## dttm (1): timestamp
##
## i Use `spec()` to retrieve the full column specification for thi
## i Specify the column types or set `show_col_types = FALSE` to qu
```

Results

Document analysis

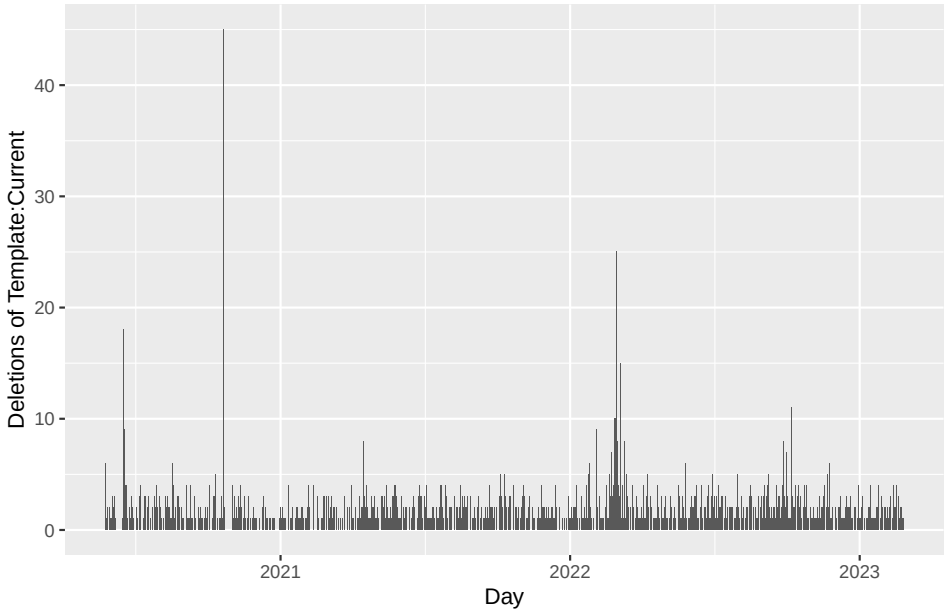
How did 'Yapperbot/uncurrenter' come about? What were the debates and discussions of the editors? What was the perceived problem the bot was supposed to fix?

Quantitative analysis

What does the bot actually do? How does that compare with what it is supposed to do?

How many pages is yapperbot uncurrenting per day?

Yapperbot usually uncurrents less than five articles a day



What was happening on that day it uncurrented 45 pages?

```
## # A tibble: 45 x 3
##      pageid title                                                tim
##       <dbl> <chr>                                                <dt
##  1 63362621 COVID-19 pandemic in Ethiopia                        202
##  2 63431783 COVID-19 pandemic in Northern Ireland                202
##  3 63181042 COVID-19 pandemic in Europe                          202
##  4 63178596 COVID-19 pandemic in Hong Kong                       202
##  5 63313047 COVID-19 pandemic in Moldova                         202
##  6 63239190 COVID-19 pandemic in Sweden                          202
##  7 64307024 Timeline of the COVID-19 pandemic in October 20~ 202
##  8 63395521 COVID-19 pandemic in Sarawak                         202
##  9 63391509 COVID-19 pandemic in Quebec                          202
```

```
## 10 63389195 COVID-19 pandemic in Sabah                                  202
## # ... with 35 more rows
```

How often does Yapperbot have to remove the template more than once?

```
## # A tibble: 7 x 2
##   count num_pages
##   <int>     <int>
## 1     1      1236
## 2     2       133
## 3     3        23
## 4     4         9
## 5     5         4
## 6     6         2
## 7     9         1
```

Which page had to be uncurrented nine times?

```
## `summarise()` has grouped output by 'pageid'. You can override u
## `.groups` argument.
```

```
## # A tibble: 7 x 3
## # Groups:   pageid [7]
##     pageid title                                                    cour
##      <dbl> <chr>                                                    <int
## 1 70168267 Siege of Mariupol
## 2 65760352 Tigray War
## 3 70161957 Siege of Chernihiv
## 4 64399515 2020–2021 Belarusian protests
## 5 70157964 Timeline of the 2022 Russian invasion of Ukraine
## 6 70160923 Battle of Kharkiv (2022)
## 7 70809573 2022–2023 monkeypox outbreak
```

Discussion: What is distinctive about Wikipedia time?

Compared to other temporal regimes

References

Assmann A (2020) *Is Time out of Joint?: On the Rise and Fall of the Modern Time Regime*. Ithaca, UNITED STATES: Cornell University Press. ISBN 978-1-5017-4245-3. URL http://ebookcentral.proquest.com/lib/uts/detail.action?docID=5964894.

Avieson B (2019) Breaking news on Wikipedia: collaborating, collating and competing. *First Monday* DOI:10.5210/fm.v24i5.9530. URL https://firstmonday.org/ojs/index.php/fm/article/view/9530.

Dijck Jv (2013) *The culture of connectivity: a critical history of social media*. Oxford ; New York: Oxford University Press. ISBN 978-0-19-997077-3 978-0-19-997078-0.

Ferron M and Massa P (2011a) The Arab Springl wikirevolutions: Wikipedia as a lens for studying the real-time formation of collective memories of revolutions. *International Journal of Communication* 5: 20.

Ferron M and Massa P (2011b) Collective memory building in Wikipedia: the case of North African uprisings. In: *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-0909-7, pp. 114–123. DOI:10.1145/2038558.2038578. URL https://doi.org/10.1145/2038558.2038578.

Ferron M and Massa P (2014) Beyond the encyclopedia: Collective memories in Wikipedia. *Memory Studies* 7(1): 22–45. Publisher: SAGE Publications Sage UK: London, England.

Ford H (2020) Rise of the Underdog. In: Reagle J and Koerner J (eds.) *Wikipedia@20: Stories of an Incomplete Revolution*. Cambridge, Mass: MIT Press, pp. 189–201. URL https://direct.mit.edu/books/book/4956/chapter/1879827/Rise-of-the-Underdog.

Ford H (2022) *Writing the revolution: Wikipedia and the survival of facts in the digital age*. Cambridge, Massachusetts: The MIT Press. ISBN 978-0-262-04629-9.

Geiger RS (2009) The social roles of bots and assisted editing programs. In: *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*. pp. 1–2.

Geiger RS (2011) The lives of bots. In: Lovink G and Tkacz N (eds.) *Critical Point of View: A Wikipedia Reader*. Institute of Network Cultures: Amsterdam, pp. 78–93.

Geiger RS (2013) Are computers merely" supporting" cooperative work: Towards an ethnography of bot development. In: *Proceedings of the 2013 conference on Computer supported cooperative work companion*. pp. 51–56.

Geiger RS (2017) Beyond opening up the black box: Investigating the role of algorithmic systems in Wikipedian organizational culture. *Big Data & Society* 4(2): 205395171773073. DOI:10.1177/2053951717730735. URL http://journals.sagepub.com/doi/10.1177/2053951717730735.

Geiger RS and Halfaker A (2013) When the levee breaks: without bots, what happens to Wikipedia's quality control processes? In: *Proceedings of the 9th International Symposium on Open Collaboration*. pp. 1–6.

Geiger RS and Halfaker A (2017) Operationalizing conflict and cooperation between automated software agents in wikipedia: A replication and expansion of'even good bots fight'. *Proceedings of the ACM on Human-Computer Interaction* 1(CSCW): 1–33. Publisher: ACM New York, NY, USA.

Geiger RS and Ribes D (2010) The work of sustaining order in Wikipedia: The banning of a vandal. In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. pp. 117–126.

Halfaker A and Riedl J (2012) Bots and Cyborgs: Wikipedia's Immune System. *Computer* 45(3): 79–82. DOI:10.1109/MC.2012.82. URL http://ieeexplore.ieee.org/document/6163451/.

Keegan B (2019) An encyclopedia with breaking news. In: Reagle J and Koerner J (eds.) *Wikipedia@ 20: Stories of an Incomplete Revolution*. Cambridge, Mass: MIT Press, pp. 55–70. URL https://doi.org/10.7551/mitpress/12366.003.0007.

Keegan B, Gergle D and Contractor N (2011) Hot off the wiki: dynamics, practices, and structures in Wikipedia's coverage of the Tōhoku catastrophes. In: *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*. Mountain View California: ACM. ISBN 978-1-4503-0909-7, pp. 105–113. DOI:10.1145/2038558.2038577. URL https://dl.acm.org/doi/10.1145/2038558.2038577.

Kennedy K (2010) Textual Machinery: Authorial Agency and Bot-Written Texts in Wikipedia. In: Smith M and Warnick B (eds.) *The Responsibilities of Rhetoric*. Long Grove, Illinois: Waveland, pp. 303–309. URL https://surface.syr.edu/wp/1.

Livingstone RM (2016) Population automation: An interview with Wikipedia bot pioneer RamMan. *First Monday* DOI:10.5210/fm.v21i1.6027. URL https://firstmonday.org/ojs/index.php/fm/article/view/6027.

Niederer S and van Dijck J (2010) Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society* 12(8): 1368–1387. DOI:10.1177/1461444810365297. URL https://doi.org/10.1177/1461444810365297. Publisher: SAGE Publications.

Pentzold C (2009) Fixing the floating gap: The online encyclopaedia Wikipedia as a global memory place. *Memory Studies* 2(2): 255–272. DOI:10.1177/1750698008102055. URL http://journals.sagepub.com/doi/10.1177/1750698008102055.

Porter E, Krafft PM and Keegan B (2020) Visual Narratives and Collective Memory across Peer-Produced Accounts of Contested Sociopolitical Events. *ACM Transactions on Social Computing* 3(1): 4:1–4:20. DOI:10.1145/3373147. URL http://doi.org/10.1145/3373147.

Rosenzweig R (2006) Can History Be Open Source? Wikipedia and the Future of the Past. *The Journal of American History* 93(1): 117–146. DOI:10.2307/4486062. URL http://www.jstor.org/stable/4486062. Publisher: [Oxford University Press, Organization of American Historians].

Twyman M, Keegan BC and Shaw A (2017) Black Lives Matter in Wikipedia: Collaboration and Collective Memory around Online Social Movements. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. pp. 1400–1412. DOI:10.1145/2998181.2998232. URL http://arxiv.org/abs/1611.01257. ArXiv:1611.01257 [physics].