

5 new learnings from our research on readers

Martin Gerlach, Yu-Ming Liou, Michael Raish (Research)

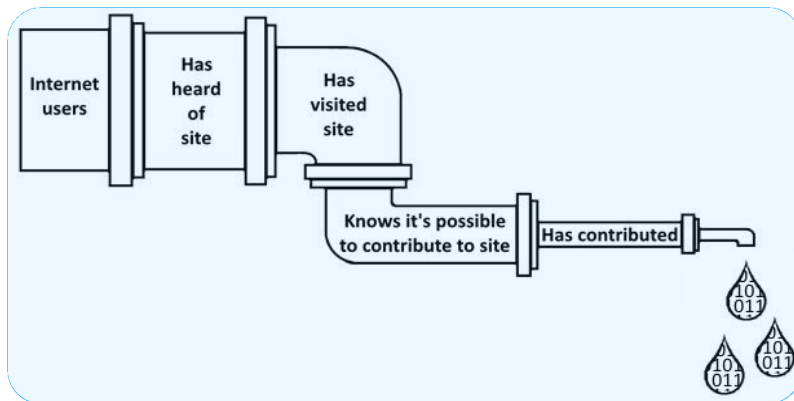
Why we need research on readers

Readers are experiencing a **changing internet**



Understanding **readers' needs and barriers** is key to serve our current & future audiences

Knowledge gaps propagate along the **pipeline of online participation**



Example: **Wikipedia's gender gap**

Readers: ~33% → Editors: <15% → Content: <20%

Previous research on readers

Previous research work



April 2022 staff meeting presentation

https://w.wiki/_vaEQ

Today's presentation



An **expanded Research Team:**

ML/AI models, Surveys, Interviews

1: Information needs vary over time

Location of readers matters. What about time?

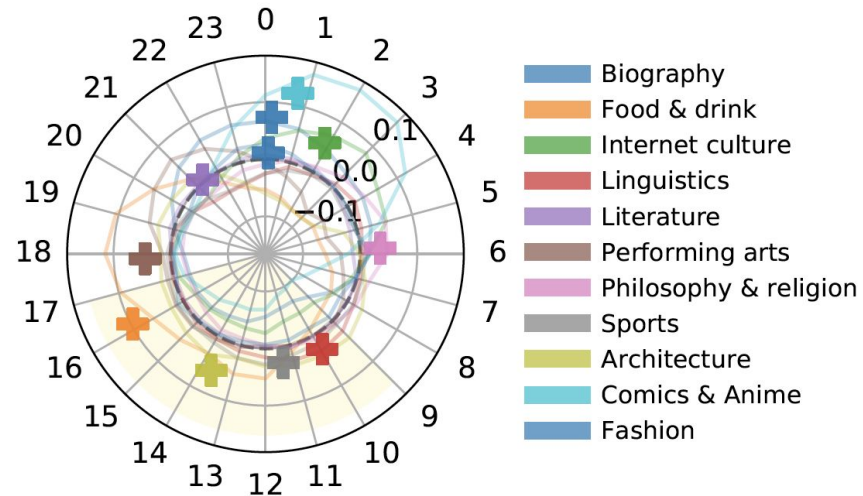
Challenge: UTC → local time of reader

Analysis: 3.5B pageloads in English Wikipedia

Overall **circadian rhythm**

Mobile peaks in the evening (+50%)

Topic preferences change over time



Capture engagement beyond pageviews counts

Use reading patterns to optimize content delivery (e.g. caching)

2: Much of the existing content is invisible

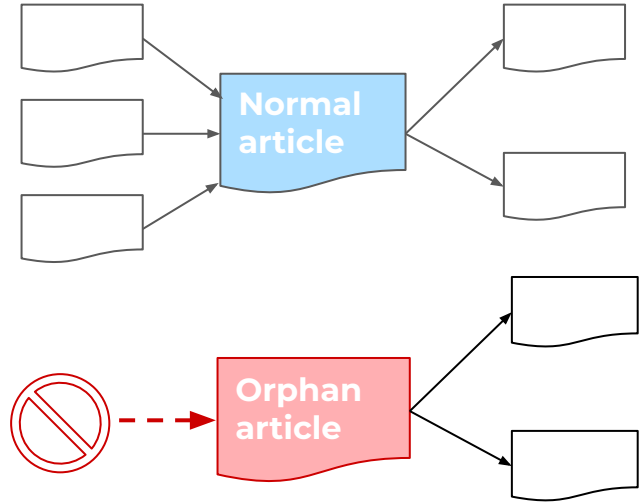
How easy is it to discover relevant articles?

Articles not visible to readers navigating Wikipedia

8.8M orphan articles in all Wikipedias (~15%)

Gender bias (women biographies overrepresented)

New link *leads to* +10% pageviews



Support editors with *de-orphanization* (e.g. via structured tasks)

Recommend orphan articles to readers

3: Readability is a major barrier

How easy is it to read and understand written text?

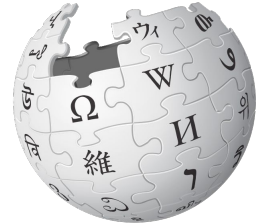
Work: a multilingual model to **score readability** of Wikipedia articles

Results: Text difficulty > Reading ability

English: **58% of articles at grade level ≥ 9**

Majority of content **too difficult for average adult**

Create customizable reading experiences using generative AI to make existing content more accessible to readers



Simple English
WIKIPEDIA



4: Our readers are disproportionately young, educated, and men

Who is reading Wikipedia?

Large-scale **reader survey in 23 wikis** (80K responses)

Regular snapshots of our readers

↓ **age** (49% under 30, includes <18)

↑ **education** (~23% with HS degree or less, 18+ only)

↑ **men** (~66% identify as men, 18+ only)

Learn more: April staff meeting (Yu-Ming)

Understand causes of readership gaps



Take a short survey and help us improve Wikipedia ×

Survey data handled by a third party. [Privacy policy](#).

5: Readers create accounts to *read*

What motivates readers to create accounts?

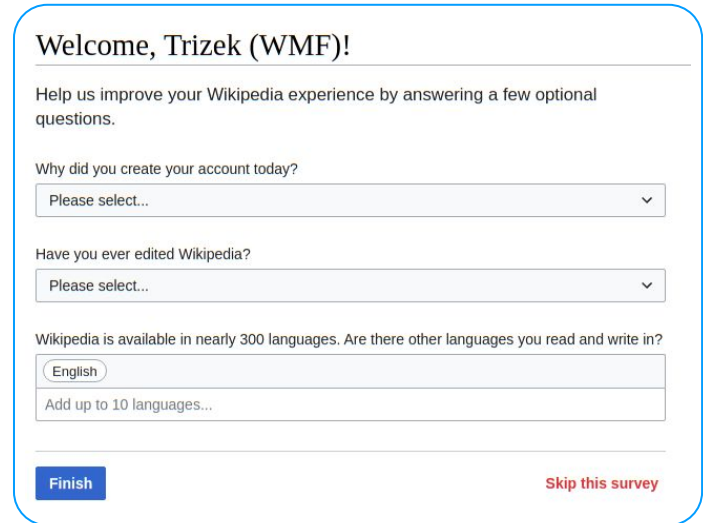
Survey + Interviews with **new account holders**

>20% didn't know they could edit

20% reason for sign-up: reading

Enthusiasm for “individual” features (e.g. create lists)

Strong interest in Wikipedia (incl. editing)



Welcome, Trizek (WMF)!

Help us improve your Wikipedia experience by answering a few optional questions.

Why did you create your account today?

Please select... ▾

Have you ever edited Wikipedia?

Please select... ▾

Wikipedia is available in nearly 300 languages. Are there other languages you read and write in?

English

Add up to 10 languages...

[Finish](#) [Skip this survey](#)

Raise awareness about how Wikipedia works

Support readers in new ways to become contributors

Thank you!

Question? Reach out: mgerlach@, yliou@, mraish@

Thanks to: Formal collaborators (EPFL, Stanford, GESIS), Research, Machine Learning, Growth, Legal

Learn more about the projects

#1 Time <https://w.wiki/97oY>

#2 Visibility <https://w.wiki/4LHD>

#3 Readability <https://w.wiki/5Kdd>

#4 Reader demographics <https://w.wiki/9Kqz>

#5 New editor survey <https://w.wiki/9Kns> / Non-editing participation <https://w.wiki/9SyW>

Credits

- Slide 2: Internet. <https://thenounproject.com/icon/internet-2126864/> (CC BY 3.0) by ibrandify
- Slide 2: Pipeline of online participation inequalities.
https://commons.wikimedia.org/wiki/File:Pipeline_of_online_participation_inequalities.png (CC BY-SA 4.0) by Aaronshaw
- Slide 3: 5 Learnings from Research on Reader Navigation.
[https://commons.wikimedia.org/wiki/File:\(Staff_meeting_2022-04\)_Learnings_from_reader_navigation_research.pdf](https://commons.wikimedia.org/wiki/File:(Staff_meeting_2022-04)_Learnings_from_reader_navigation_research.pdf) (CC BY-SA 4.0) by MGerlach_(WMF)
- Slide 4: Typical access time of topics. Figure 9 <https://arxiv.org/abs/2305.09497> by Tiziano Piccardi
- Slide 6: Logo for Simple English Wikipedia.
<https://commons.wikimedia.org/wiki/File:Wikipedia-logo-v2-simple.svg> (CC BY-SA 3.0) by Odder
- Slide 6: Logo for Vikidia https://commons.wikimedia.org/wiki/File:Logo_Vikidia_2019.svg (CC BY-SA 4.0) by Tinker Bell
- Slide 6: Logo for Txikipedia. <https://commons.wikimedia.org/wiki/File:Txikipedia.svg> (CC BY-SA 4.0) by Theklan
- Slide 7: Quicksurvey for Global reader survey. Screenshot <https://phabricator.wikimedia.org/F37697326> by DDeSouza
- Slide 8: Welcome survey.
https://commons.wikimedia.org/wiki/File:Screenshot_of_welcome_survey_2021-01-18.png (CC BY-SA 4.0) by Trizek_(WMF)