Why We Read
Wikipedia

Philipp Singer*, Florian Lemmerich*, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, Jure Leskovec
The readers

Every second, 6000 Wikipedia pages are viewed by people -- And we don’t know why!

Providing educational content and effectively disseminating it requires understanding the needs and motivations of the people behind these pageviews.
But where to start?

Surveys + Webrequest logs

CC BY SA 4.0, Singer, Philipp, et al. “Why We Read Wikipedia”
What to ask?

Survey 1: build the taxonomy (en)
“Personal interest about conflicts in middle east”

“Confirming address for shipment going to this town”

“Studying for my med school test”

“Because I am in a very boring art lesson”

“To see a movie summary”

“Interest and curiosity”

“So I can see the country’s population”

“NY Times today mentioned Operation Wetback, alluded to by Trump in debate, & wanted to learn more.”

“To find out more information about this aircraft.”

“I had previously edited it.”

“Someone came by my desk talking about The Last Man on Earth (movie). So I looked it up.”
Why are you reading this article today?

Information need

I am reading this article to
- look up a specific fact or to get a quick answer.
- get an overview of the topic.
- get an in-depth understanding of this topic.

Prior knowledge

Prior to visiting this article
- I was not familiar with the topic and I am learning about it for the first time.
- I was already familiar with the topic.

motivation

I am reading this article because
Please select all answers that apply
- the topic was references in a piece of media (e.g., TV, radio, article, film, book).
- I need to make a personal decision based on this topic (e.g. to buy a book, choose a travel destination).
- I am bored and randomly exploring Wikipedia for fun.
- the topic came up in a conversation.
- I have a work or school-related assignment.
- I want to know more about a current event (e.g. a soccer game, a recent earthquake, somebody’s death)
- this topic is important to me and I want to learn more about it. (e.g., to learn about a culture).
- Other
What to ask?

Survey 1: build the taxonomy (en)

Survey 2: assess robustness (en, es, fa)
What to ask?

Survey 1: build the taxonomy (en)

Survey 2: assess robustness (en, es, fa)

Survey 3: the grand survey (en)
The grand survey - 2016

- Duration: 1 week
- Sampling rate: 1 out of 50
- English Wikipedia, mobile and desktop
- On article pages and to those with DNT off
- Participants: 29,372
Bias

- Goal: study the motivations and behaviors representative of Wikipedia’s entire reader population.
- Method: survey a small population
- Bias
  - Coverage bias (inability to reach certain subpopulations)
  - Sampling bias (distortions due to sampling procedure)
  - Non-response bias (diverse likelihood of survey participation after being sampled as a participant)
Results
Information need

I am reading this article to
- look up a specific fact or to get a quick answer.
- get an overview of the topic.
- get an in-depth understanding of this topic.
Information need

Subgroups of users based on information need are homogenous, however, there are some notable anecdotal exceptions

- users from Asia describe their information needs significantly more as in-depth
- users interested in getting an overview use desktop more often
- topics play a certain role: fact look-ups are more often observed on sports topics
Prior knowledge

Prior to visiting this article
- I was not familiar with the topic and I am learning about it for the first time.
- I was already familiar with the topic.
Prior knowledge

Users who report familiarity with articles and topics read:

- spare-time oriented topics, such as sports, 21st century, TV/movies/novels
- popular articles
- longer articles
- articles that are more central in the link network
Motivation

I am reading this article because
Please select all answers that apply

☐ the topic was references in a piece of media (e.g., TV, radio, article, film, book).

☐ I need to make a personal decision based on this topic (e.g. to buy a book, choose a travel destination).

☐ I am bored and randomly exploring Wikipedia for fun.

☐ the topic came up in a conversation.

☐ I have a work or school-related assignment.

☐ I want to know more about a current event (e.g. a soccer game, a recent earthquake, somebody’s death)

☐ this topic is important to me and I want to learn more about it. (e.g., to learn about a culture).

☐ Other
Motivation: work/school

Those motivated by work or school projects:

- read articles about war & history, mathematics, technology, biology, chemistry, literature & arts
- read topics that are related to academic and professional activities than to leisure
- stay longer on an individual page
- are more likely to use external search engines
- use Wikipedia desktop version
Motivation: bored/rand.

Those who are bored or are randomly exploring Wikipedia:

- read articles about sports, 21st century, TV/movie/novels
- read topics that are more spare-time oriented
- spend only little time on individual articles
- come to Wikipedia more often
- switch topics frequently
- are more likely to use internal navigation
Summary - Part 1

- A robust taxonomy for characterizing use cases for reading Wikipedia
- Quantifying the prevalence and interactions between these use cases via a large-scale survey on English Wikipedia
- Enhanced understanding of behavioral patterns associated with different use cases by combining survey responses with webrequest logs

Why We Read Wikipedia, Singer et al., International World Wide Web Conference, Perth, Australia, 2017
Robustness of results

- Are the results robust over time?
- Are the results robust across languages?

Repeat the grand survey in 2017 and across more languages!
The great grand survey

- Duration: 1 week
- 14 languages: Arabic, Bengali, Chinese, Dutch, English, German, Hebrew, Hindi, Hungarian, Japanese, Romanian, Russian, Spanish, and Ukrainian
- Mobile and Desktop platforms
- Sampling rate: varied across languages
- On article pages and to those with DNT off
- Responses: 215,390

[m:Research:Characterizing_Wikipedia_Reader_Behaviour/Robustness_across_languages]
Preliminary results
2158 responses (1:10)

User: عبادـديرانية
Preliminary results. Results can change as we learn more in the coming months.
Bengali
1199 responses (1:1)
User: Hasive
Preliminary results. Results can change as we learn more in the coming months.
German
28,311 responses (1:5)

User:Lyzzy
Preliminary results. Results can change as we learn more in the coming months.
English
21,777 responses (1:40)

User:LZia_(WMF)
Preliminary results. Results can change as we learn more in the coming months.
Spanish
39,005 responses (1:5)
User:Racso
Preliminary results. Results can change as we learn more in the coming months.
Hebrew
8880 responses (1:3)
User: Amire80
Preliminary results. Results can change as we learn more in the coming months.
Hindi
3065 responses (1:2)
User:Satdeep_Gill
Preliminary results. Results can change as we learn more in the coming months.
Hungarian
2467 responses (1:2.5)

User:Tgr
Preliminary results. Results can change as we learn more in the coming months.
Japanese
20,049 responses (1:5)

User: Whym
Preliminary results. Results can change as we learn more in the coming months.
Dutch
3281 responses (1:8)
User: AWossink
Preliminary results. Results can change as we learn more in the coming months.
Romanian
3833 responses (1:2)
User: Strainu
Preliminary results. Results can change as we learn more in the coming months.
Russian
67,586 responses (1:5)
User: Kaganer
Preliminary results. Results can change as we learn more in the coming months.
Ukrainian
8053 responses (1:2.5)
User: Antananana
Preliminary results. Results can change as we learn more in the coming months.
Chinese
5726 responses (1:20)
User: Shangkuanlc
Preliminary results. Results can change as we learn more in the coming months.
Next steps

● Finish the analysis
● Brainstorm about next steps: Readers <-> Editors
● Why We Edit Wikipedia, And Why Some Don’t.
● ...

...
Thank you!

- Ongoing documentation on meta at https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour
How to ‘debias’?

Survey participants

Wikipedia readers
<table>
<thead>
<tr>
<th>Survey</th>
<th>Request</th>
<th>Article</th>
<th>Session/Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>Country</td>
<td>In-degree</td>
<td>Session length</td>
</tr>
<tr>
<td>Information need</td>
<td>Continent</td>
<td>Out-degree</td>
<td>Session duration</td>
</tr>
<tr>
<td>Prior knowledge</td>
<td>Local time weekday</td>
<td>Pagerank</td>
<td>Average dwell time</td>
</tr>
<tr>
<td></td>
<td>Local time hour</td>
<td>Text length</td>
<td>Average pagerank difference</td>
</tr>
<tr>
<td></td>
<td>Host</td>
<td>Pageviews</td>
<td>Average topic distance</td>
</tr>
<tr>
<td></td>
<td>Referer class</td>
<td>Topics</td>
<td>Referer class frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Topic entropy</td>
<td>Session position</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of sessions</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Number of requests</td>
</tr>
</tbody>
</table>