The Knowledge Gaps Taxonomy
Research @Wikimedia

Research Team

WIKIMEDIA FOUNDATION
### Motivation and problem statement

Our goal is to empower people to *share in the sum of all knowledge.*

We do not have a shared definition of *knowledge gaps.*

At least some of our decisions are at best sub-optimal because we do not see the full spectrum of the gaps we face, the relationship between these gaps, and the possibilities to address them.

Plus, we don’t have standard practices for measuring our knowledge gaps and tracking our progress (or lack thereof).

### Proposed solution

- Develop a taxonomy of knowledge gaps.
- Develop and gather metrics for measuring every level of the taxonomy.
- Define targets for closing gaps.
- Develop a composite index to measure knowledge gaps (across languages?)

### Where are we now?

We have proposed a *definition* for knowledge gaps and developed a *taxonomy* of knowledge gaps.

We are seeking your feedback to improve the taxonomy of knowledge gaps.

---

Core principle: How we arrive at the destination (*a knowledge gap index*) is as important, if not more important, than the index itself. This is a research effort and we actively seek collaborations, feedback, radical ideations, etc. We are committed to do this work with you and for you.
The global direction: Knowledge Equity
[from Movement Strategy]

**Knowledge equity:** As a social movement, we will focus our efforts on the knowledge and communities that have been left out by structures of power and privilege. We will welcome people from every background to build strong and diverse communities. We will break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge.
How far are we from reaching knowledge equity?

Operationalize knowledge equity
Identify and measure the individual components based on which we can track our progress towards this goal
How far are we from reaching knowledge equity?
Examples from research on Wikipedia gender gap

Distribution of languages by % of female biographies

Source: Wikidata Human Gender Indicators (WHGI)
How to do this more systematically?
First step is to identify individual gaps which prevents from reaching knowledge equity. The taxonomy then maps all types of gaps we can find in Wikimedia Spaces.

**Knowledge gaps:**
Disparities with respect to coverage of specific groups of readers, contributors or content across Wikimedia projects.
Taxonomy of Knowledge gaps: how we built it
Finding the taxonomy’s root dimensions
Taxonomy of Knowledge gaps
To be able to quantify gaps in knowledge equity, we first need a systematic definition of what they are.
Taxonomy of Knowledge gaps: how we built it
Finding evidence of knowledge gaps from different sources

Academic Literature

Movement Strategy and Initiatives

Community Surveys
Taxonomy of Knowledge gaps: Structure

Root Dimensions:
main actors of the Wikimedia Ecosystem
Taxonomy of Knowledge gaps: Structure

Gaps:
individual aspects of the Wikimedia Ecosystem having evidence of inequality

Root Dimensions:
main actors of the Wikimedia Ecosystem
Taxonomy of Knowledge gaps: Structure

**Facets:**
Groups of gaps referring to similar properties or objectives

**Gaps:**
Individual aspects of the Wikimedia Ecosystem having evidence of inequality

**Root Dimensions:**
Main actors of the Wikimedia Ecosystem
Readers
Readers-facet 1: Sociodemographics

Wikipedia, past and present

A snapshot of current Wikipedia users

BY KATHRYN ZICKUHR AND LEE RAINIE

As of May 2010, 53% of American internet users look for information on Wikipedia, up from 36% of internet users the first time we asked about Wikipedia usage in February 2007.
Readers-facet 2: Accessibility
Readers-facet 3: Information Need

Why the world reads Wikipedia: What we learned about reader motivation from a recent research study

By Florian Lemmerich, Bob West and Leila Zia • 15 March 2018

motivation

<table>
<thead>
<tr>
<th>response percentage</th>
<th>media</th>
<th>intrinsic learning</th>
<th>conversation</th>
<th>bored/random</th>
<th>work/school</th>
<th>current event</th>
<th>personal decision</th>
<th>other</th>
<th>no response</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Contributors
Content
Content-facet 1: Policy

Wikipedia: Core content policies

This is an explanatory supplement to the Neutral point of view, Verifiability and No original research policies. This page is intended to provide additional information about concepts in the page(s) it supplements. This page is not one of Wikipedia's policies or guidelines, as it has not been thoroughly vetted by the community.

Wikipedia's content is governed by three principal core content policies: neutral point of view, verifiability, and no original research.
Content-facet 2: Diversity

Gender
Cultural Context
Diversity
Geography
Impactful Topics

Gender Gap

A hub for resources and information about Wikimedia's gender gap.

Two kinds of gender gap exist within, and do harm to, the Wikimedia world: (a) a content gender gap (meaning that more men than women are covered in the mainspace content of our wikis), and (b) a participation gender gap, meaning that more men participate in the peer production communities of Wikimedia. This page seeks to act as a hub for resources and information about Wikimedia's gender gaps, and to adequately quantify the gender gaps, in particular by presenting studies on the subject and by collecting anecdotal evidence about why women, LGBTIQ and under-represented gender identity minorities leave or never join Wikipedia.

The gender gap mailing list is a place to talk about this with other people who are interested and
Content-facet 3: Accessibility

Readability of Wikipedia
by Teun Lucassen, Roald Dijkstra, and Jan Maarten Schraagen

Figure 3: Histogram of reading ease scores in the English Wikipedia.
Taxonomy of Knowledge Gaps
Intersections between Knowledge Gap Dimensions:
From contributors -> content

Read More: Brent Hecht and Darren Gergle. Measuring Self-Focus Bias in Community-Maintained Knowledge Repositories
Intersections between Knowledge Gap Dimensions: From readers -> contributors

Read More: Aaron Shaw and Eszter Hargittai. The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing
Intersections between Knowledge Gap Dimensions:
From content -> readers

People tend to read content that relates to their identity and context: gender, age, geography, etc.

Credits 1/2

Page 7, 8, 25: Model of engagement with the Wikimedia movement by Margeigh Novotny / CC BY-SA 4.0
https://commons.wikimedia.org/wiki/File:Wikimedia_Foundation_2019_Medium_Term_Plan_-_model.png

Page 9: A visual for the Wikimedia Movement Strategy recommendations by Riesenspatz / Svenja Kirsch, Anna Lena Schiller, riencespatz.de / CC BY-SA 4.0
https://commons.wikimedia.org/wiki/File:Wikimedia_2018-20_Recommendations_all_in_one_final.svg

Page 9, 14: Global South User Survey 2014 - Full Analysis by Report HaithamS (WMF) / CC BY-SA 4.0

Page 23 By Addshore - Source: https://tools.wmflabs.org/wikidata-analysis/20190513/geo2png/ Data:

Page 24 By Aaronshaw - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=69429179
Credits 2/2 (screenshots)

Page 5 is from the Wikidata Human Gender Indicators (WHGI): http://whgi.wmflabs.org/
Page 9 is from Shaw and Hargittai "The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing": https://academic.oup.com/joc/article-abstract/68/1/143/4915319
Page 15 is from the flyer for Kiwix: https://commons.wikimedia.org/wiki/File:New_Kiwix_Flyer_ENG.pdf
Page 16 (bottom) is from Figure 2 of the paper “Why we read Wikipedia”: https://arxiv.org/abs/1702.05379
Page 19 is from the English Wikipedia core content policy page: https://en.wikipedia.org/wiki/Wikipedia:Core_content_policies
Page 20 is the Gender Gap Portal from meta.wikimedia.org: https://meta.wikimedia.org/wiki/Gender_gap
Page 21 (top) is from the article “Readability of Wikipedia”: https://firstmonday.org/ojs/index.php/fm/article/view/3916/3297
Page 21 (bottom) is from Simple English Wikipedia main page: https://simple.wikipedia.org/wiki/Main_Page
Thank you and please give us feedback!

https://meta.wikimedia.org/wiki/Research_talk:Knowledge_Gaps_Index/Taxonomy

OR (same link; easier to copy)

https://w.wiki/d79
BACKUP
Intersections between Knowledge Gap Dimensions: From readers -> contributors

Pipeline of Participation Inequalities (United States + English Wikipedia)

Sources:
2 Aaron Shaw and Eszter Hargittai. The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing