

Growing Wikipedia Across Languages via Recommendation Systems



Leila Zia
User:LZia_(WMF)
Wikimedia Foundation



Hi!

مرحبا!

Salut!

- I am Leila. Wikidata ID: Q27826865.
- I am originally from Iran.
- I am a senior research scientist in Wikimedia Foundation.
- I am the technical lead for the Research team.
- I work in the San Francisco office. Kind of.
- I've been in the Foundation for the past 4 years.
- I  research, I  people.

I am here to ...

- Talk with you about Knowledge Gaps
- Hear from **you** about how we can support you better

Who is *we*?



Dario Taraborelli



Diego Saez



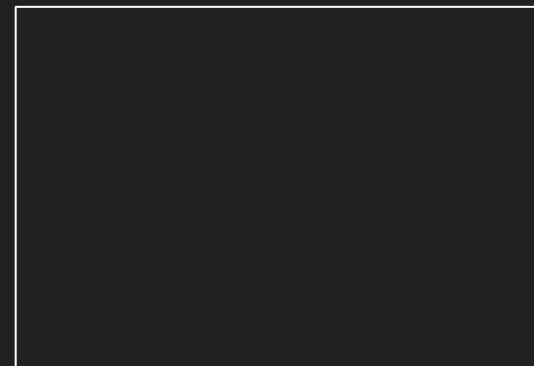
Jonathan Morgan



Miriam Redi



Erik Zachte



Bahodir Mansurov

This is interactive

Ask me questions anytime during the presentation.

- Raise your hand and someone in the room will signal to me that you have a question
- Submit your question in written using the Google URL you see in all slides

Knowledge Gaps

Knowledge gaps are ...



Readers

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.

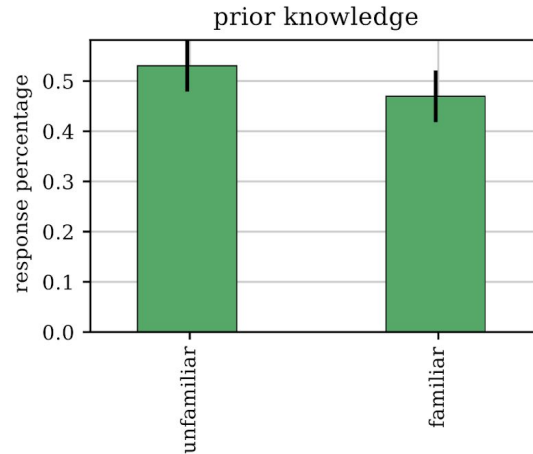
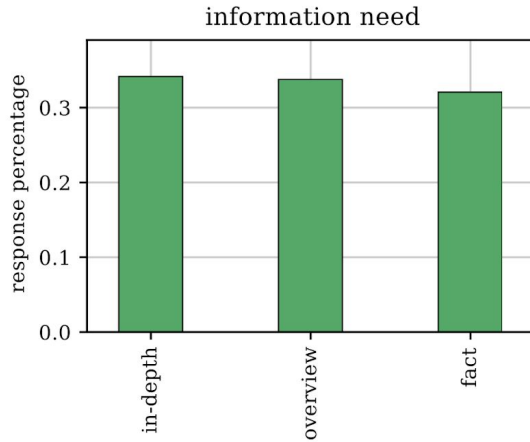
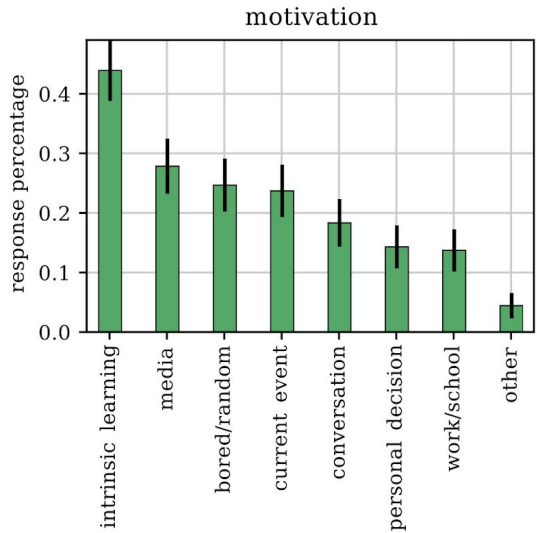
I am reading this article because

Please select all answers that apply

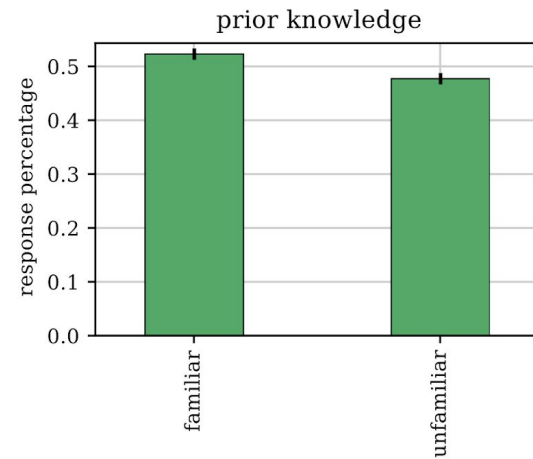
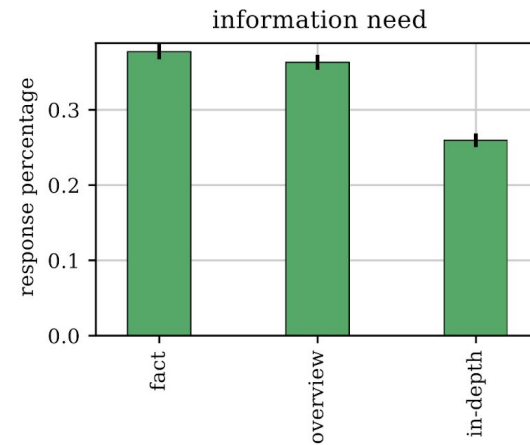
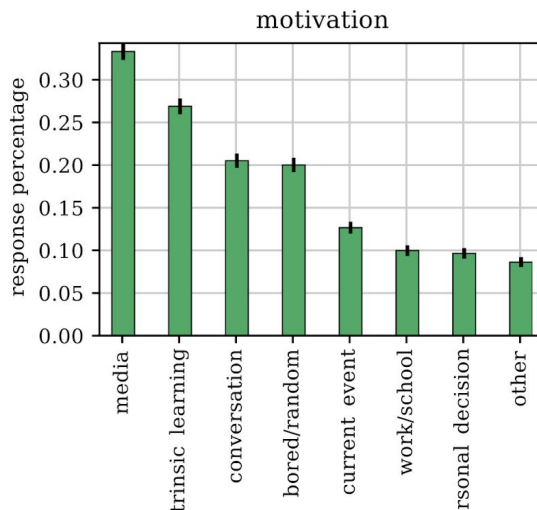
- the topic was references in a piece of media (e.g., TV, radio, article, film, book).
- I need to make a personal decision based on this topic (e.g. to buy a book, choose a travel destination).
- I am bored and randomly exploring Wikipedia for fun.
- the topic came up in a conversation.
- I have a work 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

Arabic



English



A few takeaways

For Arabic Wikipedia,

- a lot of in-depth readers. **Expand** articles.
- More than 40% report they come to Wikipedia **just to learn**. This is a very high number.
- While there are many education initiatives, only 10% of the readers report a work or school project as a motivation to come to read an article. Remember that **your reader base is very diverse**.

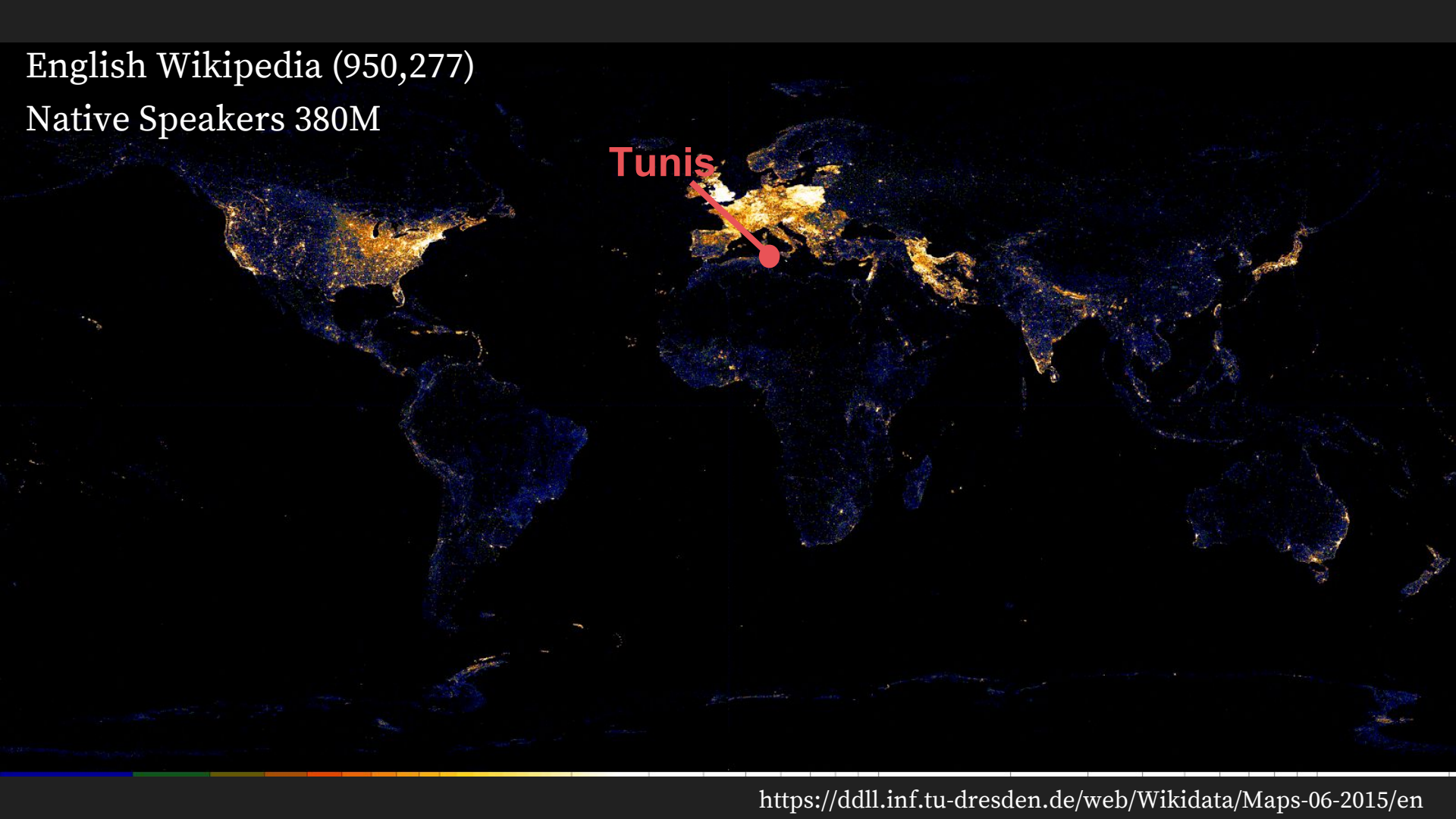
Content

2015

English Wikipedia (950,277)

Native Speakers 380M

Tunis



Arabic Wikipedia (87,017)

Native Speakers 310M



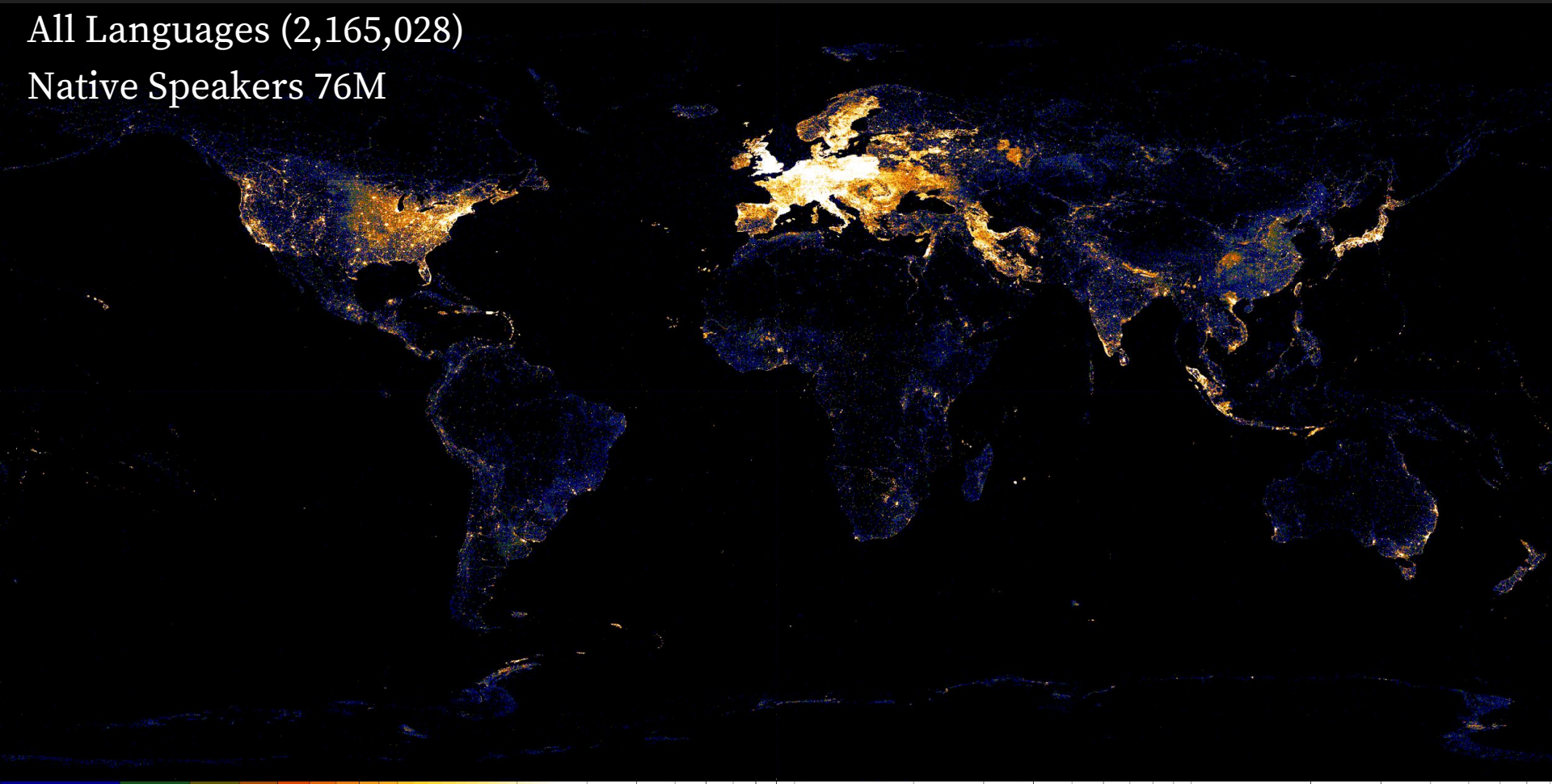
French Wikipedia (347,419)

Native Speakers 76M



All Languages (2,165,028)

Native Speakers 76M



Let's build a system
that recommends
what articles to
create next!



Bob West (EPFL)

The system



Articles



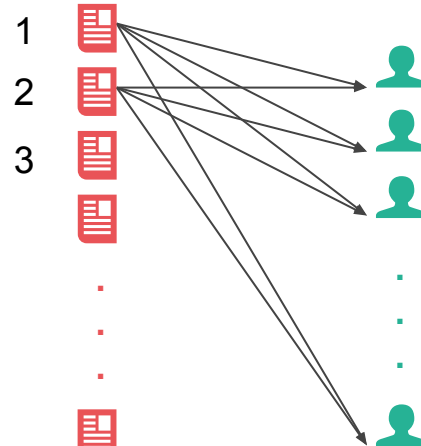
Articles



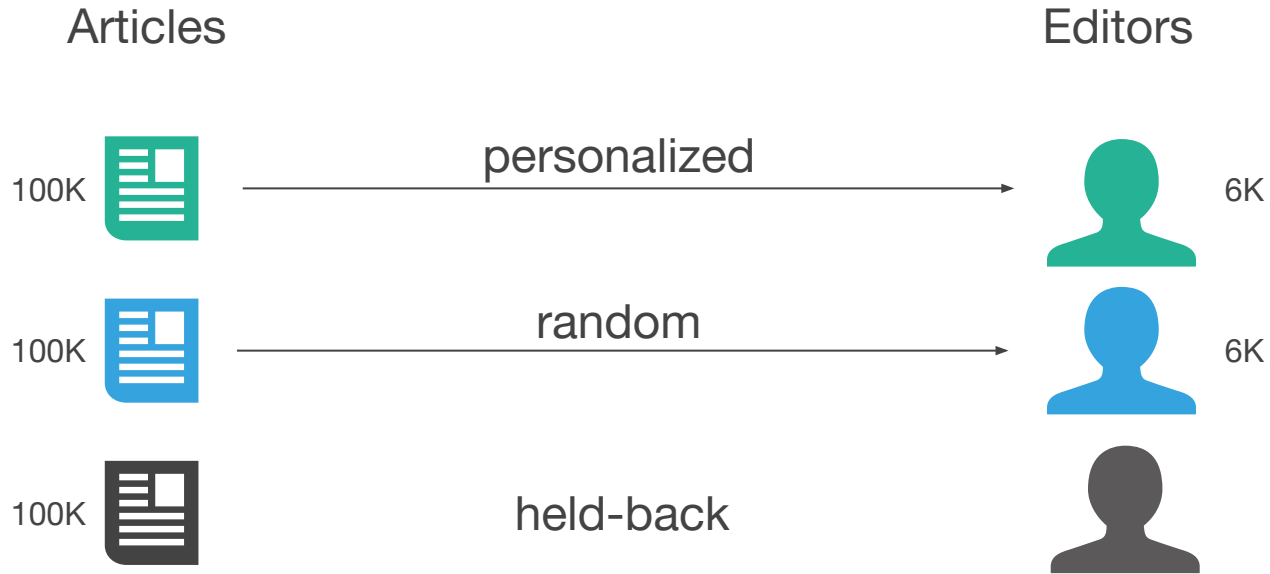
Articles



Editors



Design of the experiment



We tested the system

And learned,

with personalized article recommendations we can increase article creation rate by **a factor of 3** without compromising on quality and only with the current editor pool.

 Wikipedia **GapFinder**

English ▾

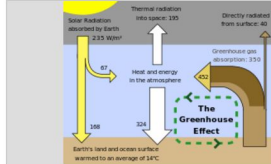
Malagasy ▾

climate change



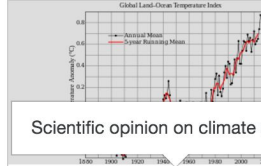
Intergovernmental P...
scientific intergovernmental body

10546 recent views



Greenhouse gas
gas in an atmosphere that
absorbs and emits radiation within

51057 recent views



Scientific opinion on climate change

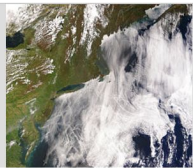
Scientific opinion on ...

8047 recent views



Effects of global war...

22475 recent views



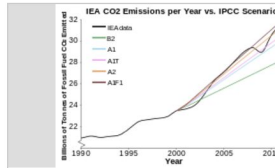
Climate change
significant time variation in long-
term weather patterns

0 recent views



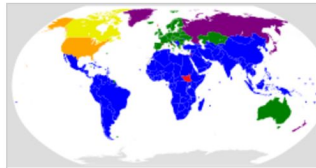
Tropical cyclone
storm system

27758 recent views



Climate change mitig..
actions to limit climate change in
order to reduce the risks of global

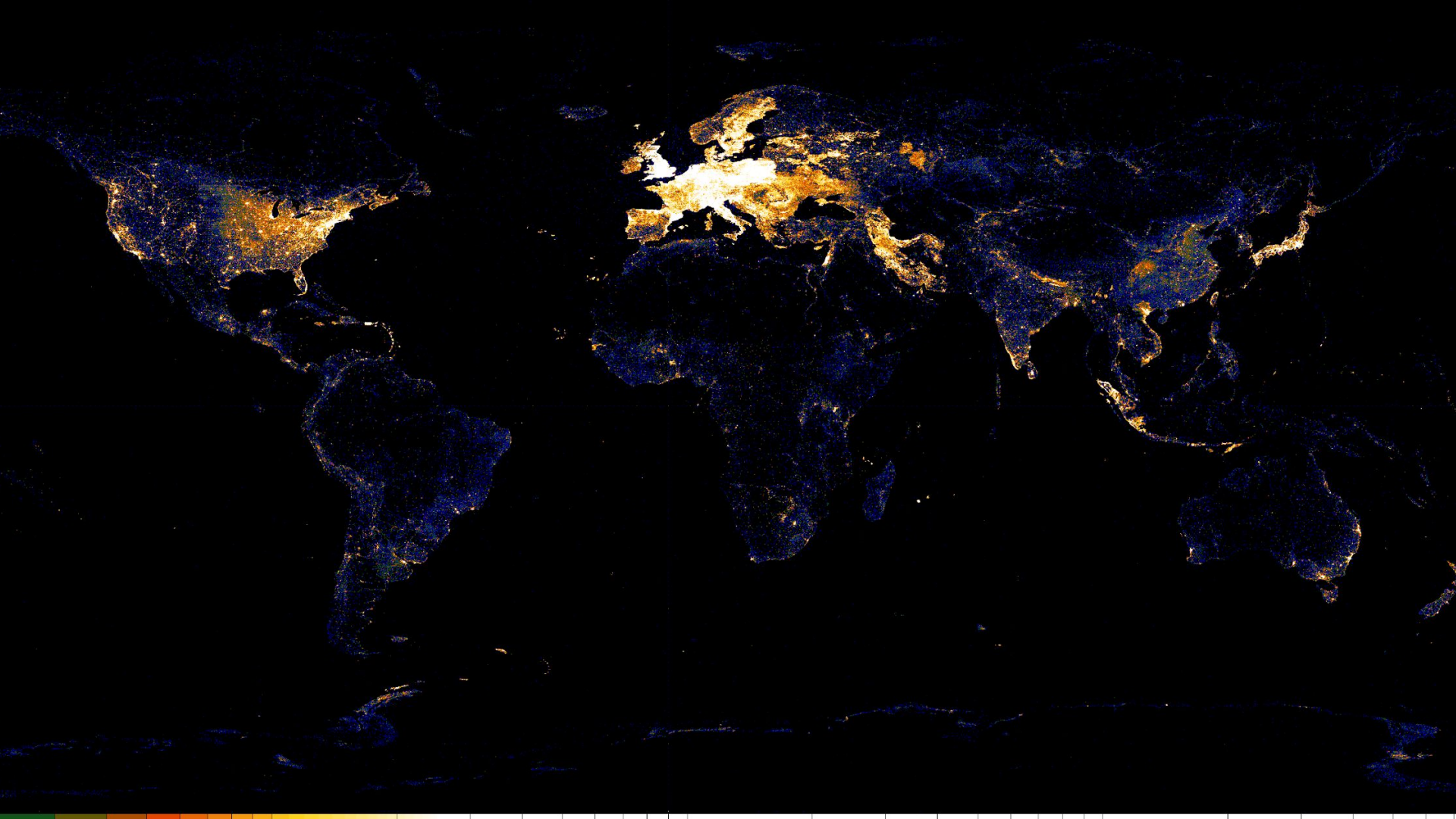
4595 recent views



Kyoto Protocol
International Treaty to reduce
greenhouse gas emissions

37088 recent views





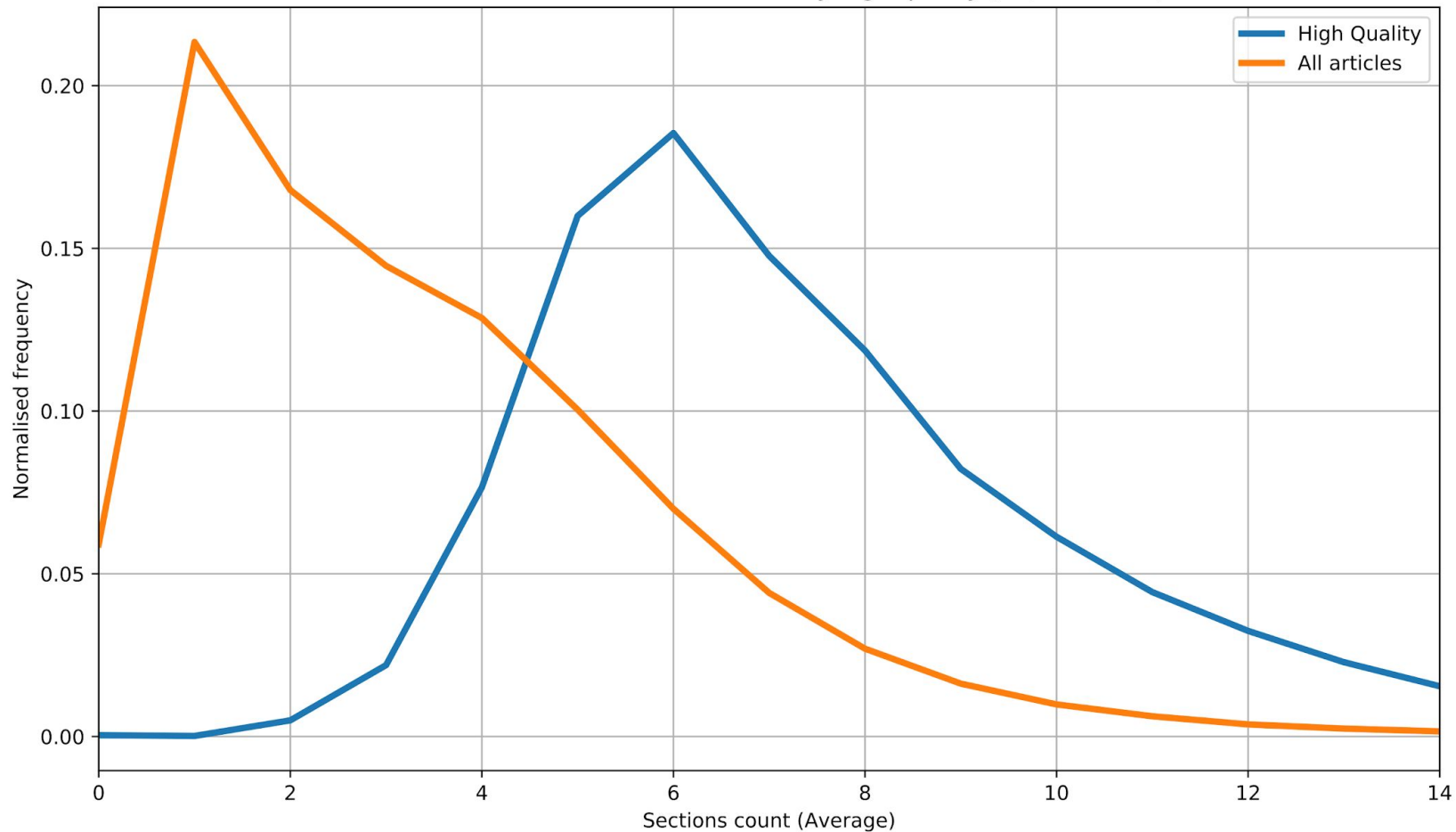
**Wikipedia articles are
incomplete!**

**Only 1% of English
Wikipedia Articles have
quality label Good or
better**

37% are stubs!

2016

Sections count for: all articles vs. only high quality (FA, GA and A)



Take-away

- We have **a lot of missing content** in Wikipedia articles
- Editathon organizers are spending **a lot of time** on manually creating templates that can help onboard newcomers
- There are initiatives that have identified these needs and are working on this problem and the challenge for them is how to **scale** the recommendations beyond very specific topics/types of articles

2016-11-17



Florence Devouard
User:Anthere

- Stubs are a real problem!
- We need to be able to easily generate
 - a list of articles that need expansion
 - templates that can help newcomers learn how to expand the article: what section to add, what characteristics the section should have, ...

Template generation

Flo: I want to be able to give you a Wikipedia category I'm interested in and your system tells me what sections should be considered for articles in this category.

We accepted this challenge. And here is how far we are:

- ▼ 0:
 - 0: "Decorations"
 - 1: 2.274393090569561
- ▼ 1:
 - 0: "Biography"
 - 1: 1.9506621222870923
- ▼ 2:
 - 0: "Early life"
 - 1: 1.700738186991944
- ▼ 3:
 - 0: "Career"
 - 1: 1.5040399153532769
- ▼ 4:
 - 0: "Legacy"
 - 1: 0.9342806840123286
- ▼ 5:
 - 0: "Early life and education"
 - 1: 0.7602971514319761
- ▼ 6:
 - 0: "Family"
 - 1: 0.6716867497513697
- ▼ 7:
 - 0: "Life and political career"
 - 1: 0.49007936507936506
- ▼ 8:
 - 0: "Death"
 - 1: 0.4580425001743164

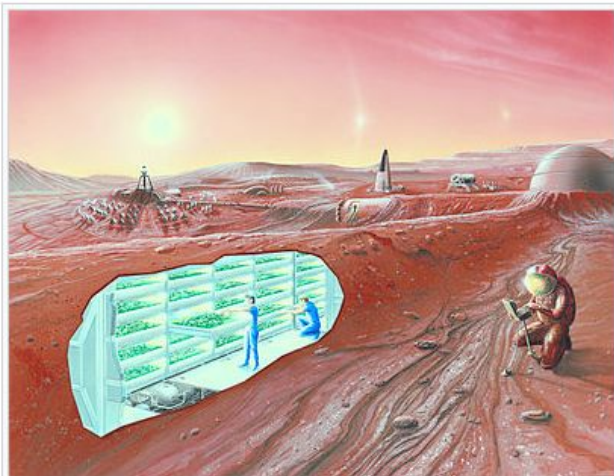
- ▼ 0:
 - 0: "Early life"
 - 1: 0.13062492806997353
- ▼ 1:
 - 0: "Career"
 - 1: 0.12751755092645875
- ▼ 2:
 - 0: "Personal life"
 - 1: 0.12314420531706756
- ▼ 3:
 - 0: "Early life and education"
 - 1: 0.12268385314765796
- ▼ 4:
 - 0: "Biography"
 - 1: 0.1115203130394752
- ▼ 5:
 - 0: "Background"
 - 1: 0.07653354816434572
- ▼ 6:
 - 0: "Federal judicial service"
 - 1: 0.07158476234319254
- ▼ 7:
 - 0: "Death"
 - 1: 0.0548969962020946
- ▼ 8:
 - 0: "Political career"
 - 1: 0.05294049948210381

Colonization of Mars

From Wikipedia, the free encyclopedia

Mars is the focus of much scientific study about possible **human colonization**. Its surface conditions and the presence of **water on Mars** make it arguably the most **hospitable of the planets in the Solar System**, other than **Earth**. Mars requires less energy per unit mass (**delta-v**) to reach from Earth than any planet except **Venus**.

Permanent human habitation on a planetary body other than the Earth is one of science fiction's most prevalent themes. As technology has advanced, and concerns about the future of **humanity on Earth** have increased, the argument that **space colonization** is an achievable and worthwhile goal has gained momentum.^{[1][2]} Other reasons for colonizing space include economic interests, long-term scientific research best carried out by humans as opposed to robotic probes, and sheer curiosity.



An artist's conception of a human Mars base, with a cutaway revealing an interior horticultural area

Sections you can add

[Relative similarity to Earth](#)

[Differences from Earth](#)

[Conditions for human habitation](#)

[Radiation](#)

[Transportation](#)

[Equipment needed for colonization](#)

[Robotic precursors](#)

[Mission concepts](#)

[Economics](#)

[Possible locations for settlements](#)

[Planetary protection](#)



[Edit to add new sections](#)

Thanks!

شكر!

Mersi!

- <https://research.wikimedia.org/>
- IRC channel #wikimedia-research on Freenode
- If you like research:
https://www.mediawiki.org/wiki/Wikimedia_Research/Showcase
- Leave me a message at
[https://meta.wikimedia.org/wiki/User:LZia_\(WMF\)](https://meta.wikimedia.org/wiki/User:LZia_(WMF))
- Email: leila@wikimedia.org

Now is your turn!
**What are the
challenges you face in
reducing knowledge
gaps in Africa?**