Audiences x 3 / 5 Year Planning Dec 4 2018

Experience	**Form Factor**
	Rich Content
	Contributors
	Customization
	Discovery
Augmentation	**Content Curation**
	Content Generation
	Governance
	Machine Translation
Scale	**Community**
	Content
	Resilience
	Ubiquity
Tools	**For Developers**
	For Organizers
	For Moderators
Culture	**Inclusion**
	Language
	Content Gaps
Trust	**Reliability**
	Accountability
	Transparency

Audiences x 3 / 5 Year Planning 2018

experience

The following are draft position papers exploring aspects of the Wikipedia user experience, based on insights from Community and WMF experts, as well as synthesios of secondary research.

Experience Overview 00

Wikimedia projects have become important fixtures in the infrastructure of knowledge sharing on the internet. However, our share of media interaction and consumption is shrinking as new populations come online, new platforms arise and the internet becomes multi-modal, more interactive, and more social.[1][2]These changes challenge both our consumption and contribution models, and to date, we have struggled to adapt to the opportunities that have arisen and the changing expectations of our users.

A user's expectations of quality are shaped by the totality of their digital experiences. To match these expectations, we have to match the quality of other experiences that users are exposed to. The simplest way to do that is to re-use and refine patterns, methods, and mental models of popular platforms.

These expectations extend to the media types they engage with. Our platform does support diverse media types including video, audio, images, and animated GIFs, and allows them to be mixed in in a single document. However, the experience of uploading and consuming this media does not match the use of media on other major informational platforms. Moreover, our content is saved and presented as a single blob of mostly text, and for some forms of knowledge, such as explanatory knowledge, we do not provide a space or tools to generate, curate or engage with that form of knowledge.[3]

On the contributors' side, Wikimedia projects compete with modern platforms that provide gentler on-boarding and guidance to new users. The competing platforms provide rich, multimedia editing tools and emotionally reward their users with explicit gratitude, meaning, and status.

Additionally, user expectations are rapidly growing with regard to tailored experiences. Software is becoming more aware of individual user needs. The likes, dislikes, and personal preferences of users are vital considerations for modern software design. We distinguish between personalization, in which a system uses what it knows about the user to determine a person's experience, and customization, which empowers users to control their experience. Through customization we can provide tailored experiences, without sacrificing our values or principles.

Data adaptability and content structure are required for the creation of modern user experiences across form factors. Users should be able to engage with Wikimedia, as consumers and creators, in the diverse variety of form factors that are the contemporary internet. And the platform must provide the flexibility to build new experiences for emerging form factors.

Finally, discovery models will be key, as will having captivating content people want to

Experience Overview 00

discover. But that discovery process must be proactive on our part. A large and growing body of research supports a key product theory about today's media - content finds the consumer.[4][5]

Our current products severely lack user awareness and interface customization for the vast majority of our potential audience. We can remedy this by following some basic modernizing principles in our user experience and development processes;

- Embrace a “factoid” paradigm; a lot of people still want to read long-form content, but a lot of people don’t.
- Go where the people already are and utilize platforms they already like.[6]
- Provide suitable content format alternatives for subjects that are not well-served by long text or require advanced levels of prior knowledge.[3]
- Purposely become a tool that empowers others to create, promote, and remix knowledge-based content in many formats.[7]
- Separate the advanced editing experience from the reading and basic editing experiences.
- Provide easier customization of information and interface to match individual needs.
- Use user-centered design to meet consumer expectation.[8]

By understanding our users needs and expectations we can modernize our products, and provide a user experience that informs and delights.

Examples

Apps Explore feed

Page previews

Popcorn video editor

Content APIs

Areas of Impact

Wikidata[9]

Commons[10]

iOS and Android apps[11]

Wikipedia[12]

Key External Factors

Social Platform dynamics

Demographics

Technological waves[13]

Google[14]

References 00

1. https://www.slideshare.net/adobe/2018-adobe-consumer-content-survey 2018 Adobe Consumer Content survey contains multiple data points that describe what people expect from digital content experiences
2. https://qz.com/333313/milliions-of-facebook-users-have-no-idea-theyre-using-the-internet/ Research indicates that millions of users say they use Facebook, but not the Internet. “In their minds, the Internet does not exist; only Facebook.”
3. Learning Styles: Concepts and Evidence” - Research shows that there’s little evidence supporting the popular idea of catering to “learning styles”, however; “differences in educational backgrounds can be a critical consideration in the optimization of instruction. [..] student’s prior knowledge is bound to determine what level and type of instructional activities are optimal for that student” and “the optimal instructional method is likely to vary across disciplines. For instance, the optimal curriculum for a writing course probably includes a heavy verbal emphasis, whereas the most efficient and effective method of teaching geometry obviously requires visual–spatial materials.”https://www.psychologicalscience.org/journals/pspi/PSPI_9_3.pdf
4. https://academic.oup.com/joc/article/68/3/636/4972617 “I Just Google it”: Folk Theories of Distributed Discovery, is fantastic and very recent (June 2018) research on how people find content on the Internet.
5. https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185 Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics: “The news-finds-me effect stems from individuals' perceptions that a) they are well informed about current events despite not purposely following the news, because b) the important information ‘finds them’ anyway, through their general media use, peers, and social connections.”
6. https://meta.wikimedia.org/wiki/Research:Student_use_of_free_online_information_resources/ResultsJonathan Morgan’s 2015 research on free online learning resources used by students
7. http://www.opensym.org/wp-content/uploads/2018/07/OpenSym2018_paper_30-1.pdf “Sharing small pieces of the world”: Increasing and broadening participation in Wikimedia Commons - recent research and interviews with people who use other image sharing platforms but aren’t heavy Commons users
8. https://www.nngroup.com/articles/aesthetic-usability-effect/ The Aesthetic-Usability Effect: “Users are more likely to want to try a visually appealing site, and they’re more patient with minor issues”
9. Wikidata can potentially provide the common vocabulary for many organizations/sites to use for referencing their content so it is discoverable by both humans and algorithms
10. Commons has to change (or be used differently) in significant ways, primarily because its user experience is far outdated and not at all what users of modern commercial multimedia tools want.
11. The apps are already utilizing many modern experiences, but further integration of social media, discovery, and multimedia will be needed.
12. Wikipedia’s article consumption experience is adequate, but there is room for improvement. Additionally, an improved editing/contributing user experience is an opportunity for attracting more people to enrich existing pages and become regular contributors.
13. The future is hard to predict, but there are definite tech trends in place today that aren’t going anywhere in the next 5 years. According to the 2018 Adobe Consumer Content survey, the top devices consumers expect to use most in the next 5 years include: Home Entertainment Streaming Devices (82%), Smart speakers/home assistants (64%), connected home appliances (44%), and wearable devices (42%)
14. Google controls many of the most powerful content discovery tools on the planet.

Experience 01

form factor

A significant challenge in making all the world's knowledge accessible to all the world's people will be to ensure that it is optimized and future-proofed for a rapidly-evolving digital consumption environment. The term Form Factor generally refers to the various entry points, devices, channels and formats that define a digital product offering. In the context of Wikimedia, form factor will include (a) the variety of devices that Wikimedia content can show up on now and in the future; (b) the size, and flexibility of the content itself. Such explicit form factor considerations are the main focus of this paper, however, other implicit aspects of form factor must be considered as well. Partnerships, for example, may require Wikimedia content to be adapted to, or deeply integrated into, third party products with assistance and guidance from the makers of those products - and several of the explicit considerations suggest this type of partnership. Likewise, ideas such as making Wikimedia content available as a utility, or layer on top of the Internet, is another implicit example of form factor, and explored in many of the concepts described in the paper.

When we talk about Form Factor we're talking about how to ensure that Wikipedia content is useful, optimized, and future-proofed for a rapidly-evolving digital consumption environment. For the purposes of this discussion, the term refers to (a) the variety of devices that Wikimedia content can live on now and in the future; (b) the size and flexibility of the content itself - we'll refer to these as *explicit* form factor considerations.

But there are some additional *implicit* form factor considerations as well:

- Partnerships: where Wikimedia content is adapted to, or deeply integrated with, third party products with assistance and guidance from the makers of those products

- Many of the topics in this document imply (but don't necessarily require) the existence of this type of partnership.
- Utility: a layer on top of the Internet. This would be enabled by Item 2, and is implied in many of the concepts listed later in this document.

This paper will focus mainly on the *explicit* aspects of form factor.

Sections

Experience 02

Devices

The future will bring new devices and screen sizes. Wikimedia content should be easy to access and easy to use on every device

In the last five years alone we've seen a remarkable rise in the number of devices people use to consume internet content. In just three years, the Apple Watch has become the best-selling wearable device in the world and a common sight. In ten years, smart appliances have become a fixture in millions of homes. Just recently, Amazon announced a new line of Echo devices boasting its Alexa technology. Most of these devices only have audio capabilities, but some include small screens as well.

Amazon's Echo is the most popular smart speaker system and has already reached 10% of US respondents in the Digital News Report. Amazon's Alexa (which powers Echo) can already read Wikipedia articles by using text-to-speech technology, but its ability to hone in on specific facts within that article is limited.

In the near future, we'll see other competing devices come onto the market. In addition to text content, these devices will need audio content to play. If I ask Alexa to play Franklin D. Roosevelt's famous "Day of Infamy" speech, I'm prompted to buy it from Amazon Music (even though it's in the public domain). Other devices, without their own massive digital storefronts, will want to be able to use that kind of content from an openly licensed source. We should ensure that we are that source.

We should also put focus on optimizing all Wikimedia content for a range of devices with screens. Home entertainment devices are a major new platform for consumption of media. 82% of consumers expect such devices to be their most used devices in the next 5 years.

Desktop experiences (and even mobile experiences) are becoming less relevant. Even though mobile usage is still high, 2018 saw the first global decline in mobile sales and app installs and opens are in a downward trend too. Wikimedia properties will be left behind without a viable platform for atomic content that can adapt to these devices..

One exception, however, may be wearable devices. The current top wearable is the Apple Watch, but with its limited capabilities and tiny screen size there may not be much we can do there.

Takeaways and things to do:

Tactical To Do's

- Decide that it is in our best interest to influence, perhaps even control, the user experience that Echo and other device users have with Wikimedia content. It's not "our" content, but we are the stewards of the systems used to create and disseminate that content, and therefore it's our responsibility to

Experience 03

ensure those systems are used optimally and appropriately
- Encyclopedic content by itself, while useful and what we're best known for, isn't all we have to offer. We have a wealth of imagery, video, audio, and instructional content that better fits the audio and visual strengths of new devices.

Technical To Do's

- Create tools and APIs that are purpose-built for the audio-based smart speaker/home assistant experience
- Improve our backend tech and documentation for thumbnailing uploaded images. We recently had an issue with the Amazon team regarding this (they were trying to download size-appropriate images linked to Wikipedia articles for the Echo Show, but were requesting files that were too big, too often)
- Improve our support for common multimedia formats, especially audio. We now have MP3 on Commons, but we should take initiative in making sure existing files are available in that format.
- Utilize Structured Data to make multimedia easier to find and easier to associate with content from a number of sources

Content

Wikimedia Content will have to adapt to accommodate different user needs

Currently, our flagship project, Wikipedia, specializes in long-form in-depth content. This should definitely be considered a strength, and one that serves most of our current user base well. Wikipedia's brand is currently (and probably solely) centered around accurate, informative, long form information.

But having only long-form information can be a problem..

Moving forward, it'll be critical for our content to adapt to shifting habits and the expectations of new audiences we begin to reach. As form factors change, the long form and complex nature of Wikimedia content may start to undermine the value of its accuracy and extensive coverage.

According to the 2018 Adobe Content Consumer Survey (US only), when content is too long, 47% of consumers stop reading, and 23% switch devices.

Our content needs to adapt to different contexts that reflect how people actually use social media and messaging

PROPORTION THAT USE FACEBOOK MESSENGER AND WHATSAPP FOR NEWS

Selected markets

Messaging Apps	Greece	Norway	United States	Australia	Finland	Argentina	Hong Kong	Malaysia
FB Messenger for news	22%	11%	7%	11%	5%	9%	8%	12%
WhatsApp use for news	4%	2%	4%	10%	10%	37%	38%	54%

Experience 04

The social messaging use case is a very important one to focus on because direct communication tools like Whatsapp and Messenger are on the rise for news, [2] particularly in emerging markets, Asia, and South America. This phenomenon barely exists in the US, but it is a huge content consumption driver in other countries.

Wikimedia projects need to become stronger here, and that means having content and systems that naturally fit with messaging usage patterns. This could include possible future features like:

- Chatbots - Imagine a Wikipedia bot on WhatsApp, Facebook Messenger, or Telegram that actually answers questions and links to citations when you talk to it. It could even be a form of interactive instruction that guides you through a topic based on your prior knowledge, available time, etc.
- Media bots - "Hey Wikipedia, show me video of World War II". Instead of taking you to a link, the bot can put customized/curated video clips right into the chat
- Have link previews that show the fact you're interested in - automated page links with <title> tags that have the text of the specific fact you want to share (so instead of seeing "Patrick Stewart - Wikipedia" as the link preview, users can see "Patrick Stewart was knighted on June 2, 2010"

Adaptive Learning [3] can help readers customize their path through content. Imagine a world where we could ask a reader what level of knowledge they have for a Wikipedia page topic, and then automatically reconfigure the facts and citations to fit the reader. For Readers with advanced knowledge, the page can automatically can skip the basics, while readers with basic knowledge can be presented with an "explain like I'm five" version. This approach also provides an on-demand customization experience which helps us avoid privacy concerns.

The Article vs. The Fact

All of the scenarios above are problematic for us right now, mostly because of one thing - our core, fundamental element is the article, not the fact.

As mentioned in the Discovery document, major Internet players like Google and Facebook are already grabbing our text content, chopping it up, and presenting it in factoid-sized chunks. We currently don't have influence over this process and the resulting user experiences, largely because we don't have any facility that we can point to and say, "do it this way, it's better and we've already done the work for you."

There are several ways we might achieve this "atomizing" of articles:

A. Automatically break up the entire article into elemental parts (sentences/passages perhaps)
B. Take the top *5* most important elements/sections from each article and atomize that (although it's unclear at the

Experience 05

moment how we would identify those top 5 elements).

C. Have the community decide which elements should be atomized for each article (essentially a new editor function/workflow)

Of these three options, option (A) is probably the most flexible, most likely to scale, and the most likely to fit every possible need we may have in the future.

The Reasonator project has made an attempt at "prettifying" Wikidata facts into human-readable form with mixed results. With better technology and techniques, we may find a better automated fact-generated system in the near future.

Takeaways and things to do:

Tactical To Do's

- Embrace a "factoid" paradigm; a lot of people still want to read long-form content, but a lot of people don't.
- Encourage and enable quick answers to discrete questions. ensure those systems are used optimally and appropriately
- Explore how we can optimally serve content in short-form environments like social messaging

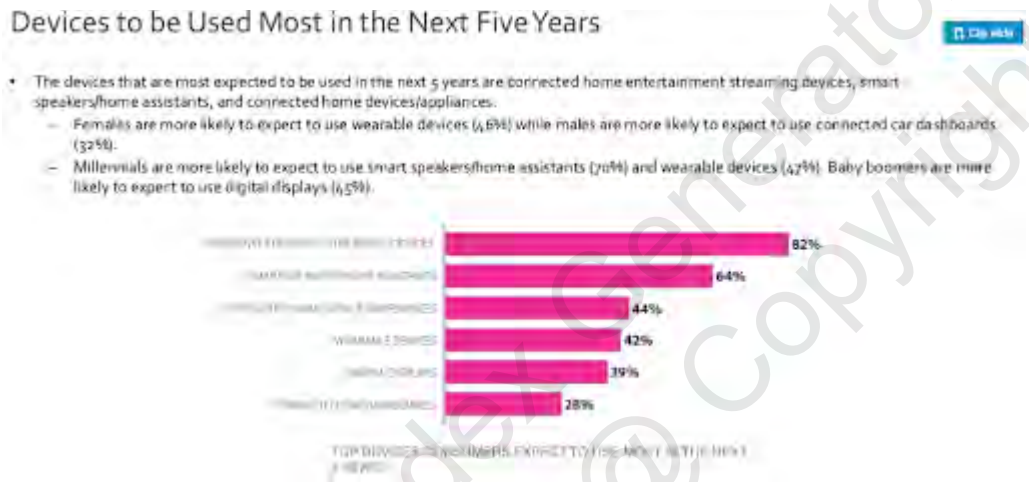
Technical To Do's

- Architect a methodology for breaking up, storing, and serving our text content into individual, atomic elements that can be paired with citations
- Explore content adaptation architecture so that pages can change their form based on context and/or reader needs
- Build our own social media/messaging APIs and improve our integration with others
- Explore automated video/audio file creation (combining multiple clips or images into one and sending it off to the user's touchpoint)
- Utilize Structured Data to help put it all together. This could potentially be Wikidata (or a new feature on Wikidata), or an entirely new tool altogether since some fact formats just don't easily fit into Wikidata right now.
- Improve our on-wiki search capabilities to enable "factoid" searches on our sites just as we would on other platforms. This could include integrating structured data into search to ensure semantic matches, improve accuracy, and enable highly focused searches (see structured data search on Commons as an early example).

Experience 06

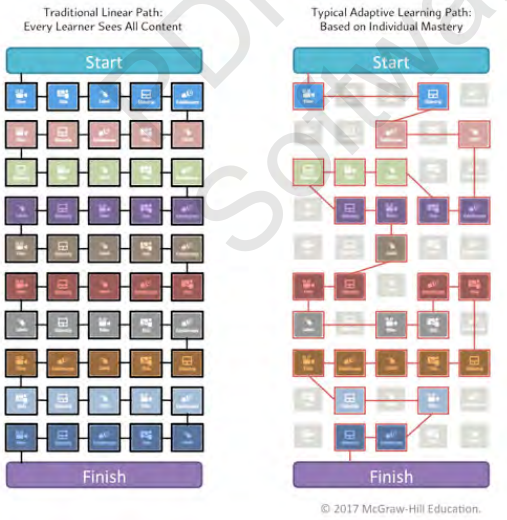
Notes

[1] 64% of respondents in Adobe's Content Consumer survey said that smart speakers/home assistants are devices they expect to use most in the next 5 years.



[2] http://www.digitalnewsreport.org/survey/2018/the-rise-of-messaging-apps-for-news/

[3] Adaptive Content Learning provides a possible framework for the future



Experience 07

Sources

R. Isler Research and Insights, Other contributors[1]: A. Baso, C. Gauthier, A. Hollender, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

1. https://www.slideshare.net/adobe/2018-adobe-consumer-content-survey
2. http://www.digitalnewsreport.org/survey/2018/the-rise-of-messaging-apps-for-news/
3. https://meta.wikimedia.org/wiki/Reasonator

Version History

VERSION	DATE	NOTES
DRAFT	12.3.2018	

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

rich content

The Open Knowledge movement encompasses much more than Wikipedia articles. Reading long-form text is not currently the only, or optimal, way people choose to gain knowledge. In fact, a recent study shows that only 20% of Wikipedia consumers are in-depth readers, no matter what language you consider.

And while Wikipedia is mainly associated with long-form, informative, encyclopedic text content, several Wikimedia projects already offer much more than that. **Commons**, while flawed, is an established source for freely licensed multimedia files. **WikiVoyage**, though largely unknown, is full of rich and useful crowdsourced travel content. **Wikisource** has a small but dedicated community of transcribers, translators, and archivists who combine imagery and text into useful digital reproductions of old publications.

Our existing projects already offer rich opportunities to expand beyond encyclopedic content and give our users useful and fulfilling experiences, so it won't be a stretch for us to continue to explore all types of media and formats to accomplish our goals.

Sections

Experience 02

Roadmap for the future?

The National Geographic Society is one of the most well-known and successful global non-profits. They began with a magazine, which stood as their only media platform from 1888 until 1964 when they aired their first television content on CBS.

Today, in partnership with 21st Century Fox, the Society still operates the magazine that got them started, but they have also branched into other forms of media including TV channels, films, a website that features extra content, worldwide events, and other media operations. After decades of exclusively being a magazine brand, today "NatGeo" is truly a successful omni-channel presence. We should seriously consider using this approach as well.

An Omni-channel approach

Disruption. It is perhaps the one word that best describes what happened to the print industry in the past 20 years. Technology didn't kill print, but it certainly gave it a mortal wound. Disruption wrecks companies, and the best defense against it is diversification.

What would National Geographic be today if they had not ventured into other media? What if they'd remained strictly a magazine company? In the early 1990s the company's flagship publication (National Geographic Magazine) boasted 15 million subscribers. That number was estimated to be closer to 3.5 million in 2015.

Although it is possible that National Geographic would still be around if they'd stuck with magazines, they wouldn't have been able to do so without a massive restructuring, and there's no question that their current setup increases their outreach to millions more people than they could ever hope to reach with just a print magazine.

How does all this apply to us?

It is doubtful that we'll see another online encyclopedia rise to compete with Wikipedia, but that's mainly because the encyclopedia business isn't exactly a growth industry these days. Sooner or later, something will disrupt our model. It might be that our donations dry up, or larger companies grab all our data and start their thing (think Wikitravel but in reverse), or Artificial intelligence algorithms perfect the automatic creation of articles based on news. Something's going to happen, and it's in our best interest to diversify so we have defenses against the inevitable and the unforeseeable.

Experience 03

Three types of knowledge

Before we focus on the content formats we should consider for the future, let's talk about "types" of knowledge. Our vision statement mentions "a world in which every single human being can freely share in the sum of all knowledge", but what does that really mean? There are least 3 types for us to target:

Factual

This is already a strength of Wikipedia and its straightforward, neutral, "citations needed" format.

However, the same things that are Wikipedia's strengths are also its weaknesses. They make it hard to find new editors who want to work in the intimidating and often conflict-laden processes of the Wikipedia world. Few people feel like factual experts, and even fewer feel like vigorously defending their claims.

Instructional

This is a weakness for Wikimedia projects. Wikipedia is very good at describing things but very bad at telling you how things work.

You can see how bad we are at this with a simple experiment. Go to Google and search for "rocket". The Wikipedia article for rocket shows up very early in the search results. Now try "how do rocks work" and see what happens (hint: you're gonna have to go to the dreaded page 2 of search results).

This isn't just an artifact of poor SEO - the rocket page really does not do a good job of simply explaining how rockets work.

Experiential

This is an important area where we again are lacking useful content. People want to know what it's like to be X or do Y or visit Z. Knowledge of someone's else's experience is just as valuable as discrete facts, and a key element of getting humans to understand each other.

Now, let's talk about the content formats we can use to convey knowledge - Video, Audio, Images, Text, and Interactive.

Video : The Elephant In the Room

"What are we going to do with video?" is a question often asked not only within the Foundation, but within our community as well. Video has become the most preferred learning method for the majority of Millenials and GenZ. In the next 3-5 years it will be crucial for us to expand the Foundation's video capabilities. [1]

Factual Video: "Explainer" videos are a popular genre. We should give users the ability to create their own videos explaining certain topics or giving highlights of their favorite Wikipedia articles.

Instructional Video: Even the most talented writer in the world couldn't write an article to teach you sign language. But it would

Experience 04

only take a decent teacher to make a useful sign language course with video. That, in a nutshell, is the advantageous power of video when it comes to instruction.

YouTube, solely through the power of its user community, has become a prime hub for learning how to do things. Although much of this content consists of frivolous things like makeup tutorials and video game tips, there is a large body of knowledge-based instructional content, from life skills like cooking to professional skills. It is even used by surgery trainees.

The Wikimedia Foundation has the unique opportunity to learn from the YouTube model and improve it with openly licensed instructional video content that is translated, vetted, and highly curated by our communities for joyful consumption by all.

Experiential Video: Imagine free, openly licensed video content that provides detailed and compelling stories of the experiences of a wide variety of people, cultures, events, and walks of life. Think "Wikipedia originals"

Documentaries are a time-tested form of information sharing that can raise awareness and enter the social consciousness. It makes natural sense for Wikimedia to explore this medium, but not in the typical way.

Other organizations already do documentary content; the BBC, Discovery Channel, and even National Geographic and Smithsonian. But all of these organizations focus on telling stories from a limited number of perspectives. Our strength is the Wiki way - allowing multiple perspectives and contributors.

Fortunately, we don't have to guess or theorize about what this might look like. There's a perfect real-world example in the form of Winter on Fire, an Oscar nominated, Emmy nominated, feature length, Netflix original documentary that chronicles the deadly anti-government protests in Kiev, Ukraine that took place in 2013.

Winter on Fire utilized footage from dozens of sources, including cell phones used by protestors during the conflict

Winter on Fire had 28 credited cinematographers, using video footage captured by ordinary people who were on the ground during the conflict. In many ways it was a crowd-sourced film.

Winter on Fire utilized footage from dozens of sources, including cell phones used by protestors during the conflict

Experience 05

From the director of the film:

"We got footage from people's phones, from GoPro cameras, from TV crews, from wherever we could. Without these volunteer cinematographers and the variety of technology available, it would have been impossible to document the movement."

The parallels to the Wikimedia processes and movement are clear.

Tactical & Technical Takeaways

Tactical To Do's

- Embrace video as a cornerstone of our media strategy going forward
- Invest in architecture, policies, and community members that support online editing tools for video
- Invest in architecture, policies, and community members that support online editing tools for video
- But also support static, immutable content that is created by collecting disparate pieces of Wiki content and combining it into a finished "work" that is greater than the sum of its parts (ala Winter on Fire)
- Answer an important question: Do we have to limit ourselves to the Internet as our only distribution method?

Technical To Do's

- Figure out how we're gonna serve and store all this dang video! :)
- Explore and invest in collaborative online video editing solutions like Popcorn.js
- Devise tools that enable admins to vet and moderate video content

Audio - the forgotten format

When most people think of audio content they think of music, but there's so much more to hear. Text-to-speech technology, in the form of personal assistants like Siri and home smart speakers like Amazon's Echo, are creating a new renaissance for audio content. And audiobooks are more popular than ever.

In their 2018 Consumer Sales Survey (US only), the Association of Audio Publishers found the following:

- audiobook sales in 2017 totaled more than $2.5 billion, up 22.7% over 2016, and with a corresponding 21.5% increase in units
- This continues the six-year audiobook trend of double-digit growth year over year.
- 54% of audiobook listeners are under the age of 45 (in other words, it's not just for the olds)
- 24% of listeners said they have listened on a smart speaker and 5% said they listen most often on a smart speaker
- The top three activities while listening to audiobooks are: driving (65%), relaxing before going to sleep (52%), and doing housework/chores (45%)

Experience 06

- 52% of people said borrowing from a library/library website was important or very important for discovering new audiobooks. 43% of listeners said they downloaded an audiobook from a library

If audiobooks are a growing form of media as a "port" of a type of long-form text content (books), perhaps we can adopt that model as well, but for Wiki articles. And perhaps we can serve as the library source for audiobooks in areas that simply don't have a lot of libraries.

Factual Audio - factoids in audio form, and beyond: What does the fox say? We can answer that question with recordings of fox sounds that are openly licensed and freely available as part of a Wiki database of animal sounds.

"Alexa, in what year was Nelson Mandela released from prison?" Audio facts like that are already being provided by products using our platforms, however, we can enhance that work with "links" to audio files. For example, with structured data we can link topics to available media on that topic, so when someone asks about Mandela getting out of prison we can programatically suggest the user listen to the public domain speech Mandela gave after he was released.

Instructional Audio - audio can teach you more than just French: Imagine a young person in India learning to speak the indigenous language of their region, or an art student in Canada listening o a public domain podcast that is produced by a group of women in Quebec and focuses on traditions of Inuit art.

And, yes, we could also have language courses so you can learn French.

Experiential Audio - listen to someone tell their story: Oral histories have been a hot topic within the foundation for some time, and they are just the kind of experiential knowledge that we've largely neglected. However, oral tellings don't just expose us to mythology and history of esoteric cultures in far off lands. They can be used to give us deep connections to topics most of us only scratch the surface of.

Imagine a CC0 collection of audio recordings from Holocaust survivors recounting their personal stories.

Tactical & Technical Takeaways

Tactical To Do's

- Don't forget audio! It is a flexible and easy to use file format
- Invest in obtaining/acquiring rich histories in audio format
- Explore what it would take to become a repository of all the sounds in the world (machine, animal, and other)

Technical To Do's

- Consider ways to use data to link topics/facts to available audio related to that fact

Experience 07

- The audio playback experience for our files on the web is terrible now. Commit to fixing it.
- File formats. We have MP3 support now (patent expired). What else might we need to do to provide good streaming audio quality?

Images: a long way to go

Commons falls far short of expectations for a modern image sharing platform. There. Now that we have that out of the way, let's talk about the kind of image content tools we should have going forward.

Factual Imagery: Photos are a type of fact that ordinary people feel comfortable adding. What exactly does the Bambino cat breed look like? It's a simple question with a simple answer (just show me a picture of one), but you won't find it on Wikipedia or Commons because we don't have an image for it.

Commons has numerous and varied knowledge gaps, and doesn't get close to being a great tool for even illustrating Wikipedia articles, let alone being a source for visual reference for all knowledge.

A large part of this problem is that Commons itself provides a poor experience as a destination. Its UI and design are outdated, it lacks features people have come to expect on platforms like Flickr and Instagram, it has almost no social component, and its processes are even more obscure than Wikipedia's.

Put simply - we can't get great, comprehensive image coverage because people don't want to use the site.

So the first step is to change that. We've enlisted the help of gurus like George Oates to give us outsider perspective on what's wrong with Commons, but changing Commons is an expensive and time consuming task (as anyone on the SDC project can tell you). The more prudent approach seems to be to forget the idea of Commons as a destination for the masses and instead use it as a piece of infrastructure maintained by people who like that sort of thing.

Imagine a world where we do image campaigns like Wiki Loves X every week instead of every quarter. Imagine a massive social media campaign like #WikiLovesWednesday, where every Wednesday we ask the whole world to donate photos of that week's topics.

But we can't do that now because directing millions of people to Commons would make them hate us and ask why we made them go to that terrible, terrible place and they're never coming back.

We propose a New Commons, which would include the creation of a purpose-built, user-friendly new presentation layers on top of Commons. This is still just in the concept phase, but has gotten some support so far.
Instructional Imagery: Images are often

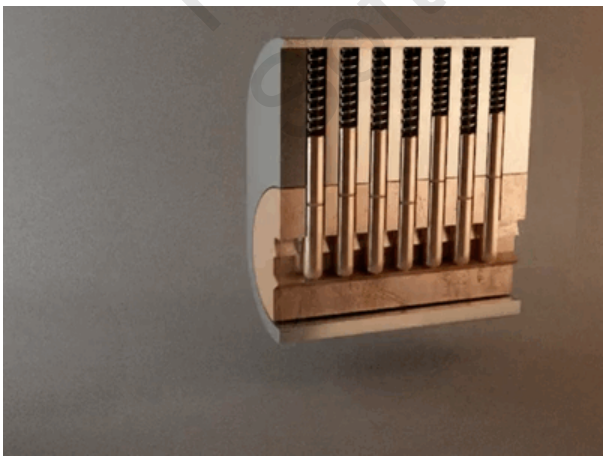
Experience 08

much more efficient at explanation than text (and usually don't need translation). Let's say you're trying to learn how ordinary door locks work.

Which experience do you find more elucidating? This text from the Wikipedia entry....

"The pin tumbler lock uses a set of pins to prevent the lock from opening unless the correct key is inserted. The key has a series of grooves on either side of the key's blade that limit the type of lock the key can slide into. As the key slides into the lock, the horizontal grooves on the blade align with the wards in the keyway allowing or denying entry to the cylinder. A series of pointed teeth and notches on the blade, called bittings, then allow pins to move up and down until they are in line with the shear line of the inner and outer cylinder, allowing the cylinder or cam to rotate freely and the lock to open."

Or this GIF explaining How a lock and key work?



Experiential Imagery: Everyone can be a photojournalist. Social photo sharing is a common activity now, with people sharing their personal experiences of travel, dining, and events both mundane and fantastical. The good news is that the Wiki way - where everyone and anyone is invited to contribute - nicely meshes with broader internet usage patterns, and other hugely popular platforms have already trained the masses to always have their smartphone camera at the ready.

Additionally, we can take a page out of the National Geographic book and encourage people to capture ordinary life in extraordinary places. We've seen some good photo essays come out of Wiki Loves Africa, but they're wasted on Commons, which simply doesn't have the reach or format to really showcase this content.

With the right tools and design and a motivated community, we can do what many photojournalistic outlets do, but at a scale they can't achieve.

Tactical & Technical Takeaways

Tactical To Do's

- Invest in at least one (possibly more!) welcoming, useful, and usable place for people to share/donate their images
- Animated GIFs have made a strong comeback. They are also fantastic for informational and instructional content. Support and explore the idea of static

images having less prominence in the future.
- Encourage experiential storytelling through imagery

Technical To Do's

- Make frictionless mobile image contribution a priority

Text : Fix it

Wikipedia isn't perfect, but it works. Its success is undeniable, and it will stand in history as a world-changing project. If you're the kind of person who loves deep dives into complex topics, and you don't mind spending time with text that can be challenging, Wikipedia's current format totally works for you. There are certainly parts of its formula that work and should be left alone.

But there are other Wikimedia projects that are heavily text-based and far more accessible than Wikipedia. They are in drastic need of some attention and fixes.

Factual Text: Facts matter, but we don't have all of them yet. Wikisource is a fantastic...well...source, for all sorts of information that just wouldn't work on Wikipedia. Want to read a biology text for kids that was published in 1875? It's there! Want to read the One Thousand and One Nights story in the original Arabic? It's there!

There are countless rich and engaging pieces of public domain or freely licensed text content out there. Some of them are digitized already, but many are not. A lot of stuff is locked away in archived books sitting in vaults (Charlotte's Culture Outline touches on this). There's a large opportunity for us to facilitate the process of freeing this text and bringing it to the people.

Sometimes, however, the content we want won't be under a license we like. Recent out-of-print content can still be under copyright, but perhaps there's room for us to fund the process of purchasing the rights for valuable content and then releasing it to the public domain or CC0 license.

Instructional Text: Video is king, but text is still like...a duke or something. Video is still the most engaging and powerful medium if you want to tell someone how to do something, but video is not always easy to create, much harder to edit, and not as portable as well-written text instructions.

As we've discussed earlier, Wikipedia is terrible at instructions. But Wikisource and Wikivoyage are pretty good at it! You can find an entire book on how to teach yourself Chinese on Wikisource, and Wikivoyage has a wealth of content like how to buy a kimono in Japan.

Experiential Text: Good writers can make text descriptions come alive. This is another area where Wikisource and Wikivoyage can shine if we let them. Wikipedia's neutral point of view rule makes the content fair

and less prone to bias (although not impervious to it), but that rule also makes the content bland and no fun to write.

Investing in our other projects with less stringent content rules will help attract new readers looking for something less dry, and contributors who can really write and want to use that skill for a good cause.

Tactical & Technical Takeaways

Tactical To Do's

- Spend money on Wikisource
- Spend money on Wikivoyage
- Spend money on acquiring rights to books, articles, and other text content we want (then make it free)

Tactical To Do

Much of Wikisource is held together with string, gum, and hope. Volunteer developers have kept it working with many disparate hacks over the years. We'll have to either commit to fixing it or, perhaps even better, create a new platform that is purpose-built for this use case.

Interactive

This is a complex topic, best described in slide deck form. See Future of Commons Deck, Path #3.

Priorities

This is our recommendation for how to prioritize investment in the rich media types described in this paper:

1. The first thing we need to invest in

Notes

[1] The Next Generation of Learners
[2] xxx
[3] xxx

Sources

R. Isley Research and Insights; Other contributors[1]: A. Baso, C. Gauthier, A. Hollender, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

https://meta.wikimedia.org/wiki/Research:Characterizing_Wikipedia_Reader_Behaviour/Human_development_index_and_Wikipedia_use_cases

Five things you need to know about millennial media habits
September 22, 2017 | By Damian Radcliffe, Carolyn S. Chambers Professor in Journalism–University of Oregon @damianradcliffe
https://digitalcontentnext.org/blog/2017/09/22/five-things-need-know-millennial-media-habits/

27/11/2017 Huffington Post
The Media Habits Of Millennials, Generation Z, And The Rest Of Us: In Five Key Chartshttps://www.huffingtonpost.co.uk/entry/the-media-habits-of-millennials-generation-z-and-the-rest-of-us-in-five-key-charts_uk_5a149436e4b0815d3ce65ac5?guccounter=1&guce_referrer_us=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_cs=hkCzYDU0_kr_tHiPSqxiAg

Ana Gotter — October 24, 2017

Desktop vs. Mobile Video Consumption: What You Need to Know

https://www.business2community.com/video-marketing/desktop-vs-mobile-video-consumption-need-know-01937184
Surgical education on YouTube

https://pdfs.semanticscholar.org/fbc3/96b2d3f4fcd3a3844e2e02866992204c6032.pdf

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Experience 01

contributors

In order to reach our 2030 knowledge equity goal it will be necessary to diversify the pool of people contributing to the projects, to expand the modalities of engagement and to open the door to new types of content that can be contributed. That is, it must be possible for the next wave of contributors to provide raw text, speech, images, video and other multimedia formats; to contribute new data in a structured way; and to perform discrete editing tasks to improve existing content. This next wave of editors must also be empowered with the tools necessary to *storify* (assemble and add context to) the raw content uploaded by others. The next wave of contributors must be able to create short form content (e.g. parts of articles) as well as *rich articles*. We're going to need to both make our contributing experiences both cheaper and easier as well as richer and more complex.

Wikipedia is currently a reading *and* editing tool, but in the next three to five years these experiences must be separated and optimized for their purpose. In this near term future Wikipedia will become a reading and recruiting entry point (driven by search traffic, and reinforced by citations across the internet) while another more purpose-built experience will be created to optimize the contributors' experience. This new experience must focus on better supporting the myriad of tasks related to uploading, labeling, editing and monitoring contributions.

Finally, we're going to need to make sure that the moderation experiences around the next contribution modalities and types are going to work for existing and new contributors (satisfying "Risker's checklist"[1]) ensuring that the content creators are happy with the new content created and it meets reader needs for quality and trust.

Sections

Experience 02

Modalities of Content

To diversify the pool of contributors and perspectives we must diversify the modalities of content we accept. This means that we must "open the aperture" and start managing contributions that are both richer and more complex than existing models, but also cheaper and easier.

With diverse modalities comes the new responsibilities of storifying these contributions, creating new types of editing tasks, even as other types of content creation (see augmentation et al) reduce traditional editing tasks.

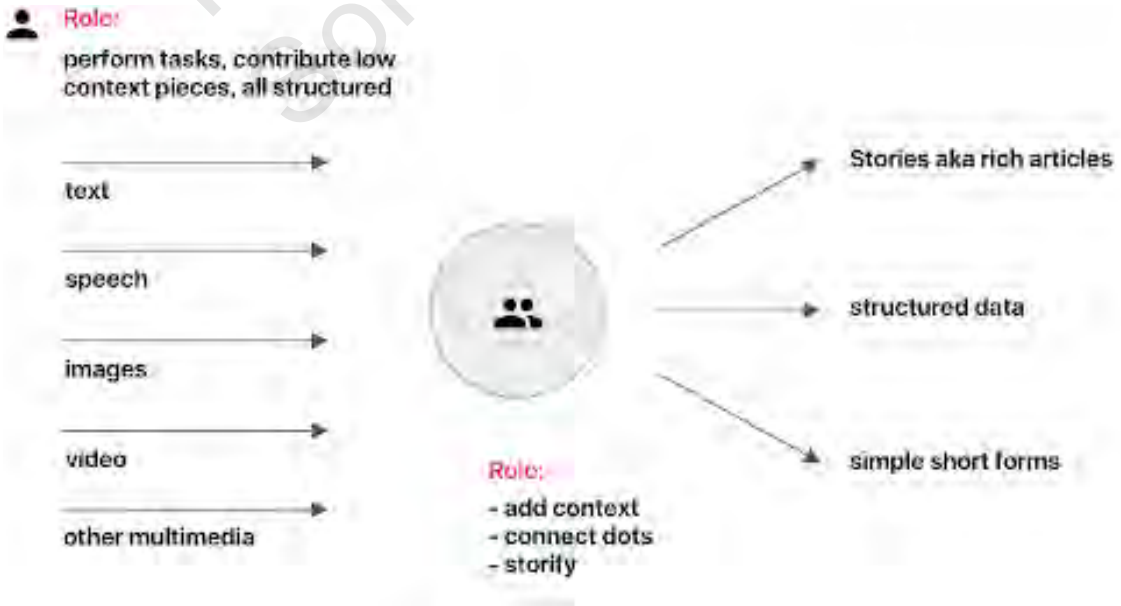
Speech

New methods like speech-to-text are vital for many use cases and audiences:

- Local languages : Languages which are less compatible with technology or where entry is difficult with traditional tools like keyboards or small screens [3]
- Oral histories (could we really write this document without mentioning oral histories?)
- Tools to support voice consumption and navigation.
- Older than average users: The typical user of voice assistant applications is a 52 year old woman [1]
- Mechanisms to augment existing content with audio, animations or other forms of more dynamic content (but not maps)

Video

Our own research has shown there is demand and need for rich content on Wikipedia [citation needed]
Our commitment to open media formats has held us back for years and we may need to find a way to find a compromise between the open source values of the



Experience 03

community and the modern expectations of web users.

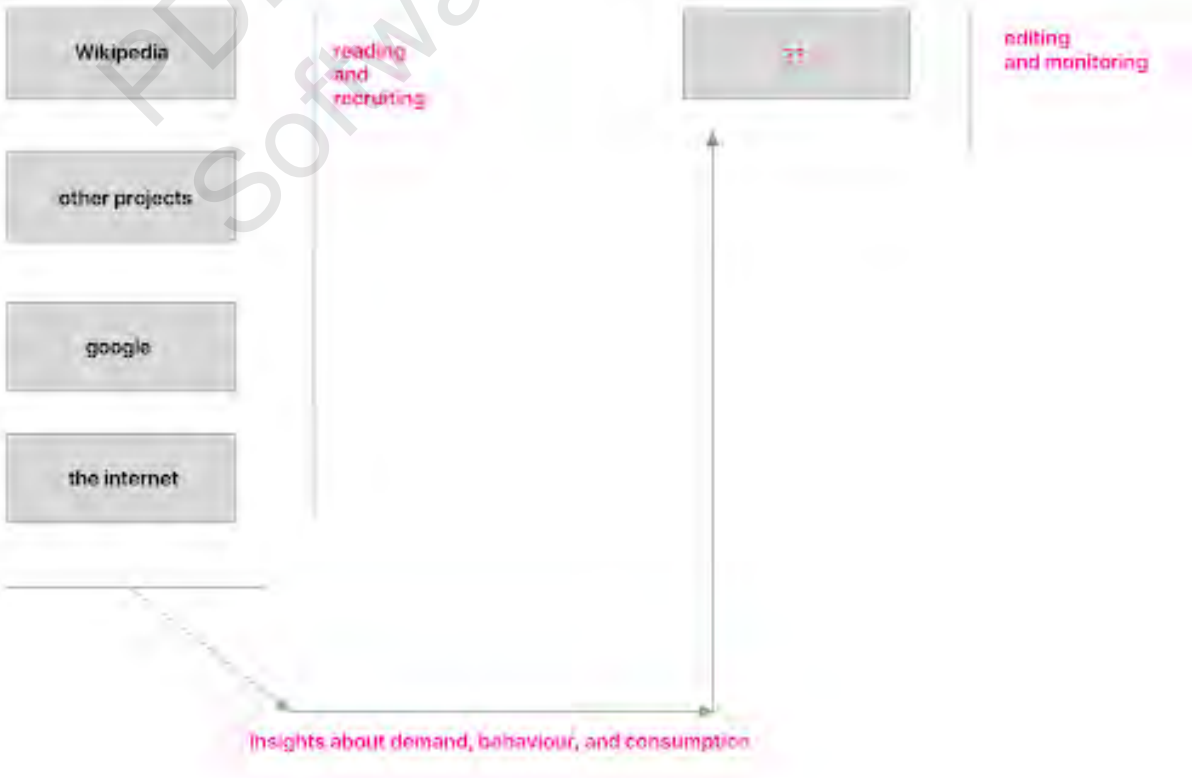
Youtube is a surging platform for procedural learning. Research shows using TAM (Technology Acceptance Framework)[1] framework, the user acceptance of this behaviour is sufficiently high enough to call it a leading place for learning. [2]

- Wikimedia should seek partnership opportunities with Youtube or Youtube like service
- Youtube could serve as a potential bypass to open source policies for video distribution on Wikipedia projects

The contribution experience

We use Wikipedia as reading and editing tool
In next 3 to 5 years, we need to separate these
Reading and editing have drastically different intentions
The funnel problem can be still solved by following method
The wall can still be tore down by surfacing information

We will need to design editing (and moderation) experiences for the rich media experiences described in the experience section. It seems likely that a new type of contributor as well as tools



Experience 04

will be necessary to product the type of content necessary for an omnichannel experience.

Potential framework

Wikipedia and the internet becomes a reading and recruiting place. Part of wikipedia becomes a purely editing place.

These experiences are not desktop or mobile experiences but modes of contribution that are customized to their context. They need to provide a mental map of the contribution process to make the knowledge creation process transparent and navigable to all users, new and experienced.

Machine learning, personalization and customization are fully engaged across these experiences.

The workstation

The workstations is fully customizable to the need and the role of the contributor.
Bret Victor, noted design strategist notes that the current state of digital software as it is, is a medium where you do not manipulate your environment to match your needs. Take your desk, a carpenters workbench or any other working area for example. - Bret Victor. The workstation is similar to an integrated IDE where we double down on the community's use of add-ons and specialized tools.

- Ability to create pipeline (if.... Then.... that...) workflows
- Machine Learning to cut down gruntwork
- Citation suggestions and insertions
- Seamless Integration with media and structured data
- Advance discovery mechanisms to find issues and gaps
- Clear consistent APIs for extensions.

The launchpad

The launchpad is a more intuitive, simpler contribution tools. It's a place for newcomers to get help and see what impact they are making. More experienced users have personalised feeds of things that need to be done.

We will need to provide tools that connect other knowledge related activities such as reading, browsing, researching, discussing, taking photographs, recording sounds, downloading sensor data, etc and make them available for immediate

Power editor experience - The workstation	**New editor experience - the launchpad**
Focus on productivity	Focus on impact
Discovery based on customization	Discovery based on personalisation
Share details of what happened because of others	Share details of what happened because of you
Interface is tailored and workflows are co-designed	Interface is easy to use and workflows are assigned

Experience 05

content creation -- think of a “publish to wikipedia” button that carries a fact and citation information straight to the article. Loose content could also be stored in the workstation for integration into content at a later time.

Authors should always be able to contribute to discussions, review other edits quickly and easily so they feel connected to the community and the project all the time.

contributions for vandalism and spam and identify tasks for later. A necessary part of increasing the number and diversity of contributors is increasing the tasks and contexts where people can contribute.

The contribution reward

Social networks have succeeded because they variably distribute the most evolutionarily compelling reward possible: approval [4]

Our ecosystem does not give consistent reward for participation in the knowledge creation process nor is there any signposting for how to progress in skill and responsibility.

The community has created some mechanisms to do this but it is not particularly accessible and the look and feel is a bit dated. We need to retain the community control but clearly support a clear model of editing and provide mechanisms and rewards that make editing and rewarding and sticky experience.

Experience 08

Notes

[1] https://en.wikipedia.org/wiki/User:Risker/Risker%27s_checklist_for_content-creation_extensions

[1] https://en.wikipedia.org/wiki/Technology_acceptance_model

[1] The Past, Present, and Future of Speech Recognition Technology

https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf

[2] User acceptance of YouTube for procedural learning: An extension of the Technology Acceptance Model

https://www.sciencedirect.com/science/article/pii/S0360131512002229

[3] W. Knight, 2016 Powerful speech technology from China's leading Internet company makes it much easier to use a smartphone.
https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces/

[4] We're Primed to Be Addicted to Social Media
https://zandercutt.com/2018/09/18/were-primed-to-be-addicted-to-social-media/

Sources

N. Pangarkar Research and Insights, Other contributors[1]: A. Baso, C. Gauthier, A. Hollender, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

customization

Customized and Personalized user experiences are the new norm in consumer products, bu are they appropriate for Wikipedia? A Personalization-focused strategy would conflict with Wikipedia privacy policy, product principles and Movement equity goals, but Customization could contribute to greater usability for readers, communities and editors and for this reason must be considered as part of product modernization overall. In terms of the reading experience, the platform should support a set of user-modifiable customization options and a set of community-modifiable customization options (to allow for language-specific and culture-specific preferences). In terms of the editing experience, the platform should support customization based on common usage patterns and contributor activities at scale, and across whole groups of wiki projects. By making customization options part of the platform, it is possible to serve a more diverse set of needs and preferences without forking the main product. This approach will make scaling much more achievable, and the process of integrating new customization features less dependent upon the technical resources of regional communities.

Sections

Experience 02

Customization vs. Personalization

Customization and Personalization are sometimes used interchangeably but are fundamentally different in terms of user experience. **Personalization** is the automatic adaptation of a system to the behaviors and preferences of a user. Because Personalization requires much less explicit input from the user, it is appealing to non-expert users - their experience is automatically tailored without additional effort on their part. But a truly satisfying bespoke experience can only be delivered with a more technically sophisticated system, and requires sign-in, or another form of user identification, and the ongoing collection of user behavior data over time. This approach would be unfeasible and incompatible with the principle of intentional transparency.

Customization refers to the degree to which a user can tailor their use of a product through overt controls and settings. Customization features allow a user to optimize their experience through preferences, defaults and conditionals. But unlike Personalization, Customization requires the user to *take action* in order to have a more optimal experience and hence is most appropriate in products where users are highly motivated to make such adaptations. Customization tends to result in greater satisfaction among power users, and lower satisfaction among among non-power users. [3]

The Reading Experience

Customization For Individual Users

The *Customized* aspect of consumption experiences may not be differentiating or particularly memorable, but multiple studies [2] [3] suggest customization features are a factor in user satisfaction, enjoyment, and perception of usefulness.

In the context of Wikipedia, it is tempting to treat customization for individuals as a matter of agency (i.e. that it is self-evident that a user *should* have control over what they are being shown, how information is presented, and how the context they're in affects the modes of presentation). However while basic levels of customization in reading experiences have been shown to result in quick wins, [x] the considerable effort required to implement them may not result in a commensurate impact. So while the product landscape may have set an expectation of agency over...

- Aesthetics
- Accessibility
- Data control
- View modes for comfort, by screen
- Short form/long form reading
- Implicit topical interests
- Control of notifications
- Abuse filters

Experience 03

such settings will not necessarily result in *sustained* or *significantly greater* satisfaction with the product.

Customization For Communities

Reading habits and aesthetic preferences are drastically different in different cultures. [3] communities should be able to customize the reading experiences for their language wikis drastically as well There is a small subset of contributors who work towards reading audience, (main page designers, maintainers, template designers) We need to give more choices and agency over how they present content to their communities [not edited at all]

when but it at the same time fragments experiences which is a big issue in scaling. Thus customization must be considered in parametric ways within frameworks that are well defined and scalable [not edited at all]

So while Customization *can* be a good thing for readers and communities, in an ecosystem with limited resources, the return on investment (in terms of change in user satisfaction) must be considered against the potential hit to scalability. When should considerations around agency prevail over considerations of scalability? when but it at the same time fragments experiences which is a big issue in scaling.

Reco: Thus customization must be considered in parametric ways within frameworks that are well defined and scalable

The Editing Experience

Customization For Individual Editors

In spite of the fact that customizing how a system behaves is a superficial change in the medium, it has been shown to lead to greater retention and satisfaction. [1]

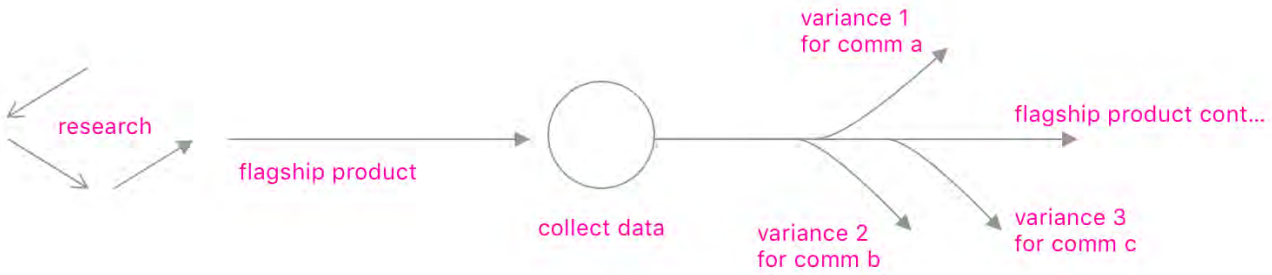
Customization for Communities

Create community personas that show similar behavior, stage, reactions etc.

Build main product conceived by research, which solves problems that are lowest common denominator

Collect information on specific communities on usage of the product

Fork and Modify the main product to match community needs



Experience 04

These will be specific efforts and intervention for target communities to see a particular change in that community

fork when community needs are determined to be different enough to warrant a shift.

Notes

[1] Because Personalization requires a high degree of technical sophistication it would likely only be implemented in communities with a high degree of technical acumen and the developer resources to enable and support it. K. Sweet and K. Wirth

[2] Customization leads to perceived ease of use and perceived ease of use leads to perceived usefulness. A user's perception of control positively affects their attitudes toward the product and the creator of the product. H. Lee and E. Chang

[3] This pattern holds in "low privacy" contexts, but the opposite is true in "high privacy" contexts. S. S. Sundar and S. S. Marathe

Sources

N. Pangarkar : Research and Insights, Other contributors[1]: A. Baso, C. Gauthier, A. Hollender, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

K. Sweet and K. Wirth : One-to-One Personalization in the Age of Machine Learning

H. Lee and E. Chang : Consumer Attitudes Toward Online Mass Customization: An Application of Extended Technology Acceptance Model

S. S. Sundar and S. S. Marathe Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage

Exploring the Relationship between Reading Habits and Aesthetic Preferences in Different Cultural Contexts and Design Practices

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Experience 01

discovery

In today's world, content finds the consumer. Over the past decade, social media have had a profound impact on the way people discover content on the internet: time and attention are scarce resources and users have become increasingly accustomed to consuming information surfaced and filtered by friends and family via social feeds. In the current paradigm, information automatically flows toward readers – it simply shows up as part of whatever journey they're on. [1]

"paying active attention to the news was unimportant, because such information was "in the air" as an ambient part of daily life." [2]

In this way, social media channels have (re)defined the expectations and habits of users all over the world. Users now expect relevant information to find them as a result of their preferences, feed settings and serendipitous browsing. Facebook [3] and WhatsApp have become primary entry points for new users accessing the internet, and are, for many emerging communities, simply conflated with "the Internet". For these communities, the page-based mental model [4] of the Internet will effectively never have existed. So to meet our newest users where they are we must recognize the interdependence of form factor and discoverability in a consumption model not driven by search.

Sections

Experience 02

Social Media Facilitates Discovery

In a global user study by SDL, when asked "How do you typically discover new and interesting things online?" respondents indicated they turn first to social networks for content discovery, and then to online and customizable newsfeeds. More traditional means, like email and search engines, ranked far behind.

That same study found, on average, millennials share six pieces of content via social media a day, which has overtaken email with five shares a day as the de facto channel for sharing content.

Content discovery and sharing is driven by social media, and if Wikimedia content isn't there, we're missing an opportunity to increase our reach and be part of the global conversation.

This is particularly true in emerging markets, where social media is sometimes the entirety of a user's online experience. This is especially true in emerging markets where Facebook has reached low-income mobile users by partnering with cell networks to provide Facebook access for free (Facebook Zero inspired the ill-fated Wikipedia Zero).

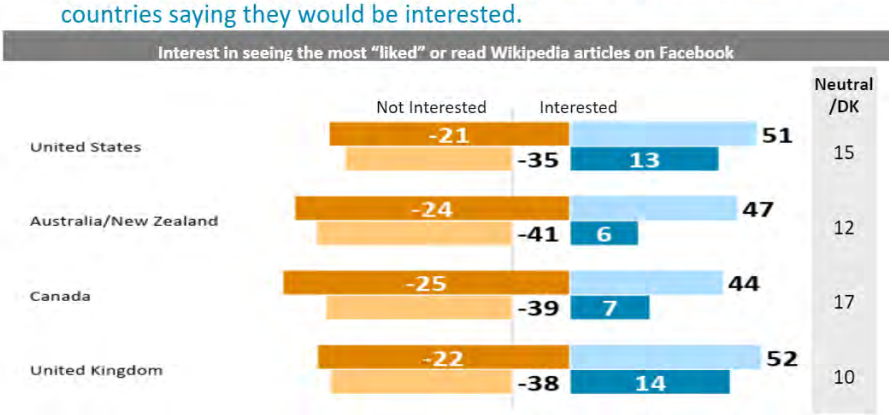
Researchers evaluating how Facebook Zero shapes information and communication technology use in the developing world found that 11% of Indonesians who said they used Facebook also said they did not use the Internet. Essentially, for them, "Facebook is the Internet"

Statistics show the number of shares on social media are down for many types of content, but Wikimedia has an advantage - "evergreen" content

In his *Content Trends 2018* report, Steve Rayson,found data showing that "evergreen" posts have resisted shifts in user behavior, tastes, and changes in Facebook algorithms. Despite Facebook share traffic being down overall, evergreen content remains

According to Rayson, articles that qualify as evergreen have at least one of the

Readers are more interested in seeing the most "liked" or read Wikipedia articles on Facebook, with approximately half of readers in all five countries saying they would be interested.



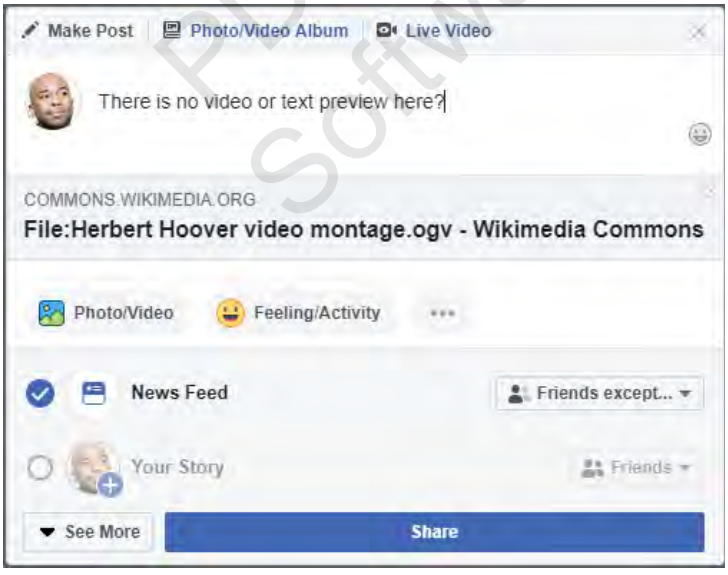
following characteristics. They are:

- Research-based
- Reference-style
- Topics that are relevant over time
- Updated regularly

These attributes are perfect, natural matches for Wikimedia content and show a clear opportunity for us to gain reach with a strategy for encouraging wider sharing of our content.

"In this new world of content saturation and falling social shares, the big winners are sites that have built a strong reputation for original, authoritative content.
The majority of content gets zero backlinks but authoritative research and reference content continues to gain links. In particular, authoritative evergreen content consistently gains shares and links over time."

- Steve Rayson, Content Trends 2018



Improve discoverability of content

Feeds, top articles, and relevance

On the mobile Wikipedia apps, top stories have become a promising avenue. Jonathan Morgan's 2017 research on the Top Articles feature in our apps found the following:

- do people prefer lists that are based on recent page views, or trending edits?
- On average, raters reported that they would be more interested in reading the articles in the 'top read' list than the 'trending' list. The results were consistent across groups, and (marginally) significant for India-based raters.

On average, raters reported that they would be more interested in reading the articles in the 'top read' list than the 'trending' list. The results were consistent across groups, and (marginally) significant for India-based raters.

Additionally, recent (though perhaps skewed) research about Wikipedia readers consistently found that around half of respondents were interested in seeing most liked or read Wikipedia articles on Facebook. These data points indicate that we're missing an opportunity to promote popular content at key places, either on-Wiki or on social media. Not everyone is interested in Top Articles, but a lot of people are and we can enhance their experience with functionality that is optional

Experience 04

and unobtrusive but very useful for users who want it.

Discovery of Multimedia

Discovery of multimedia content should also be a key component of our future strategy. With a focus on:

- Helping editors find useful multimedia to add to articles
- Helping readers find multimedia related to what they've read

As Ubiquitous as Internet Advertising

Advertising is an omnipresent force on the Internet today. It's not just tolerated, but expected, and as more people in developed nations become "cord cutters" and move away from traditional TV, internet advertising's omnipresence increasingly becomes a primary way that people discover new content, products, and news.

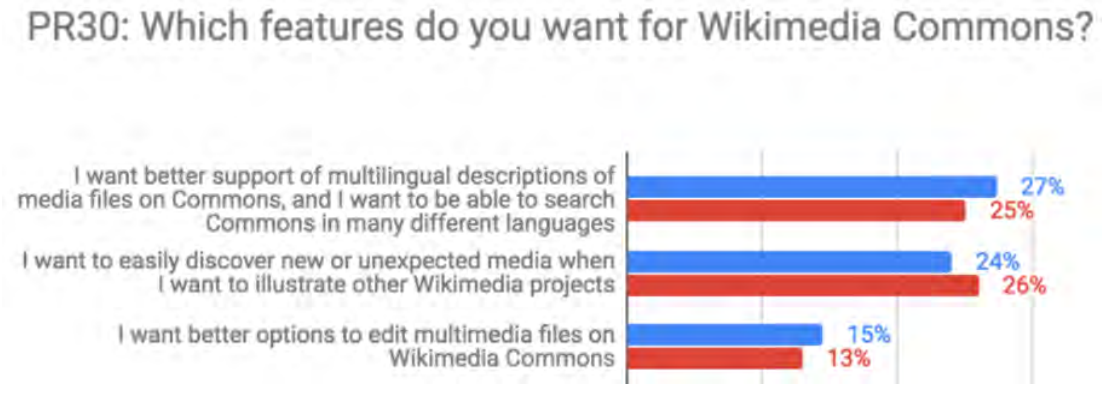
However, growing trends in consumer trust and content saturation suggest that there is room for a new omnipresent force on the Internet - facts.

Major players like Facebook, Amazon, and Google have already integrated Wikimedia content into their platforms. The Wikimedia experience is slowly being integrated into the user experience of other major platforms, but without our input.

Still, the ultimate goal companies like Facebook and Google are trying to achieve is a good one. Misinformation has become a top consumer concern on the Internet. The *Reuters/Oxford Digital News Report for 2018* [1] found:

Over half of those polled (54%) say they are very or extremely concerned about what is real and 'fake' on the internet. This is highest in countries like Brazil (85%), Spain (69%), and the United States (64%) where polarised political situations combine with high social media use.

Modern Internet users aren't sure what to believe. This is becoming a defining element of the Internet usage experience in modern times. In 2018, Facebook and Youtube began using content from Wikipedia to help combat this problem.



Experience 05

Notes

[1] Like the Burma Shave signs motorists encountered on US roadways in the 1950's and 60's, modern day internet users tend to consume whatever information shows up along whatever route they've chosen.
[2] need ref from Ramsey
[3]<ref>https://www.statista.com/statistics/346167/facebook-global-dau/</ref>
[4] [https://en.wikipedia.org/wiki/Mental_model mental model]
[5] Reuters/Oxford Digital News Report for 2018
http://www.digitalnewsreport.org/survey/2018/misinformation-and-disinformation-unpacked/

[2] By "2020, experts estimate 2.95 billion people to access social networks regularly. The majority of this growth is projected to come from mobile devices, as emerging markets catch up on online connectivity." https://www.statista.com/statistics/346167/facebook-global-dau/

[3] Like Burma Shave signs motorists encountered on US roadways in the 50's and 60's, modern day internet users consume the information that shows up along whatever route they've chosen. https://en.wikipedia.org/wiki/Burma-Shave#Roadside_billboards

Sources

R. Isley Research and Insights; Other contributors[1]: A. Baso, C. Gauthier, A. Hollender, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

J. Hempel, Wired, 2018 What Happened To Facebook's Grand Plan To Wire The World? https://www.wired.com/story/what-happened-to-facebooks-grand-plan-to-wire-the-world/

https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

I Just Google it:" Folk theories of Distributed Discovery
https://academic.oup.com/joc/article/68/3/636/4972617

Study by SDL
https://www.itbusinessedge.com/slideshows/global-study-finds-social-media-drives-content-discovery-with-millennials-02.html

Facebook is the Internet
https://qz.com/333313/milliions-of-facebook-users-have-no-idea-theyre-using-the-internet/

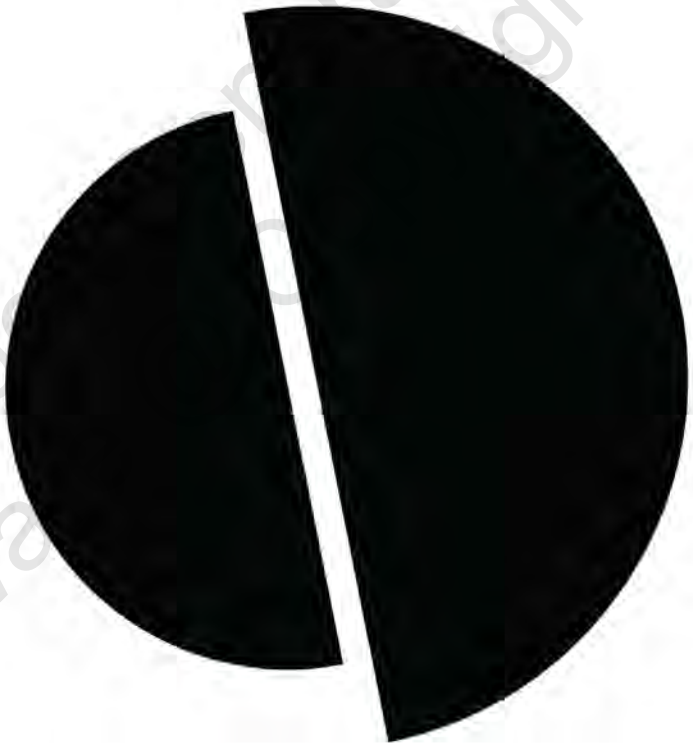
Content Trends 2018 https://buzzsumo.com/blog/content-trends-2018/

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Audiences x 3 / 5 Year Planning **2018**

augmentation

The following are draft position papers exploring aspects of machine learning and AI based on insights from Community and WMF experts, as well as synthesis of secondary research.



01
Content Curation

02
Content Generation

03
Governance

04
Machine Translation

Augmentation Overview 00

The Wikimedia movement wants the sum of all knowledge to be available to everyone in the world. We also want the process to assemble that knowledge to be inclusive, balanced, and safe for all participants. But there is too much knowledge needed, in too many languages, for humans to do this alone. As an example, if we assume that a Wikipedia that covers a substantial amount of knowledge has 2 million articles (likely a low estimate), and we believe that 300 languages should have access to that knowledge, we should expect there to be 600 million articles. There are currently only 48 million articles,[1] which is 8% of the way there. There are simply not enough potential human editors, especially in smaller languages, to get there. Whether or not we believe that long-form articles will be the medium of the future, this illustrates the problem we face.

Augmentation for contribution activities is our path to closing these gaps. Augmentation refers to any technology that helps humans do their work, and wikis have been using augmentation almost since their beginnings: Rambot created 34,000 articles from Census data in 2002,[2] Twinkle has been automating repetitive tasks since 2007,[3] ClueBot has been reverting vandalism since 2011,[4] and the Content Translation tool has employed machine translation to help generate content since 2016. Over the next three to five years, human editors will need to increasingly wield augmentation tools, especially those that incorporate artificial intelligence, to create content, curate content, and maintain a safe environment on the wikis. Artificial intelligence will not replace human editors -- it will allow human editors to focus on the most impactful and fulfilling work, and, if used correctly, will open up more avenues for more contributors.[5]

But although artificial intelligence is a powerful editing aide, it also has the potential to powerfully magnify the problems of bias and unfairness[6][7] that already exist in the wikis, and has the potential to discourage new editors.[8] Therefore, the role of human editors will change in the future to focus on wielding these tools safely to guard the wiki values that only humans understand.[5] In pursuing any augmentation technology, we should stick to the principles we apply to code and content: transparency and the ability for anyone to contribute. We should build closed-loop systems that essentially make augmentation "editable" by community-members, even non-technical ones. By making it possible for members of all communities to audit augmentation tools, contribute training data, flag errors, and tailor tools to their wikis, we will ensure that wikis are not unduly influenced by the smaller set of people who build the tools, while also opening up a new avenue of contribution.

In terms of capabilities we need to build, the Wikimedia Foundation should do two main things:

1. Build an infrastructure platform for many people to contribute augmentation tools.
2. Provide interfaces that make it possible for non-technical editors to apply, adjust and contribute to those tools.

The former would likely be pursued by the Technology department, while the latter would be pursued by the Audiences department. The Audiences work will create on-wiki tools that allow non-technical editors to record training data, identify errors in existing algorithms, and tune algorithms to fit their wiki's culture; surfacing those tasks as first-class wiki work that other editors can see. Through these interfaces, the shepherding of augmentation tools will become a new, major way of contributing that will ensure that machines are fair and healthy contributors to every wiki.

Assembling the platform and the interfaces that allow a feedback loop are the most important parts of this strategy -- more important than the particular applications of augmentation. That said, particular augmentation tools will generally fall into three aspects: content generation, content curation, and community conduct. We will need to develop design principles in each of these aspects that ensure augmentation tools are transparent and editable; and that ensure augmentation respects the boundaries between human work and machine work. These principles should also govern the ways we incorporate augmentation resources from third parties not controlled by the Wikimedia movement, such as machine translation services.

And finally, in order to be successful with this strategy, we will need to continuously recognize and embrace augmentation as a major way to contribute to the wikis. We can do this through community capacity building, holding events, providing training, and encouraging discussion in the community.

Examples

Rambot (content generation)

Twinkle (content curation)

ClueBot (content curation)

SuggestBot (content generation)

HostBot (governance)

Bot approval process (governance)

ORES models in RecentChanges and Watchlist (content curation)

Content Translation tool (content generation)

Article Placeholder (content generation)

Areas of Impact

Wikidata[9]

ORES[10]

Experienced editors[11]

Volunteer developers[12]

Key External Factors

The rate of improvement to artificial intelligence, especially machine translation.[13]

Efforts by other tech companies to automatically translate English Wikipedia, or to otherwise make massive amounts of information available.[14]

The movement's ability to get top talent to work on these issues as staff or volunteers.[15]

References 00

Sources

Bohannon, John and Dharnidharka, Vedant (2018). **Quicksilver: Training an ML system to generate draft Wikipedia articles and Wikidata entries simultaneously.** [Video from Wikimedia Research Showcase August 2018]. Retrieved from https://youtu.be/OGPMS4YGDMk.

Chisholm, A., Radford, W., & Hachey, B. (2017). **Learning to generate one-sentence biographies from Wikidata**. EACL.

Halfaker, Aaron. 2017. **Interpolating Quality Dynamics in Wikipedia and Demonstrating the Keilana Effect**. In Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17). ACM, New York, NY, USA, Article 19, 9 pages. DOI: https://doi.org/10.1145/3125433.3125475

Halfaker, Aaron, et. al. **ORES: Facilitating re-mediation of Wikipedia's socio-technical problems**. From Wikimedia Commons.

Kaffee LA. et al. (2018) **Mind the (Language) Gap: Generation of Multilingual Wikipedia Summaries from Wikidata for ArticlePlaceholders**. In: Gangemi A. et al. (eds) The Semantic Web. ESWC 2018. Lecture Notes in Computer Science, vol 10843.

Notes

1. Wiki segmentation 2018
2. ↑ Lih, Andrew (2009). The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia. p. 102. New York: Hyperion. ISBN 978-1-4013-0371-6.
3. ↑ History of Twinkle
4. ↑ History of ClueBot
5. ↑ Jump up to:5.0 5.1 This book **The Second Machine Age** examines the likely economic implications of artificial intelligence by looking at the effects of the Industrial Revolution. It makes the case that in the near and medium terms, artificial intelligence can create more jobs for humans than it replaces. Rather than replacing humans with machines, smart businesses will overhaul the way they work to incorporate new technologies as tools wielded by humans with increasingly sophisticated skill sets. This is already the case with companies who have successfully adopted modern IT practices. Brynjolfsson, Erik and McAfee, Andrew (2016). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W. W. Norton and Company. ISBN 978-0-393-35064-7.
6. ↑ JADE/Intro blog/Short story by Aaron Halfaker Gives a short description of how models can reinforce bias in the Wikimedia setting. Academic sources also exist.
7. ↑ Buowamlini, Joy (2018). **The Dangers of Supremely White Data and The Coded Gaze** [Video from Wikimania 2018]. Retrieved from https://www.youtube.com/watch?v=ZSJXKoD6mA8
8. ↑ Halfaker, A., Gieger, R. S., Morgan, J., & Riedl, J. (2013). **The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline**. American Behavioral Scientist 57(5) 664-688. Identifies automated curation systems as a key factor in de-personalizing the wikis and driving away new contributors.
9. ↑ Wikidata has the potential to be the abstract database of facts from which artificial intelligence could create content. We should decide whether we want this to be the case, and if so, to put resources behind Wikidata.
10. ↑ ORES and the way it is architected is the proof-of-concept for an open and auditable artificial intelligence abstraction in the wikis. It could either continue to grow to encompass more tasks, or it could serve as a model for future systems.
11. ↑ Experienced editors will need to continuously adjust their perception of what it means to do wiki work, as technology gives them increasingly powerful tools for content generation, content curation, and governance.
12. ↑ Volunteer developers will have a new way to contribute to the wikis beyond just software and content. They will be able to contribute algorithms.
13. ↑ There are many varying estimates for how quickly artificial intelligence will be able to take on human tasks. It is possible that capabilities will increase so quickly that the wikis are operating fundamentally differently within five years. Or that may not happen for 30 years. We should err on the side of expecting changes sooner, otherwise the wikis may be eclipsed by other, less open and fair, projects.
14. ↑ As machine translation improves, major tech companies and startups may attempt to make information, such as English Wikipedia, automatically available across all languages. The risk is that those companies would not have the same inclinations toward openness and fairness as the Wikimedia movement. If companies become suppliers of information before Wikimedia projects do, the world may wind up with an inferior dominant source of information.
15. ↑ If we see artificial intelligence as a critical path toward our movement's goals, we should be mindful of the difficulty of getting top talent to work on it. People who work on artificial intelligence are in high demand at the most elite high-paying companies in the world, but we will need them as volunteers and staff for Wikimedia projects.

content curation

Content curation is the aspect of wiki activity related to editing, refining, and cleaning up content that has been generated. The Wikimedia movement's ambitious aspiration to make the sum of all knowledge available to everyone in the world means that the movement has a tremendous amount of work to do with respect to making judgments about what information belongs, and how to organize, phrase, and cite it. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles. As all that content gets added, the curation workload will increase beyond what humans are capable of doing.

Augmentation is a potential pathway to curating the massive amount of information needed in the Wikimedia projects. By applying algorithms and artificial intelligence in the right ways, human editors can be aided in making the most important judgments about the content in the wikis, allowing the content to be well-organized and reliable. This kind of human-machine partnership is not new in the wikis. Tools like Twinkle and Huggle have been helping to automate the tasks of reviewing recent changes and patrolling for vandalism since 2007 and 2008. ClueBot has been independently reverting vandalism since 2011. And in more recent history, ORES machine learning models have begun to surface the edits and pages most in need of attention.

As humans and machines work together to curate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may just direct human attention to important curation needs. In other scenarios, the human may review tasks completed (e.g.

edits reverted) by an algorithm. This paper explores some specific examples of content curation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Because bias and unfairness already exist in the contents of the Wikimedia projects, algorithms have the potential to magnify and exacerbate those problems. The Wikimedia movement should confront this with the same principles that have led to our success in the past: transparency and the ability for anyone to contribute.

Sections

Augmentation 03

Augmentation strategy summary

This is a summary of the overall strategy for augmentation, which this document applies to the specific aspect of content generation.

In order to meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of the future of the wikis. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and for even non-technical editors to participate in training and tuning those algorithms. These principles would apply to all types of augmentation, whether it is in the aspect of content generation, content curation, or governing interactions between people.

Definition of content curation

Content curation is the aspect of wiki activity related to editing, refining, and cleaning up content that has been generated. This is in contrast with content generation, which is about adding new facts, writing, translations, or images to the wikis. It is also separate from governance, which is about the interactions and communications between wiki editors.

Aspiration

The Wikimedia movement wants the sum of all knowledge to be available to everyone in the world. That ambitious goal means that the movement has a tremendous amount of work to do. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles. Assembling all that knowledge is about more than just compiling it -- it means curating it: making judgments about what information belongs, and how to organize, phrase, and cite it. This makes the sum of all knowledge more accessible, and also makes it more trustworthy. There is going to be too much curation work in the future for humans to do it unassisted.

Augmented content curation

Augmentation is a potential pathway to curating the massive amount of information needed in the Wikimedia projects. By applying algorithms and artificial intelligence in the right ways, human editors can be aided in making the most important judgments about the content in the wikis, allowing the content to be well-organized and reliable. This kind of human-machine partnership is not new in the wikis. Tools like Twinkle and Huggle have been helping to automate the tasks of

Augmentation 02

reviewing recent changes and patrolling for vandalism since 2007 and 2008. ClueBot has been independently reverting vandalism since 2011. And in more recent history, ORES machine learning models have begun to surface the edits and pages most in need of attention.

As humans and machines work together to curate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may just direct human attention to important curation needs. In other scenarios, the human may review tasks completed (e.g. edits reverted) by an algorithm. Below are some specific examples of content curation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Activity	Algorithm role	Human role
Identifying vandalism	Flag edits that are likely vandalism	Review the flagged edits
Identifying unsourced content	Flag parts of articles that make claims that do not appear to be sourced	Review the flagged content and correct or delete
Checking sources	Check sources for the facts cited from them	Review any flagged citations found by the algorithm
Identifying copyright violations	Check edits for whether their contents appear in other sources	Review flagged edits and correct or revert
Grouping tasks	Assemble related individual curation tasks into a prioritized queue	Use the queue to work more efficiently
Routing tasks	Route individual curation tasks to the editors who are most likely to be interested or capable	Receive and take action on routed tasks
Improving article composition	Make automatic improvements to the tone, style, grammar, and spelling of written content	Review automated improvements

Augmentation 03

Content curation strategy

The human-machine partnership scenarios described above are easier said than done, and the technical effort to build such algorithms is only part of the challenge. The more important challenges are how to build a technical framework and establish design principles to ensure that as algorithms play a growing role in content curation, they are a force for unbiased and fair curation.

Content curation as done by humans necessarily reflects the biases that the humans have. For instance, one human editor's preference for writing with a certain style might accidentally exclude edits done by members of other demographic backgrounds. Because bias and unfairness already exist in the contents and practices of the Wikimedia projects, algorithms that learn from human work have the potential to magnify and exacerbate those problems. The Wikimedia movement should confront this with the same principles that have led to our success in the past: transparency and the ability for anyone to contribute.

Concretely, for algorithms that participate in content curation, these things should be true to ensure transparency and the ability for anyone to contribute:

- Algorithms should be able to be created and deployed by anyone. If content curation algorithms are only contributed by a select group, the way they curate content will reflect the biases of that select group.

  Example: if the creators of algorithms for identifying vandalism are all from the English-speaking world, the algorithms might do a poor job at identifying vandalism in other languages.

- It should be clear what work is being done by algorithms and where those algorithms come from.

  Example: if an edit is reverted by an algorithm, the user who initially made the edit should know that it was reverted by an algorithm. This will increase transparency, and potentially encourage human editors to improve the result.

- We should always build "closed loop systems" with humans in the loop. This means that any content curation done by an algorithm should involve a human to edit, improve, and audit the work.

  Example: in the Recent Changes feed, ORES models suggest edits that need attention, but do not automatically take action.

- Our "closed loop systems" should allow corrections made to machine work to be fed back into the system to improve the algorithm going forward.

  Example: the Recent Changes feed currently lacks a way for humans to flag

ORES suggestions that are incorrect so that ORES can be improved.

- Shepherding, tuning, and training algorithms should be an important wiki role that non-technical editors can take on. Any editor should be able to wield, adjust, and provide data for improving augmentation. This work should "count" as wiki work, as actual edits, and editors should find their way to this augmentation niche.

  Example: if the Recent Changes feed were to include a way to flag ORES judgments as incorrect, flagging those judgments should count as an edit.

As described in the overall augmentation theme strategy, the Wikimedia Foundation should do two concrete things to make the above possible:

1. Build an infrastructure platform for many people to contribute augmentation tools, coupled with Wikidata (or something like it) to serve as a repository of facts.
2. Provide interfaces that make it possible for non-technical editors to adjust and contribute to those tools.

Augmentation 05

Notes

[1] TBD

Sources

M. Miller and J. Klein : Research and Insights, Other contributors[1]: D. Garry, A. Halfaker, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

Twinkle https://en.wikipedia.org/wiki/Wikipedia:Twinkle

Huggle https://en.wikipedia.org/wiki/Wikipedia:Huggle

Cluebot https://en.wikipedia.org/wiki/User:ClueBot_NG

ORES Machine Learning Models https://www.mediawiki.org/wiki/ORES

Recent Changes Feed https://en.wikipedia.org/wiki/Special:RecentChanges

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Augmentation 01

content generation

Content generation is the aspect of wiki activity related to adding new facts, writing, translations, or images to the wikis. The Wikimedia movement's ambitious aspiration to make the sum of all knowledge available to everyone in the world means that the movement has a tremendous amount of work to do with regard to content parity across all wiki projects. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles.

Augmentation is a potential pathway to closing the gaps described above. By applying algorithms and artificial intelligence in the right ways, human editors can be assisted in generating the most important content for the wikis, allowing us to close the most important gaps fastest. This kind of human-machine partnership is not new in the wikis. As early as 2002, Rambot [1] generated 32,000 stub articles in English Wikipedia using Census data, and now in 2018, thousands of articles are translated between languages with the help of machine translation algorithms. On the horizon are technologies like Quicksilver,[2] which detects facts about entities from news articles and collates them for human editors to turn into needed articles.

As humans and machines work together to generate content, we can think about that interaction on a spectrum of how much work the human editor does and how much work the machine does. In some scenarios, the machine may just suggest a task that the human editor does in entirety. In other scenarios, the human may edit and improve on work done primarily by a computer. This paper explores some specific examples of content generation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Because bias and unfairness already exist in the contents of the Wikimedia projects, algorithms have the potential to magnify and exacerbate those problems. The Wikimedia movement should confront this with the same principles that have led to our success in the past: transparency and the ability for anyone to contribute.

Sections

Augmentation 02

Augmentation strategy summary

This is a summary of the overall strategy for augmentation, which this document applies to the specific aspect of content generation.

In order to meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of the future of the wikis. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and for even non-technical editors to participate in training and tuning those algorithms. These principles would apply to all types of augmentation, whether it is in the aspect of content generation, content curation, or governing interactions between people.

Definition of content generation

Content generation is the aspect of wiki activity related to adding new facts, writing, translations, or images to the wikis. This is in contrast with content curation, which is about editing, refining, and cleaning up content that has been generated. It is also separate from governance, which is about the interactions and communications between wiki editors.

Aspiration

The Wikimedia movement wants the sum of all knowledge to be available to everyone in the world. That ambitious goal means that the movement has a tremendous amount of work to do. Most of the hundreds of languages in the world have Wikipedias with less than 10% the number of articles that English Wikipedia has, and even the largest Wikipedias have serious gaps in terms of the depth of their articles, and the subject matter covered by their articles.

Augmented content generation

Augmentation is a potential pathway to closing the gaps described above. By applying algorithms and artificial intelligence in the right ways, human editors can be aided in generating the most important content for the wikis, allowing us to close the most important gaps fastest. This kind of human-machine partnership is not new in the wikis. As early as 2002, Rambot generated 32,000 stub articles in English Wikipedia using Census data, and now in 2018, thousands of articles are translated between languages with the help of machine translation algorithms. On the horizon are technologies like Quicksilver, which detects facts about entities from news articles and collates them for human editors to turn into needed articles.

As humans and machines work together to generate content, we can think about that interaction on a spectrum of how much

Augmentation 03

work the human editor does and how much work the machine does. In some scenarios, the machine may just suggest a task that the human editor does in entirety. In other scenarios, the human may edit and improve on work done primarily by a computer. Below are some specific examples of content generation activities that can exist in the future, drawing from all along the spectrum of the human-machine partnership.

Content generation strategy

The human-machine partnership scenarios described above are easier said than done, and the technical effort to build such algorithms is only part of the challenge. The more important challenges are how to build a technical framework and establish design principles to ensure that as algorithms play a growing role in content generation, they generate high-quality and unbiased content.

Because bias and unfairness already exist in the contents of the Wikimedia projects, algorithms have the potential to magnify and exacerbate those problems. The Wikimedia movement should confront this with the same principles that have led to our success in the past: transparency and the ability for anyone to contribute.

Concretely, for algorithms that participate in content generation, these things should be true to ensure transparency and the ability for anyone to contribute:

- Algorithms should be able to be created and deployed by anyone. If content generation algorithms are only

Activity	Algorithm role	Human role
Suggesting articles	List which articles should exist but do not	Create the articles listed
Suggesting information	Find new sources, distill facts, and surface the ones not yet in the wikis	Integrate the missing facts into the wikis
Suggesting updates	Identify and flag when information may be out of date	Correct the flagged information that needs to be updated
Initial translation	Create an initial translation of content from one language to another	Improve translations and remove bias
Starting articles	Create the beginnings of articles from sources	Correct and expand machine-generated stubs
Multimedia	Reassemble information from one medium (e.g. long-form article) into another (e.g. visual slideshow)	Correct machine-generated content

Augmentation 04

contributed by a select group, the content they generate will reflect the biases of that select group.

Example: if the creators of algorithms for suggesting notable female scientists are all from the English-speaking world, it is possible that the algorithm neglects notable female scientists from outside the English-speaking world.

- It should be clear what work is being done by algorithms and where those algorithms come from.

  Example: if an article is primarily generated through machine translation (such as through the Content Translation tool), that characteristic should be clear to readers and editors. This will increase transparency, and potentially encourage human editors to improve the result.

- We should always build "closed loop systems" with humans in the loop. This means that any content generation done by an algorithm should involve a human to edit, improve, and audit the work.

  Example: in the Content Translation tool, users are required to review and correct the automated translation done by the algorithm.

- Our "closed loop systems" should allow corrections made to machine work to be fed back into the system to improve the algorithm going forward.

  Example: after a human editor reviews and corrects a machine translation from the Content Translation tool, that correction should be used to improve the machine translation algorithm for the future.

- Shepherding, tuning, and training algorithms should be an important wiki role that non-technical editors can take on. Any editor should be able to wield, adjust, and provide data for improving augmentation. This work should "count" as wiki work, as actual edits, and editors should find their way to this augmentation niche.

  Example: if an algorithm were to exist to suggest notable topics for articles, and the algorithm suggests some topics that are not notable, non-technical editors should have an on-wiki way to mark those topics as non-notable so that the algorithm can improve.

As described in the overall augmentation theme strategy, the Wikimedia Foundation should do two concrete things to make the above possible:

Build an infrastructure platform for many people to contribute augmentation tools, coupled with Wikidata (or something like it) to serve as a repository of facts.
Provide interfaces that make it possible for non-technical editors to adjust and contribute to those tools.

Augmentation 05

Notes

[1] Meet the ‘bots’ that edit Wikipedia https://www.bbc.com/news/magazine-18892510

[2] Quicksilver was developed by Primer.ai

Sources

M. Miller and J. Klein, Research and Insights , Other contributors[1]: D. Garry, A. Halfaker, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar,

The History of Bots on Wikipedia https://en.wikipedia.org/wiki/Wikipedia:History_of_Wikipedia_bots

https://quicksilver.primer.ai/

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

governance

In order to meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of the future of the wikis. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and for even non-technical editors to participate in training and tuning those algorithms. These principles would apply to all types of augmentation, whether it is in the aspect of content generation, content curation, or governing interactions between people.

Governance is a word meant to capture a broader scope than "Code of Conduct". It refers to all the ways that people interact with each other on wiki projects, in both constructive and unconstructive situations. Current newcomers rarely contribute past their initial edits because of bad reactions to quality control mechanisms, algorithmic tools (bots) or policy. We see augmented governance practices as the vehicle that will safeguard, and simultaneously empower the Wikimedia community to become that desired safe haven for knowledge discourse through a set of human-centered principles.

Sections

Augmentation 02

In order to meet our movement's goal of making all the world's information available to everyone, we have more work to do than human editors can do alone. We need help in the form of augmentation, which is when humans and algorithms work together. Though augmentation in the wikis is not new, it will be a growing part of the future of the wikis. To ensure that the contributions made by algorithms are productive, unbiased, and fair, we will need to stick to our movement's principles of openness, transparency, and the ability for anyone to contribute. We should build closed-loop infrastructure and interfaces that allow anyone to contribute new algorithms, and for even non-technical editors to participate in training and tuning those algorithms. These principles would apply to all types of augmentation, whether it is in the aspect of content generation, content curation, or governing interactions between people.

Definition of governance

Governance is a word meant to capture a broader scope than "Code of Conduct". It refers to all the ways that people interact and communicate with each other on wiki projects, in both constructive and unconstructive situations. This is inclusive of mentorship, harassment, and straightforward conversations about content generation and content curation.

Aspiration

The Wikimedia movement wants the sum of all knowledge to be available to everyone in the world. We also want the process to assemble that knowledge to be inclusive, balanced, and safe for all participants. Current newcomers rarely contribute past their initial edits because of bad reactions to quality control mechanisms, algorithmic tools (bots) or policy. We see augmented governance practices as the vehicle that will safeguard, and simultaneously empower the Wikimedia community to become that desired safe haven for knowledge discourse through a set of human - centered principles .

Augmented governance

Wikipedia wants to attract and retain a person who edits in good faith and has a relatively high quality of edits (a desirable newcomer). Therefore, our goal for an augmented governance is to leverage artificial intelligence to create a safe space for humans to collaborate with each other to craft and compile the worlds' knowledge.

Currently the interplay between augmentation and governance has been explored through bots such as HostBot, which welcome new contributors to Wikipedia, and through processes such as the Bot Approval Process. The reception of these machine-generated greetings and process guidance has been initially cold due to impersonal and sometimes ineffective guidance. However, the way forward is leveraging the technology to create a partnership spectrum in which bots and algorithmic tools support and enhance the role of humans within the Wikimedia communities.

Below are some specific examples of augmented governance activities that can exist in the future, drawing from all along

Augmentation 03

the spectrum of the human-machine partnership.

Activity	Algorithm role	Human role
Welcoming	Identify newcomers and send welcome message	Follow up machine-generated messaging with a real offer for mentorship
Flag content for quality review	Identify and list poor quality articles with associated reason	Correct machine-generated reviews
Suggest content revisions for inappropriate behavior	List common ways to rectify the poor quality condition(ie: vandalism, harassment, sockpuppetry)	State if recommendation was useful and rectify the content dispute within the article
Provide wayfinding	Identify in-context opportunities and provide recommendations for users to consult resources (ie: norm guidance or policy disputes)	Follow wayfinding recommendations and consult resources.
Connect like-minded contributors	Identify when contributors are inhabiting similar spaces on the wikis and suggest contact	Accept suggestion and make human contact
Expertise and good behavior recognition	Identify when contributors have routinely provided their expertise within a wiki and provide a micro-credential such as a badge assertion.	Endorse micro-credentials. Once issued micro-credential, choose how to display it.

Governance Strategy

In the 1987 film RoboCop[1], a picture is painted of a dystopian city on the decline that is dealt with by employing an army of robots to brutally police the city. This is not the future of governance that we want at Wikimedia. The majority of AI systems and related tools that are being created in the world today are being put in place with minimal oversight, few accountability mechanisms and little research into their broader implications. Although artificial intelligence is a powerful editing aide, it also has the potential to powerfully magnify the problems of bias and unfairness [2] that already exist in the wikis, and has the potential to discourage new editors. [3] Currently, there are no internally agreed-upon methods to measure and assess the social implications of governance augmentation. Therefore, to ensure that the systems that we use and create are responsive to the complex social domain of Wikipedia communities, we will need to develop ways to measure, audit, analyze and improve them.

Augmentation 04

There are two ways that we can concretely do this:

1. The Wikimedia community needs to generate a set of guiding principles to ensure that the wikis are authentic and sincere representations of the worlds' knowledge. The following are topics that will be addressed:
   - **Human-centered AI:** AI should be "human centered", meaning that it nudges and opens the door to human connection instead of closing it. Wikimedia augmentation champions you – your security, privacy, and the quality of your online activity within our tools. Wikis got where they are today because of people working together, so there is something good about human connection that we should be mindful of preserving.
   - **Learnability as a core metric:** Wikimedia is optimizing for learning , which is a fundamentally different use of algorithms than recommendation systems built for revenue. Therefore, the core metric for success for all of the augmentation tools within Wikimedia's tools and projects is learnability, how well did the the tool or service assist the learning experience?
   - **Transparency:** While augmentation tools (hopefully) make sensible decisions for you that respect your time, data, and attention, you should always be made aware when augmentation tools are being utilized.
2. Developers and administrators will create and utilize open governance algorithms. If anyone can contribute to these tools, and if Wikimedia makes it possible for non-technical editors to watchguard them so that they aren't allocation or representation biased, we have the opportunity to define the role of human computer interaction within the context of governance within the communities that we create and run.

Guiding Questions

- What do WE mean by fairness?
- How might augmentation help us detect 'fake news' and misinformation campaigns from powerful malicious actors, like nation states?
- How can we ensure that individual editors (new, experienced) feel that their contributions are still VALUED in a wiki where so much is created and curated by machines?
- How can machines enhance the governance work done by editors?

Augmentation 05

Notes

[1] https://en.wikipedia.org/wiki/RoboCop
[2] Basic description of JADE by Aaron Halfaker
[3] Halfaker, A., Gieger, R. S., Morgan, J., & Riedl, J. (2013). **The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline**.
[4] American Behavioral Scientist 57(5) 664-688.

Sources

M. Marshall and J. Klein Research and Insights, Other contributors: A. Halfaker, D. Garry, D. Horn, J. Katz, J. Minor, J. Morgan, T. Negrin, M. Novotny, N. Pangarkar

Aaron Halfaker. 2017. **Interpolating Quality Dynamics in Wikipedia and Demonstrating the Keilana Effect**. In Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17). ACM, New York, NY, USA, Article 19, 9 pages. DOI: https://doi.org/10.1145/3125433.3125475

Halfaker, A., Gieger, R. S., Morgan, J., & Riedl, J. (2013). **The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline**. American Behavioral Scientist 57(5) 664-688.

Buowamlini, Joy (2018). **The Dangers of Supremely White Data and The Coded Gaze** [Video from Wikimania 2018]. Retrieved from https://youtu.be/ZSJXKoD6mA8.
https://www.mediawiki.org/wiki/File:Conversations_Gone_Awry_(slides).pdf

Augmentation 01

machine translation

Machine translation like any other machine augmentation, could accelerate our vision for the sum of all knowledge to be available to everyone in the world. Specifically, it means work done in one language can be made available in others. This could speed up the transmission of information dramatically, making more knowledge available and freeing contributors to work on original research. At the same time, machine translation from a third party, represents a threat to our model, potentially promoting English dominance, siphoning traffic from our wikis, offering a poor experience, or discouraging global contribution. We cannot ignore this trend, but must adapt to it. No matter how fast humans create or translate content, we will not be able to create or, crucially, update content faster than a competitor who uses machine translation. We need to adopt a proactive, long-term approach to machine translation that serves our users, aligns with our values and supports our ecosystem. If we do this correctly, machine translation could represent the ingredient that allows us to realize our vision of global knowledge sharing.

Sections

Augmentation 02

Machine translation is here

Imagine logging onto the internet, conducting a search, and seeing no results in your language. It is hard for English speakers to imagine, but it is the experience of many Internet users today. Lack of relevant content was cited as one of the top reasons people don't read online by our New Reader's research.

Automated translation is changing that. It is changing how people read, write and relate to each other. In late 2016, Google announced that it was now using neural networks to power it's translation, as these had quickly surpassed traditional, algorithmic models.[1] As a result, content translations are now just 1-click away in more than 100 languages. As of 2017, Google translate served 200M users a day and, as of 2018, translates 143 billion words a day.[2] The Facebook platform now performs 6 billion translations a day.[3]

Machine translation like any other machine augmentation, could accelerate our vision for the sum of all knowledge to be available to everyone in the world. Specifically, it means work done in one language can be made available in others. This could speed up the transmission of information dramatically, making more knowledge available and freeing contributors to work on original research. At the same time, machine translation from a third party, such as Google represents a threat to our model, potentially promoting English dominance, siphoning traffic from our wikis, offering a poor experience, or discouraging global contribution.

For some time, we have believed Google or someone else would soon use these tools to provide translated versions of Wikipedia if we didn't.[4] It is now clear that Google will be implementing a pilot of this in Bahasa Indonesia very soon, offering up English versions of pages, machine translated into Bahasa Indonesia if there isn't already a Bahasa Indonesia version. [5]

We cannot ignore this trend, but must adapt to it. No matter how fast humans create or translate content, we will not be able to create or, crucially, update content faster than a competitor who uses machine translation. We are currently working with Google to ensure that translated pages provide an option to modify the translation and save it as a page on Bahasa Indonesian Wikipedia.[6] However, this short-term solution would lead to static forks of pages that do not update over time.

We need to adopt a proactive, long-term approach to machine translation that serves our users, aligns with our values and supports our ecosystem. If we do this correctly, machine translation could represent the ingredient that allows us to realize our vision of global knowledge sharing.

1. Short Term Necessity

If Google's pilot is successful, we can expect it to roll out in other languages. Google search users will be offered machine-translated English Wikipedia articles as the default experience. As mentioned above, this potentially promotes English dominance, siphons traffic from our wikis, offers a poor experience, or discourages global contribution. One could

easily see how this could shrink communities and make entire countries read-only receivers of English-written perspectives. Ideally Google would let us control this experience, but that is not an option and our licensing does not allow us to enforce it.

Instead, we are working with Google to address the issues of experience and community health. We are asking them to ensure that our users know that the content is not written by humans and offer them a way to modify the translation for addition to the wiki in their language. In the long-term, we hope they will choose to use articles from other languages when appropriate, but this is not currently on the table.

This will change how people read Wikipedia globally. However, the long-term implications are just as important.

2. A Path to Leveraging Machine Translation

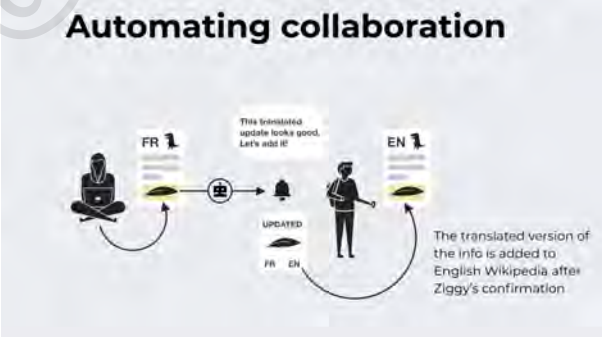
While machine translation is potentially a threat to Wikipedia communities. If we approach machine translation correctly, it offers the promise to spread knowledge much more quickly than traditional methods while accelerating content creation globally.

To take advantage of the opportunity, we will have to adapt our current approach from one in which content translation is a single fork of content, to one in which content flows readily from one wiki to another.

For example, today, the article about Genetically Modified Organisms (GMOs) might be brought from English to Hebrew. However, from this point on, the development of the two articles is forked. In English, several paragraphs might be added about the history of GMOs. In Hebrew, someone might add a paragraph about the economics of GMOs. At this point, neither wiki is benefiting from the scholarship of the other.

As articles change, their counterparts in other languages should have the option to import the new material. For example, editors would be informed about changes to articles in other languages (sections added, facts, etc). [7]

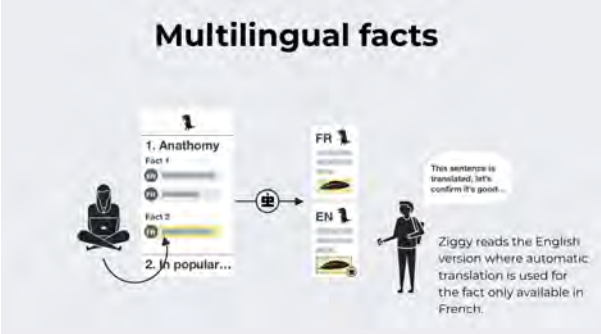
Here is a symbolic flow of what that might look like:



(From Pau Giner's "A Multilingual Wikipedia")

Another approach would be to focus on generating facts which can be migrated from language to language via machine translation. Whether or not Wikidata, already a global repository, is used as a semantic storage and mediation platform between wikis is somewhat contested.

Augmentation 04



(From Pau Giner's "A Multilingual Wikipedia")[8]

In either approach, we are taking the best of augmentation: using machines to replicate existing efforts and bringing in humans to confirm. This will require a shift in the kinds of work that needs to be done and the kinds of people who work on the encyclopedia and the kinds. For every 1 writer who has a book in their hand and cites it, there need to be 100 other editors whose job is to import the new content into the appropriate place in the destination wikis.

Machine translation will impact different wikis in different ways at different times. This approach focuses mostly on the wikis for which machine translation is good enough, as measured by our user.[9] Other segments won't be as immediately impacted by machine translation.

3. The Glorious Future

At some point machine translations will be good enough that it will allow real-time collaboration across languages. Indeed, Facebook and Google have already built this into their services. Instead of simply porting content from one language to another, this allows contributors from multiple languages to discuss and contribute to the same piece of work.

The future is already here — it's just not very evenly distributed.

-William Gibson

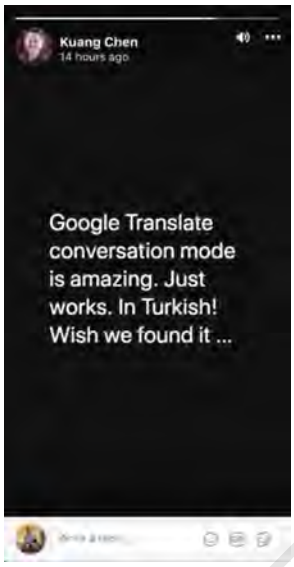
This is already happening on platforms like facebook. Here is an example where a woman was able to share the news of her father's death one-time in Hebrew and have her English-speaking and Hebrew-speaking friends discuss on the same thread:

[10]

The next step for facebook would be to automatically translate other languages into the language of the reader. This began earlier this year in select places.[11]

Augmentation 05

Google implemented real-time translation of speech in 2015 [12]and integrated these as a key feature into the earbuds they introduced in 2017.[13] A recent friend travelling around the world just posted this from Turkey:



[14]

Eventually, through this technology, it is possible that we would arrive at a single knowledge corpus, offering a variety of perspectives and written and consumed in many languages. In fact, such a system was recently proposed in a paper by Denny Vrandečić and shared on wikimedia-l.

The idea of a more global Wikipedia was also promoted by a community member in response to our announcement of a new machine translation service being available:

"I believe that if in the future it is possible to write the wikipedia articles in a universal language (I am not referring to english, latin or something like that, but something for computer), and any change made in any language was visible to all..."

This would be a more efficient way to amass the sum of human knowledge, but whether or not this idea is actually feasible will depend in large part on the interests of our communities. The time for that conversation is far away.

Guiding principles

While it might be tempting to simply adopt or reject machine translated content outright, adopting a set of principles we use to guide our machine translation strategy will allow for a more adaptive, fluid strategy.

- Here is a set of principles we stand by when it comes to machine translation
- More knowledge availability is good
- Contributions from people all over the world are necessary for capturing valuable perspectives and because a single region cannot capture all the world's knowledge
- Machines are neither good nor bad, but they require oversight.
- All wikis should have the chance to export content other languages. Default translations should be derived from the most compelling content from whichever (translatable) language provides it.
- Wikimedia Foundation decisions regarding MT need to be informed by what readers and community members want with respect to machine translation usage. Deciding based only on our values or perception has the potential to exacerbate unhealthy global power dynamics.

Risks and Open Questions

We still have many questions about machine translation that we will have to address over the coming months and years.

Risk: Disintermediation. Even if Google agrees to provide the users options to circle back to edit articles on their destination wiki, it is possible the mechanism will not be sufficient and local wiki communities will stagnate. Similarly, if users never reach Wikimedia-hosted site, we will not be able to fundraise.
Defense: Work with partners to establish the necessary entry points into our system As rapidly as possible fill out wiki content using machine translation to augment human effort.

Risk: Machine translation enforces bias. If we continue to rely on a 3rd party machine translation services, we are subject to the unknown biases built into those tools. In the case of language, some specific examples include using the male form of "prime minister" by default in gendered languages.
Defense: There aren't great options here. Work with tool providers to ensure that feedback is registered and to promote transparency. Using Wikidata's label system might help here. Another response might be to create our own, open machine translation tool.

Risk: Dependence on a 3rd party tool makes us vulnerable. If a fact is created on Wikipedia and then is replicated 100x using machine translation, most of the work is now being done by a third party tool. This dependency means we risk losing the tool at any time and would be vulnerable to the demands of a few key players, such as Google or Yandex.
Defense: There aren't great options here. Work with tool providers to establish terms up front. Push towards storing content as structured data. Another response might be to create our own, open machine translation tool.

Risk: Low quality machine translation. If machine translation is pushed on users by a third party and creates incomprehensible text or even promulgates falsehoods, we risk harming our users and Wikipedia's reputation.
Defense: Work with Google to push back where necessary. Research tools and listen to user feedback. We are currently planning an investigation into how machine translation is perceived.

Summary

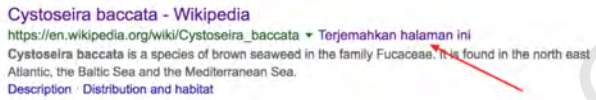
Machine translated content is being read by humans at an accelerating rate. This will soon extend to Wikipedia. In the short term, we need to work with third parties to ensure our values our being met and our ecosystem can continue to thrive.

In the longer term, we need to multiply our efforts on the translation front, by using machine translation to promote the flow of knowledge across multiple projects. This will mean moving beyond article translation into translating changes across articles.

1. Real-time translation: A third party is using machine translation to provide meaningful default articles where the users' primary language version is non-existent (or maybe even a short stub), using the English article. Eventually move to some notion of article quality and translation quality to choose the source language. We use default articles as funnel to contribution. "Improve this translation". A key component of this is working with Google.
2. Synching articles: Harness machine translation to make cross-wiki collaboration better. Move from forking articles to synching articles, Expand the use cases so that contributors do not need to know more than one language to propagate changes from one wiki to another.
3. Potential Global Corpus: Eventually we approach a state where there is a global corpus and when someone does the research to expand an existing

> Wikipedia entry, that research and writing doesn't necessarily need to be manually reinvented in 200 separate places. We are able to examine whether we want different perspectives to be reflected along ethnic or philosophical lines, rather than along language lines (which have, by necessity,served as a convenient, but imperfect proxy for "perspective").

Today, Amisha, a biology student in Indonesia, looks up Cystoseira baccata, a species of brown seaweed, on Google. There is no Wikipedia article in Bahasa Indonesia for this topic, so Google provides the Wikipedia article in English in the results page.



There is an offer to translate it. If she instead clicks on the link, she is taken to an article in English and her browser may or may not offer to translate it.

There is a Google effort to improve this experience. In the near future, if Amisha conducts the same search, Google will by default offer up a translated version of the English page, hosted by Google with a Google header, yet with of our site's appearance and branding.

Here is one possible view:

This offering has been tested, and Google has suggested that readers reacted very positively to it. They plan on doing this for any page in which there isn't a suitable Bahasa Indonesian article. Bahasa Indonesian is the only language they are currently applying this to, because it is a fairly well-structured and easy-to-translate language. However, it is obvious that Google's ambitions do not stop at Indonesian or with Wikipedia. Their goal is to make the world's knowledge available in every language on the web.

Augmentation 06

Notes

[1] https://www.blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/
[2] https://www.businessinsider.com/sundar-pichai-google-translate-143-billion-words-daily-2018-7
[3] There is also evidence that machine translation will have improved results on the long-tail of languages for which there isn't a rich corpus of translated works.
[4] https://office.wikimedia.org/wiki/User:Jkatz_(WMF)/Language_in_15_years
[5] For more context, see section below.
[6] Pau's approaches to mediating Google's intervention
[7] There's already a phabricator ticket for this
[8] More variations on this theme here
[9] This research from 2018 translating idiomatic English phrases to 102 languages using google translate, suggests at least some directional evidence of what those languages are. More context linked from Quora.
[10] Screenshotted from Facebook, pulled November 1st, 2018
[11] https://newsfeed.org/facebook-will-automatically-translate-your-pages-and-groups-posts/, https://techcrunch.com/2018/05/01/facebook-messenger-translation/
[12] https://techcrunch.com/2015/01/14/amaaaaaazing/
[13] https://www.youtube.com/watch?v=oQVQVt5H2QM
[14] Screenshotted from Facebook, pulled November 9th, 2018

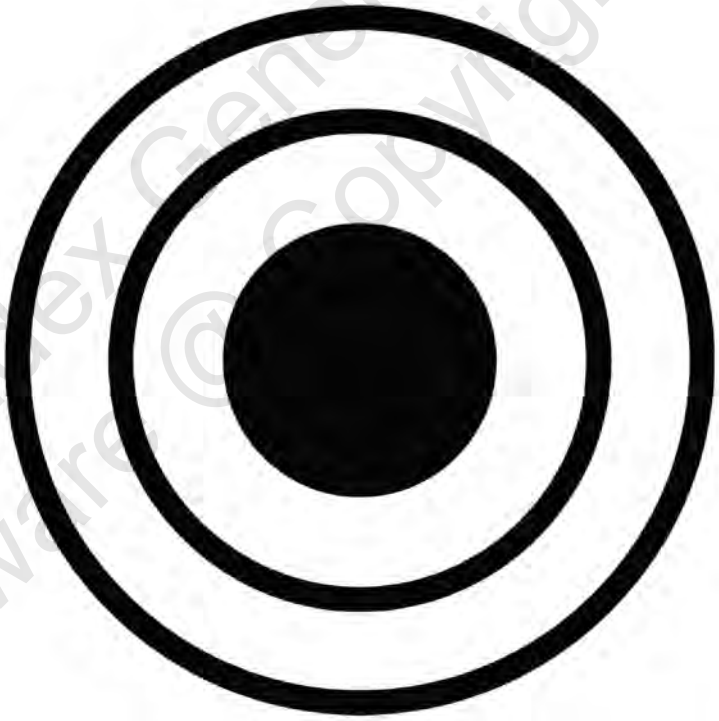
Sources

J. Katz : Research and Insights, Other Contributors: M. Miller

Audiences x 3 / 5 Year Planning 2018

scale

The following are draft position papers exploring imperatives and implications of scaling the projects based on insights from Community and WMF experts, as well as synthesis of secondary



Scale Overview 00

By 2030, up to 90% of the world's population will be using the internet[1]. They will bring new languages, new customs, and ways of communicating - and the technologies we have today will have to evolve to account for their needs. The internet will change and the patterns and interactions of today will become less relevant. Some will grow, transform and reinvent themselves. Others will fade into obscurity.

It is difficult to envision a future where Wikimedia projects, in their current form, continue to be essential to the needs of new internet users exposed mainly to social media, short-form text, and multimedia across a variety of platforms.[2][3][4][5] It is equally difficult to envision Wikipedia restructured as a social network or atomized into a database providing knowledge throughout every corner of the internet while retaining its active readers, communities, and donors. [6][7]

For years we have established a baseline for quality content for the world's internet population. Yet as the identity of this population changes and our content gap widens, we are found increasingly wanting. If our goal is to increase readership in new markets, or even to provide the content readers are interested in within existing markets, we must focus on not only the size, but also the relevancy of our content. An increase in locally-relevant content can not only bring in new readers, but provide for them an opportunity for representation that has so far been sparse not only within Wikimedia projects, but within all media. In addition, allowing communities to create different types of content can make accessibility for a variety of different audiences much easier.

Yet growing relevant content works under the assumption that the supply of content is equal to the demand from readers and, unfortunately, this does not apply to our current structure. Along with our readers, we must grow our communities by focusing on decreasing the barriers between readers and editors.[8] and ensuring new editors have the support they need to begin providing quality content to projects they are interested in.[9]

Additional focus must be placed on the content itself. While we are not capable of predicting the needs of all of our future users, we can ensure that our content is adaptable to any technical trends that may occur and support our communities by providing them with the tools necessary to create, curate, and moderate such content. We can focus on building relationships between projects and communities so that people looking to find, or contribute to, different types of information can do so with ease.[10]

One change that may seem inevitable is syndication across other platforms - providing the ability for partners to use our content and for others to access it. We must note that such a future, if implemented without proactive management, can put the sustainability of our communities at risk. Without a steady

Scale Overview 00

rate of visits to the site, less readers become editors and, over time, the quality of our content will suffer. To account for this imbalance, we can explore the relationship between content creation and syndication and focus on building tools that will allow content creation to continue in an increasingly dispersed network. We can expand our presence on other platforms while continuing to navigate users back to our projects.

Success in the above spheres will provide equitable growth to our projects and communities and ensure ubiquity of our content throughout the fabric of daily technology. Yet it does not address our vulnerability to external threats or offer us protection. Defense to such threats must also be treated as a priority. We can explore options such as making censorship and security threats more expensive for those who wish us harm, exploring different ways for accessing our content, and supporting other organizations that stand against censorship.

If we hope to become the "essential infrastructure of the ecosystem of free knowledge" and to allow "anyone who shares our vision [to] be able to join us", we must focus on providing knowledge-seekers with content relevant to their needs and interests, sustainably growing healthy and diverse communities, and ensuring our continued presence throughout the fabric of the internet. We must also focus on protecting ourselves and ensuring we are resilient to internal and external threats.

Examples

Structured Data

Global tools

External contribution models

Identifying content gaps

Platform-agnostic content

Areas of Impact

All wiki projects

Community Relations

Community developers

Partnerships

Infrastructure

Research

Resources

Modeling crowdsourcing as collective problem solving - models productivity for crowdsourcing tactics https://www.nature.com/articles/srep16557

A study of inequality in content among different language projects. Identifies common growth patterns among wikis https://arxiv.org/abs/1610.06006

A study of multiple cross-language Wikipedias to determine connections between articles in different languages are not directly correlated to their existence or quality in English Wikipedia https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0070-8

A study of versions of articles available across 26 wikis. Identifies that project growth seems unrelated to linguistic or cultural factors. http://rsos.royalsocietypublishing.org/content/4/10/171217

A case study of community participation in the design and implementation of algorithmic systems (namely moderating bots on Wikipedia). https://spir.aoir.org/index.php/spir/article/view/1383

References 00

1. Looks at different article categories to determine patterns of editor interaction. Conclusions were that “infrequently referred articles tend to grow faster than frequently referred ones, and articles attracting a high motivation to edit counterintuitively reduce the number of participants” https://arxiv.org/abs/1510.06092 https://meta.wikimedia.org/w/index.php?title=File:The_World_in_2030_-_Presentation_for_movement_strategy_discussions.pdf&page=30
2. https://dl.acm.org/citation.cfm?doid=3173574.3173929 Studies the rise and decline of Wikia wikis and identifies a similar pattern for these wikis, as observed in the original Rise And Decline paper by Halfaker, et. al. indicating that a wiki may have a roughly predetermined life cycle.
3. New readers research - which highlights beginner reader needs and expectations in emerging markets. https://meta.wikimedia.org/wiki/New_Readers/Findings
4. NY Times editorial summarizing the predominance of multi-media content https://www.nytimes.com/interactive/2018/02/09/technology/the-rise-of-a-visual-internet.html
5. https://econsultancy.com/why-we-need-to-stop-repeating-the-50-by-2020-voice-search-prediction/ Less hype-prone look at the rise of voice search, as an example of the way users interactions with internet content are changing
6. The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies - shows that Google’s clickthrough rates drop significantly when not surfacing Wikipedia content, but that Wikipedia’s pageviews drop when Google displays such content
7. Examining Wikipedia with a broader lens: Quantifying the value of Wikipedia's relationships with other large-scale online communities Looks at Wikipedia’s relationship with Stack Overflow and Reddit: “Wikipedia provides substantial value to both communities, with Wikipedia content increasing visitation, engagement, and revenue, but we find little evidence that these websites contribute to Wikipedia in return. “
8. https://www.journals.uchicago.edu/doi/abs/10.1086/689816 Is a study that explored Coleman’s social norms in the context of Wikipedia: “dense networks provide an opportunity structure to reward those who punish norm violators, leading to more frequent punishment and in turn fewer norm violations.”
9. The New Editor Experiences research provides insights on these barriers and suggests ways to reduce them
10. Displays that knowledge is more effective when actively sought https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185

Scale 01

community

The Foundation is currently working on features designed to bring more people into our communities.[1] What will all of these new residents need in order to find their place, and what does the existing community need, to deal with this influx of new strangers? The population of English Wikipedia has famously dropped since hitting a peak ten years ago, going from a high of around 28,000 active editors in 2007 down to about 15,000 by 2013.[2] Because this figure has remained more or less consistent since then, it may suggest the number of people the current structure of the site can support. But if the goal is to grow the active population of Wikipedia by attracting and assimilating a large number of new strangers, then a new conceptual model is required- this paper posits a model based on urban planning theory, conceptualizing each contributor community as if it were a city. Such a model would require tools and support to help established residents and newcomers be more visible to one another, and interact. That is, each community must be reconsidered in terms of scale - like a city is organized by neighborhood - in order to ensure that newcomers land in a place they can identify with, among others with similar interests and motivations. Such a model would ensure greater cohesion within contributor communities, transfer of knowledge between related contributor communities, and the opportunity for more visibility and awareness of the activities of others (both in terms of recognizing positive contributions and policing negative behavior). This new model begins to suggest role structures that are more flexible and nuanced (official, ceremonial, interest-based, activity-based, time-based etc.).

Sections

Scale 02

The Problem of Strangers

Growing from a population of 15,000 active editors to 150,000 is like moving from a small town to a busy city. New York City isn't just a small town that got bigger; the scale creates new levels of complexity. In a small town, you see the same people every day, and you can keep track of all the important happenings around town.

But you could walk around New York for a week, and not see the same person twice. In a big city, you're constantly surrounded by strangers, and there's far too much going on to keep track of, which means that there are different expectations around the way that people behave.

For established residents of the community, Wikipedia still feels like a town -- they see a lot of the same people, and they know where all the important meeting points are. But it's a town that's overrun by strangers -- there are 15,000 active editors per month, and around 350,000 people making 1-4 edits.

And for visitors who enter the community by making an edit, it's like stepping into a dark, empty street, where their first interaction might be a stranger jumping from the shadows and bullying them. The newcomer doesn't know if anyone is around to stick up for them, or help them.

Eyes on the street

In a city, both the residents and the newcomers want to feel safe among all of these strangers, and that feeling arises from the natural, active use of the city's streets and sidewalks. In a safe neighborhood, people are sitting on the steps, looking out of the windows, and hanging out in front of the stores. People are less likely to commit crimes or bully people, if there are other people watching.

In The Death and Life of Great American Cities, Jane Jacobs identifies this as the key to public safety:

"There must be eyes upon the street, eyes belonging to those we might call the natural proprietors of the street. The buildings on a street equipped to handle strangers and to ensure the safety of both residents and strangers, must be oriented to the street. They cannot turn their backs or blank sides on it and leave it blind." [Jacobs, ch 2]

This amateur surveillance doesn't need to be organized and explicit; it happens naturally, because people enjoy the sight of other people.

"Nobody enjoys sitting on a stoop or looking out a window at an empty street. Almost nobody does such a thing. Large numbers of people entertain themselves, off and on, by watching street activity." [Jacobs, ch 2]

But the street needs to be active, in order to be safe:

"*The sidewalk must have users on it fairly continuously, both to add to the number of effective eyes on the street and to induce the people in buildings along the street to watch the sidewalks in sufficient numbers."* [Jacobs, ch 2]

In *City: Rediscovering the Center*, William H. Whyte says that the same strategy applies to other public spaces:

Scale 03

"The best way to handle the problem of undesirables is to make a place attractive to everyone else. The record is overwhelmingly positive on this score. With few exceptions, center city plazas and small parks are safe places." [Whyte, ch 10]

But this strategy works on a local level, street by street. What Jacobs refers to as "the natural proprietors of the street" are the people who feel some ownership and responsibility -- the people who live on that block, the people who own and work at the local businesses, and the regular visitors who have a connection to the neighborhood.

You can't keep an eye on the whole city at the same time, and nobody really wants to. People like to watch places that are busy, and places that they care about.

Wiki Neighborhoods

The scale of a big city is comprehensible because there are neighborhoods, smaller areas with their own characters and specialties. There's a big difference between Greenwich Village, a bohemian artist's neighborhood, and Wall Street, an international financial center. In these neighborhoods, there are different schedules, and different expectations about how people talk, dress and behave. The kind of people who feel welcome there will be different. On Wall Street, everyone wears a business suit, and they all go home at 6pm. In Greenwich Village, the neighborhood doesn't really get started until the middle of the afternoon, and things are open all night.

Currently on Wikipedia, the closest analogue to a neighborhood is the WikiProject, topic-based project pages where editors coordinate editing work around a shared interest. Editors join the project by adding their username to a list of members, and active projects organize communal events and work toward shared goals. These should be places where the "natural proprietors" of a topic area can watch people walk by, and perform both functions of the city street -- protecting the neighborhood from bad-faith strangers, and making sure good-faith strangers are treated well.

Unfortunately, for the most part, WikiProjects don't perform those helpful functions, because the project pages are static -- explaining what the project is and how to get involved, but not providing any signs of activity that would encourage visitors to come back. These are buildings that turn a blank wall to the street, creating empty plazas that don't inspire people to take action.

Dynamic Environments

The active WikiProjects on English Wikipedia go out of their way to make sure that there's new activity to look at, often near the top of the project page:

- WikiProject Women in Red has a tally at the top of the page that shows the percentage of biographies about women which is updated each week, and has

recent announcements and events listed right under that.
- On WikiProject Military history, there's a monthly newsletter that comes out more often (and appears to be better-loved) than the general-interest Wikipedia Signpost.
- WikiProject Medicine has a regularly-updated Discussions module, which automatically lists current talk page conversations about pages in the WikiProject's subject area.

Topic areas could become functional neighborhoods that help to solve the problem of strangers, but the people who work in that topic area need a street to watch. There should be topic-specific Recent changes feeds, where people who care about that subject can see the activity happening on the relevant pages. The feed could be based on the existing WikiProject categorization, with new articles added automatically, based on a proposed expansion of ORES machine learning.

Making people more visible

Seeing other people on city streets also helps visitors and new residents to fit into the existing neighborhood. Watching other people helps people understand how to behave in a new environment. If we want thousands of new strangers to assimilate to the Wikipedia communities, then they need to see other people interacting successfully.

Currently, Wikipedia articles are designed as if the existence of editors was a closely-guarded secret. Besides the button marked "View history", there's no way for a reader to understand what editors do on Wikipedia, or how they behave.

The received wisdom is that well-written article pages should look professional and encyclopedic, with all visible signs of human activity tucked away on the talk page and history page. However, there are clues about editor activity on pages with issues -- warning templates explain that "The neutrality of this article is disputed," or "This article has no lead section." This means that visitors are only made aware of editing activity when something has gone wrong.

This same urge to make things look clean and professional occurs to city planners as well, as Jacobs points out:

"This last point, that the sight of people attracts still other people, is something that city planners and city architectural designers seem to find incomprehensible. They operate on the premise that city people seek the sight of emptiness, obvious order and quiet. Nothing could be less true. People's love of watching activity and other people is constantly evident in cities anywhere."

If we want the active population of Wikipedia to grow by attracting and assimilating a large number of new strangers, then we need to design as if Wikipedia was a city, rather than a small town. We need to help established residents and newcomers to see each other, and interact.

Scale 05

Notes

1. Mobile editing tools, more onboarding features, and an easier-to-use communication system.
2. The estimates of active editor participation comes from stats.wikimedia, using 25+ edits/mo as the definition of "active"

Sources

D. Horn: Research and Insights, Other contributors[1]: A. Amaroni, C. Gauthier, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

Jane Jacobs, 1961 The Death and Life of Great American Cities, Random House

William H. Whyte, 1988, City: Rediscovering the Center, University of Pennsylvania Press

A. Kittur, B.Pendleton and R. E. Kraut, 2009, Herding the Cats: The Influence of Groups In Coordinating Peer Production" [pdf]

H. Ung and J-M. Dalle, 2010, Project Management in the Wikipedia Community [pdf]

J. T. Morgan, M. Gilbert, D. W. McDonald and M. Zachry, 2013 Project talk: Coordination work and group membership in WikiProjects, [pdf]

J. Solomon and R. Wash, 2014, Critical Mass of What? Exploring Community Growth in WikiProjects [pdf]

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Scale 01

content

Wikimedia is many things: a software platform, a global movement, a collaborative community. But for the vast majority of our daily users[2] Wikimedia means one thing: informational content. Readers come to Wikimedia (and largely Wikipedia) for many reasons[3], primarily to satisfy an intrinsic curiosity, or to become more informed about something they see in other media. But no matter the motivation their satisfaction rests, finally, on one thing: relevant content.[4] Satisfying this need for new users in new markets will be the key to encouraging growth in readership, just as it did in the early growth phase of Wikipedia.[5]

This core user need also aligns with our strategic direction. That is, locally relevant content is not only a potential engine of growth in new markets, but filling these gaps in the content is core to combating the larger inequities in the knowledge that historically has been stored and shared on Wikimedia. By encouraging and enabling new content and topic growth in previously excluded areas, Wikimedia can drive not just growth for its own sake but equitable growth: growing specific audiences and content which have previously not been able or allowed to participate in global knowledge production and distribution.

Sections

Scale 02

Intro

"topics about the global south are not as strong in English Wikipedia... [Getting more content in these gaps...] that has an important effect for us as a movement, broadly. the more content there is in Wikipedia that is relevant to people in a certain part of the world, the more likely they are to use it and engage with it. It's sort of a self feeding cycle." [1]

Wikimedia is many things: a software platform, a global movement, a collaborative community. But for the vast majority of our daily users [2] Wikimedia means one thing: informational content. Readers come to Wikimedia (and largely Wikipedia) for many reasons, [3] primarily to satisfy an intrinsic curiosity, or to become more informed about something they see in other media, but no matter the motivation, their satisfaction rests, finally on one thing: relevant content. [4] Satisfying this need for new users in new markets will be the key to encouraging growth in readership, just as it did in the early growth phase of Wikipedia. [5]

This core user need also aligns with our strategic direction. That is, locally relevant content is not only a potential engine of growth in new markets, but filling these gaps in the content is core to combating the larger inequities in the knowledge that historically has been stored and shared on Wikimedia. By encouraging and enabling new content and topic growth in previously excluded areas, Wikimedia can drive not just growth for its own sake but equitable growth: growing specific audiences and content which have previously not been able or allowed to participate in global knowledge production and distribution.

Content and Participation

One significant barrier to wider participation and filling of content gaps by new participants with content that is relevant to new readers is the asymmetry between the experience and tools of our current editing community and the reading community whose needs they fill. Currently only about 5% of edits are made on mobile devices. However nearly 60% of our total device access comes from mobile devices. This means there is a disconnect between the people writing and curating the content and the people who consume it. Although this affects issues like presentation and content form very directly, it also means that the people writing Wikipedia do not reflect the reader population, its context or experience. It is key that we enable participation on the devices and in the context where content is consumed.

Although our ultimate goal is to fill these gaps, and satisfy the information needs of users around the world, Wikimedia also relies on a dedicated community to create and most importantly for this discussion, quality control, the information. In order to preserve the trust and reliability we must also balance the pressure between content growth and content quality and moderation systems. Merely bringing in new eyeballs with click-bait for fake news might create growth, but it undermines the value of that same content. This means that as we encourage new content contributions and the growth of new topics, we must monitor and support the curators and

administrators. However, we also must overcome the significant bias and inherent

exclusionary nature of certain policies and current community attitudes. Qualitative research[6] and user reports[7] suggest that policies, particularly around notability and reliable sourcing are especially problematic.

Regional Relevance

One way to provide relevant content for many users is to replicate the existing content in their language. This resolves one barrier for users (ie. the content is at least in a language they read and write). However, many many topics of local importance and interest may not exist on any Wikipedia. Additionally for many users English (or other large colonial language) remains the primary language of the internet and of education more generally, and users expect to search and read about their topics of interest in this global language. This means that we cannot fill these information needs and expectations purely by translating content from large to small languages. It means that we will need to ensure large global wikis like English accommodate and, indeed, encourage a multi-cultural tolerance of difference and variation, and get support for curation tools that enable this tolerance and cooperation.

For most users,[8] coming to Wikipedia to have your information needs met starts not on Wikipedia but on Google. Their journey begins by searching for keywords. If these keywords are found on Wikipedia, there is a good chance [9] they will see that result and come to us to satisfy their information need. This results in increased readership, which in turn, should result in additional contributors and content growth. Encouraging this virtuous cycle between search, content and knowledge generation applies energy to the flywheel that is at the heart Wikimedia's content engine. By identifying and filling content gaps, in English and across languages, we add more search keywords that help readers find us. Some of those readers share and contribute, expanding the movement. These new contributors keep our collaborative community diverse and active, adding and improving content. And all this comes back to help satisfy readers needs for the sum of all knowledge.

There are many potential ways to improve and encourage this cycle, including some already underway.

Software interventions: directing interest with recommendations, improving inclusive interactions

Programmatic interventions: interest groups, project tiger, content campaigns,

Scale 04

Notes

[1] D.Scott, Lead Organizer of Wikimania 2018 https://www.youtube.com/watch?v=TTtb4dEypQk at roughly 22 minutes in

[2] We count approx 200,000 contibutors a month, and roughly 1B devices. The means 99.98% of use is non-contributory consumption. This undercounts as it doesn't account for the readers reached through massive re-publication channels, such as the Knowledge Panel.

[3] https://blog.wikimedia.org/2018/03/15/why-the-world-reads-wikipedia/

[4] "Interestingly, one of the barriers to adoption that this report cites is a lack of local content. "In trying to connect the unconnected to the internet, content has for many years been the forgotten ingredient, with efforts prioritised in expanding coverage and lowering the cost of ownership. These are, of course, fundamental, but so too is the question: is the internet relevant for me?" https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Sources/Considering_2030:_Future_technology_trends_that_will_impact_the_Wikimedia_movement#cite_note-8 pointng to https://www.gsmaintelligence.com/research/?file=357f1541c77358e61787fac35259dc92&download slide

[5] There are a number of papers and books which examine the network effect and symbiotic growth between Google and English Wikipedia in the early years of the project. Anrew Lih's is prbably the most narrative. Academic version: The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies

[6] New Editors and New Readers research both make the case that learning and understanding policies and the suitability of those policies for other cultures or underserved topics may present significant barriers.

[7] In the commentary that follows the quote that opens this document, for example, notability and reliable sources policies are cited as barriers for African participants in English Wikipedia for example. Interestingly Asaf Bartov recently claimed in a related discussion that notability is not the core problem faced by emerging communities, but rather reliable sourcing. In either case these are community policy issues.

[8] Search referral traffic is X% (I think like ⅔) of daily traffic.

[9] Note about search rank and CTR

Sources

J. Minor Research and Insights Other contributors[1]: A. Baso, B. Davis, A. Halfaker, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, O. Vasileva

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Scale 01

resilience

In order to reach its 2030 goals, the Wikimedia product platform must prepare for rapid scaling of development, of contributors and of content. In the process it will be critical to design for resilience – the ability to engender sustainable growth and to fend off threats. For example, it will be necessary to define countermeasures against external threats such as censorship, misinformation [1], climate and policy related threats, as well as attacks on security or privacy by state actors. It will also be necessary to anticipate and countermand threats that could undermine the projects from within: communities or affiliates turning against one another, communities turning themselves [2] and communities turning against the Foundation. And finally, perhaps the most critical existential threat is relevance: what barriers to entry could be erected to prevent loss of mind share? What pre-emptive measures must be taken to guarantee mind share as new communities come on line? This paper explores each type of threat and offers a set of economic, cultural, and technical countermeasures. As the incumbent nonprofit internet presence defending a neutral point of view and access for all, it is critical that Wikimedia maintain and strengthen itself to preserve a future with truly free knowledge.

Sections

Scale 02

Intro

Resilience refers to Wikimedia's ability to engender sustainable growth and to fend off threats. As the incumbent nonprofit internet presence defending a neutral point of view and access for all, it is critical that Wikimedia maintains and strengthens itself to preserve a future with truly free knowledge.

This paper recommends a number of countermeasures to be supported by the Wikimedia Foundation's Audiences department in order to bolster Wikimedia's resilience. A synthesized version on Officewiki is forthcoming as part of the Audiences department's 3-5 year planning FY 2018-2019.

External Threats

There are four major external threats to Wikimedia:

- Censorship
- Misinformation, principally from state actors or sophisticated PR firms
- Climate- and policy-related disasters
- Attacks on security or privacy by state actors

Internal Threats

There are also several ever present internal threats:

- Communities turning against each other
- Communities turning against newcomers
- Communities turning against themselves
- Communities turning against the Foundation
- The WMF turning against communities
- Wikimedia becoming irrelevant

Countermeasures

The following alternatives seek to address a number of the threats listed above. There are no one-size-fits-all countermeasures for the threats, and thus a set of the alternatives would likely need to be applied for a robust defense.

1. Consolidate to one domain name. Consolidate Wikimedia production traffic under one domain name. This will discourage DNS poisoning and make DNS poisoning and TLS negotiation-based blocking more evident when it does occur.

2. Give Huggle a hug. Support growth and diversity of the editor ecosystem through targeted product enhancement: adapt (possibly mainstream) tools like Huggle with low BRD (Bold, Revert, Discuss) reciprocation rates. Make these tools run on additional contemporary platforms, adding features to streamline guidance to good faith editors, with integrated follow-up discussion, and promoting praise of edits going through this BRD cycle. Shepherd appropriately sized coalitions of users focused on the new platform tools and updated approaches.

3. Decentralized internet distribution. Work with key experts and OS and browser vendors to build a secure protocol stack for decentralized distribution that

- Ensures availability
- Maintains content integrity and recognizable URLs (e.g., Signed HTTP Exchanges)
- Shields reading habits from intermediaries (e.g., inbuilt browser tunneling or use of trusted peer nodes)
- Shields metrics logging from intermediaries (e.g., opaque out-of-band logging)
- Reasonably accommodates protecting readers from outdated reverted material for the common consumption case. [3]

This is in addition to other resilient Wikimedia hosted solutions. Forthcoming enhancements to core protocols (e.g., DNS over HTTPS and ESNI coupled with proxying through critical hosting intermediaries) present additional opportunities to raise the costs of eavesdropping and denial of service.

4. **Database copies to more cloud storage providers and mirrors**. More proactively place Wikimedia dumps on BitTorrent, Github, Gitlab, BitBucket, AWS/S3/Cloudfront, Azure, GCP, Rackspace, Akamai, and Cloudflare. Also foster more mirroring relationships with a global network of universities. Consider coordinating with Google, Cloudflare, and Bing to serve as hosts for AMP as a fallback of last resort in case of widespread system outage or blockade. Apply cryptographic signatures to these distributions. This would provide redundancy and would create obstacles to censorship while allowing experts to better verify edit histories.

5. **Two factor authentication.** Add support for two-factor authentication for all interested users. Holding all other factors constant (no pun intended!), this is one of the surest ways to confound a broad class of attacks on security and privacy.

6. Invest in AI. Consider further investment in AI resources for:

- Liar, outlier, and bias detection
- Machine vision and speech-to-text
- Labeling and model tuning

This will be necessary for combatting bad faith state actors and PR firms. It will also be necessary to support a probable influx of multimedia content that needs moderation (and tagging and translation). Product opportunities for high value micro-contributions abound here as well.

7. Wikipedia all up. Begin streaming of algorithmic or volunteer curator (or both) selected content via one or more of the following means. Consider a consolidated global Wikipedia brand. Offer language content in one to thirty languages, depending on the format.

- Internet radio
- Global radio frequencies
- YouTube (with permissive syndication)
- Multicast for broadcast and cable television
- Satellite TV

This helps in further establishing a global brand presence as an information utility, swinging the door open for further future investment. It also creates an outlet for Foundation and Movement thought leaders to explain how Wikipedia works and why. It also raises the costs of censorship at comparatively lower costs of support. Finally, it is an opportunity for forging collaborative user groups for durability and a global brand.

It should be noted this concept could easily be applied in native fashion on various consumer appliances as well, although that is a separate product question.

8. Structured markup. Embrace distribution on syndicating and interactive agent platforms, utilizing partnership conversations for bespoke treatments where appropriate. A broader presence not only keeps Wikimedia relevant, it makes suppression harder - for two reasons: (1) when Wikimedia is part of the fabric of life people won't take kindly to it disappearing, and (2) when Wikimedia is everywhere it's technologically harder to suppress. Employ five principles:

A. Use of structured markup. As specific next steps, (i.) add Schema.org support to TemplateWizard and (ii.) conduct a consultation with the Wikidata and major wiki communities about the Wikidata community modeling templates using Schema.org and weaving that modeling into the non-Wikidata projects (by mainstreaming of Parsoid markup). This is an opportunity to build trust between communities and help define some functional roles for the future.
B. Ability to measure impact. It's important to know if and to what extent distribution is helping the cause.
C. Clear branding. This is important for brand presence and enforcement.
D. Attribution. This is important for compliance and staying true to Wikimedia's values.
E. Positive contribution feedback loop. Not all distribution platforms will have this capability, but contribution should always be intentionally encouraged, and ideally, co-designed.

9. Add Node.js and Python support to Templates. Add support for Node.js, and possibly Python, to Scribunto. Scribunto supports the Lua language, which is not widely used. It should support Node.js, and possibly Python, which has a huge developer following.

Steer volunteer engineering toward

A. template (Scribunto) scripting, gadgets, and bots
B. improving MediaWiki Core

This places a higher emphasis on growing content and workflows for the wikis in a more sandboxed fashion while simultaneously making basic MediaWiki more excellent software for collaborative knowledge production (a global ecosystem form of resilience). Further investment in first class global templates, ideally with a mechanism to fuse data with Wikidata, is complementary. These new technologies are an opportunity to consider more contemporary code contribution workflows.

Scale 05

10. Fund anti-surveillance and anti-censorship research. Provide funding to 1-2 reputable anti-censorship / anti-surveillance firms (or fund incrementally internally). This lets more sophisticated forms of distribution and protection be developed.

Summary

The following table is a guide to the countermeasures, how they address the major threats, their relative cost, and how the countermeasures might complement other efforts

Countermeasure	Threats addressed	Cost and horizon	Complements
1. Consolidate to one domain name Content and APIs are all served from wikipedia.org. Censorship of one language is censorship of all, which is costly for censors.	→ Censorship → Attacks on security or privacy by state actors	Medium, two year project with cross-functional team at 50%	Brand unification under Wikipedia
2. Give Huggle a hug Get Huggle on Android and iOS. Improve its UX. Invest in productive in-app feedback loops. People work nicely with each other, more editors stay around, the ecosystem flourishes.	→ Communities turning against newcomers → Communities turning against the Foundation	Medium, three year project for one team	Making wiki projects bustling neighborhoods
3. Decentralized internet distribution Host Wikipedia in a decentralized fashion with secure tunneling and digital signing. Wikimedia is accessible even when servers are down or blocked. This is in addition to other resilient Wikimedia hosted solutions.	→ Censorship → Climate- and policy-related disasters → Attacks on security or privacy by state actors → Wikimedia becoming irrelevant	Big, five year project for one small team with support from several other teams. Incremental milestones.	Eventually, offline editing
4. Database copies to more cloud storage providers and mirrors Digitally verifiable database dumps become more pervasive. It's even harder to erase Wikimedia and its chain of edits.	→ Censorship → Misinformation, principally from state actors or sophisticated PR firms	Small, one year project with one additional dedicated FTE	Academic research outreach

Countermeasure	Threats addressed	Cost and horizon	Complements
5. Two factor authentication Anyone who wants it gets the option of two factor authentication. Account compromise becomes much harder.	→ Attacks on security or privacy by state actors → The Foundation turning against communities	Medium, 18 month project with three dedicated FTEs and recurring SMS fees	Potentially, scoring and brand positioning
6. AI Investment in liar, outlier, and bias detection; machine vision and speech to text; labeling and model tuning. It's easier to spot the bad guys. It's also easier and more fun for users to interact with and moderate content	→ Misinformation, principally from state actors or sophisticated PR firms → Communities turning against each other → Communities turning against newcomers → Wikimedia becoming irrelevant	Large, 5 year project with paradigm shift for Audiences and Technology - varying levels of commitment by team.	Translation, scoring, mobile contribution and AI training, multimedia contribution, oral history
7. Wikipedia All Up Content is streamed online, over the airwaves, and by satellite. People can catch Wikimedia anywhere. Wikimedia is a trusted brand everyone knows will always be there, even for those without computers or smartphones. It's harder to block an omnichannel presence.	→ Censorship → Misinformation, principally from state actors or sophisticated PR firms → Wikimedia becoming irrelevant	Medium, 3 year project with small team with escalating brand penetration	Brand unification under Wikipedia
8. Structured markup Using structured markup and partner management, Wikimedia content is further embedded online, with impact measurement and Wikimedia values. Interactive agents automatically rely on the structured markup.	→ Censorship → Wikimedia becoming irrelevant	Medium, 3 year project with small team with escalating brand penetration	Granulariz-ation of the article, translation
9. Add Node.js and Python support to Templates Would-be template editors no longer need to use Lua for scripting (Scribunto), they can also use programming languages they know and love. This allows a key piece of the ecosystem to grow and thrive.	→ Wikimedia becoming irrelevant	Medium, 2 year project with one dedicated FTE and one code review/tester	Global templates
10. Fund anti-surveillance and anti-censorship research The next round of privacy tools gets researched and built while we pursue efforts on the current tools.	→ Censorship → Attacks on security or privacy by state actors	150K-300K annual investment	

Other Considerations

The following items are arguably not major Audiences efforts in and of themselves, but represent potential opportunities for other departments.

Cooperate with Technology on a continuity plan in case both primary data centers go down for an extended period due to climate or policy disaster.
Explore international governing body action on censorship on the basis of anti-competition (e.g., most blocks have corresponded with unfairly positioned state-supported alternatives) or adverse health and safety externalities (medical information and other critical information has become unavailable). This is a longshot, and the consequences for scrutiny on the content and the positioning as an NGO would need to be considered, but it may provide a defense.

Conclusion & Other Opportunities

The countermeasures preempt future, and in some cases squash current, threats. You'll notice that they are also oriented around the space where the Wikimedia Foundation is uniquely positioned to take action, as these are large and difficult efforts requiring personnel. The recommendations don't fully capture the range of discussions or feedback I received during late September and early October 2018 as part of the 3-5 year planning process.

Much more is said in other theme and subtheme documents as part of the Audiences 3-5 year planning point of view about potential community or feature interventions, but the following, which is heavily informed by my recent conversations, are examples of how to aid in resilience in various other ways. Some of them overlap with material in other documents. They principally speak to creating the content and ecosystems that can activate and sustain growth, which is germane to the general theme of SCALE, as well as several other themes.

- Abuse filters
- Creating spaces to inform editors where there is surging demand or probable surging demand (based on algorithms) for topics and those topics do not yet meet a particular quality bar.
- Driving programs to encourage bilingualism. Exploring with professors the concept of translation proofreading as coursework.
- Ensuring inflows of translations into English Wikipedia and other major wikis.
- Investing in generalized work backlog solutions, catered for various personae and form factors.
- Emphasizing product experiences for mobile that are catered principally for AI training.
- Supporting federated SSO with major social identity providers, and flowing contribution activity back to user social channels.
- Scaling analysis of interventions by further integrating with academics in our data analysis.
- Partnering with a provider such as Coursera on a free course such as Programming Wikimedia: APIs, Bots, Gadgets, and Template Scripting.
- Supporting content snapshots (i.e., branded, perma-linkable, countable, attributed hypermedia fragments) for

embeddable content. This would be a complement to the summary endpoint and context cards.

Notes

[1] Principally from state actors or sophisticated PR firms
[2] For example, veteran contributors working against newcomers.
[3] Note: risk concerning potentially infringing content, perhaps avoided by simply obfuscating discovery, needs analysis.

Sources

A. Baso and O. Vasileva Research and Insights, Other contributors: B. Davis, A. Halfaker, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

Early onset of structural inequality in the formation of collaborative knowledge, Wikipedia

Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity

Modeling crowdsourcing as collective problem solving

Robust clustering of languages across Wikipedia growth

The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to popularity is causing its decline

Chris Dixon essay Why Decentralization Matters and scanned Build Your First Ethereup DApp, reading selected sections.

ubiquity

[unedited] In order to achieve ubiquity, we need to focus on the following themes:

Making our content adaptable to all platforms instead of focusing on future trends
Structuring our content
Unifying content through global tools
Investing in multimedia
Growing and connecting our projects

Investing in syndication without losing the essence of our projects
Studying the current effects of syndication
Demanding attribution
Making the risks to our future known to companies and individuals invested in or dependent on our content
Investigating models of accountability for companies using our content
Investigating alternate modes of content creation

Ensuring we have constant production of relevant content
Growing smaller communities
Addressing knowledge gaps
Balancing gaps in interest - creating content that readers are interested in

Sections

Scale 02

Platform agnosticism

"Stroll through Sanlitun, a bustling neighborhood in Beijing filled with tourists, karaoke bars, and luxury shops, and you'll see plenty of people using the latest smartphones from Apple, Samsung, or Xiaomi. Look closely, however, and you might notice some of them ignoring the touch screens on these devices in favor of something much more efficient and intuitive: their voice." [1]

It seems the Chinese language, as many other languages, was not built for typing tiny letters on a small screen. But that's okay because technology, as it usually does given large-enough demand, is making its way around such technical difficulties. In this particular case, the answer might be voice search, but there are other cases where AI or messaging play bigger parts.

Many are coming to the internet with new needs, new languages, and new modes of expression and it is certain that their arrival will change the fabric of the ways and forms in which knowledge is created, shared, and used. As internet usage in growing economies rises, the internet will become a more diverse place and will be required to adapt to the needs and motivations (our own research shows that motivations vary greatly based on project) of its new users.

If Wikimedia projects want to be present along with this growth or even further and our goal is to "break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge", we must ensure adaptability to any platform or mode of usage.

Yet predicting trends can be tricky and the risks that have prevented us from being at the forefront of technical innovation so far still apply. Unlike Google, who have the resources to do everything-everywhere-all-the-time, we do not have the luxury or expertise in taking large risks, especially if they do not come from the ground up, from our communities.

For us, ubiquity must mean adaptation - skipping the guessing game of what will be big in the future, investing in the needs our current and potential communities and making sure that our content is prepared for use in any future trend and for presentation on any device.

Thoughts on ways we can achieve this:

- Structuring our content so that it can be reusable by ourselves, our communities, and other platforms
- Creating tools that our communities can use for structure such as establishing global templates or article templates. Global customizable templates will make the creation of APIs easier. It can allow individual wikis to structure their content as they wish while ensuring consistency in new features and presentation. Article templates can structure our content by section or idea - making it easier to use portions of content inside and outside of our platforms.

- Providing the building blocks for technical tools to be built by the community. If we provide the framework for community-built tools, we can ensure that they will work smoothly with the remainder of our content. This will empower communities to build tools that are necessary for their own projects while allowing us to focus on the structure of projects themselves.
- Investing in technologies currently used by emerging economies.
- Investing in internal platform and project agnosticism by decreasing the separation between our projects. (Creating workflows that will smoothly move an interested user from the Wikipedia article on Istanbul, to the Wikivoyage guide, to images and videos about the city from Commons)

Syndication

"Wikipedia content appears to play a substantially more important role in the Internet ecosystem than anticipated, with other websites having critical dependencies on Wikipedia content."

"Google becomes a worse search engine for many queries when it cannot surface Wikipedia content (e.g. click-through rates on results pages drop significantly) and the importance of Wikipedia content is likely greater than many improvements to search algorithms. Our results also highlight Google's critical role in providing readership to Wikipedia. However, we also found evidence that this mutually beneficial relationship is in jeopardy: changes Google has made to its search results that involve directly surfacing Wikipedia content are significantly reducing traffic to Wikipedia." [2]

So far, Wikipedia's relationship with Google has been fairly symbiotic. We provide a trusted source they can show at the top of the page; they provide an increase in pageviews and, in turn, an increase in donations, in new editors, and in the continued creation of quality content they can then show to users. Everybody wins and information is distributed freely.

Yet exposing more information outside of the site, such as in Google's knowledge panels, has decreased pageviews to Wikipedia. It is unfortunate that this is an issue. We still completed our goal of providing the information a reader seeked. Yet without the extra pageviews we face not only a decrease in funds, but eventually a decrease in quality. Potential editors never see the site and current editors have less motivation to continue writing. Over time, we're in trouble.

But, so is Google. The study quoted above displayed, clearly, that Google is a worse search engine in a world without Wikipedia. Wikipedia's importance is so large that the "mere presence of Wikipedia links may have an effect approximately 80 times larger than the difference between a good ranker algorithm and bad one (for many queries)". Similar patterns have been found for other online websites such as Reddit and StackOverflow, where Wikipedia content is widely shared.

Scale 04

Thus we find ourselves in an odd paradox where our current ubiquity is becoming a threat to itself. One option would be to take a purely defensive stance and work towards preventing any information from usage outside of the site but it is needless to say that that goes against the morals of our movement severely. The other option would be to take syndication for granted - to imagine our content spread throughout the fabric of the internet, and shift our content creation and revenue model to such a future.

Thoughts on ways we can achieve this:

- Opening conversation with our partners on these relationships asap
- Working with our partners to ensure proper attribution of information - not only for pageviews but also so readers are not misled by the presentation of information taken out of context.
- Creating standards that will hold larger organizations accountable, financially or legally, to the growth of the project they have become dependent of
- Inspiring organizations to invest in the future of Wikimedia projects by becoming editors or inspiring others to do so - for example, one of the papers included suggestions such as getting karma points on reddit for improving linked articles
- Working with our partners to create symbiotic ways of using our content (such as exploring features like context cards, which share a preview of wikipedia articles with direct links back to the original article page, outside our projects)
- Working outside the boundaries of our platform - creating workflows of contribution from other places where our content is used.

Content Relevancy

"In the English Wikipedia, articles of strong insufficient quality alone receive close to half of the pageviews, and in the Russian Wikipedia, they receive more than half." [3]

For our projects to be ubiquitous, we must provide relevant content to all of our users. Not all wikis are the same, nor do they grow in similar fashions and users of different projects have widely varied motivations for reading. For example, our research shows that readers in Western-language wikipedias are more likely to focus on quick-fact information whereas the speakers of languages in growing economies are more likely to use Wikipedia for deeper learning and for work or school purposes. To be able to cater to the needs of individual wikis or groups of wikis, we must be able to distinguish their needs. Features that might work great on one wiki, might not work on another. Similarly, content that might be notable for a particular community, might not be for another.

Focusing on targeting our work to match our unique audiences as well as providing them with the tools to build according to their needs will help us cover the entire range of content that our current and future readers will require.

Thoughts on ways we can achieve this:

- Creating tools that will aid growing and established communities to identify and create the content that readers are interested in such as suggestion algorithms for sections and other type of content, article templates, improved content translation, and reader requests. These tools can also be used to focus on content diversity and, eventually, on editor diversity.
- Focusing product initiatives on targeted groups of communities or geographic areas. We can focus on assisting projects according to their needs (for example, larger wikis that tend to be less friendly over time can gain from focusing on editor retention, while smaller wikis working on establishing their communities can focus on converting readers to editors).

Bits and bobs

Other thoughts on the subject of ubiquity:

- Sustainable models of growth - modeling the needs of readers, editors, and moderators to highlight imbalances in projects (for example, quality content but low readership, or large amount of content but low quality)

Notes

[1]https://www.technologyreview.com/s/600766/10-breakthrough-technologies-2016-conversational-interfaces/

[2] The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies

[3] https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15623/14799

Sources

O. Vasileva: Research and Insights, Other contributors[1]: A. Baso, B. Davis, A. Halfaker, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

https://econsultancy.com/why-we-need-to-stop-repeating-the-50-by-2020-voice-search-prediction/

The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies [20]

Examining Wikipedia with a broader lens: Quantifying the value of Wikipedia's relationships with other large-scale online communities [22]

Creative commons non-commercial license
https://meta.wikimedia.org/wiki/Free_knowledge_based_on_Creative_Commons_licenses#8._

Can_NC-licensed_content_be_used_in_the_Wikipedia_project?_%E2%80%93_No._Wikipedia_contents_are_being_used_commercially.

Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information [23]

Wikimedia movement strategic direction
https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Direction

A study of versions of articles available across 26 wikis. Identifies that project growth seems unrelated to linguistic or cultural factors.

Robust clustering of languages across Wikipedia growth

Studies multiple cross-language wikipedias to determine connections between articles in different languages are not directly correlated to their existence or quality in enwiki
https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-016-0070-8
Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

A study of inequality in content among different language projects. Identifies common growth patterns among wikis https://arxiv.org/abs/1610.06006

The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to popularity is causing its decline https://www-users.cs.umn.edu/~halfaker/publications/The_Rise_and_Decline/halfaker13rise-preprint.pdf

A visualization that displays the differences in participation between various projects - could be used as an indicator of project growth cycleshttps://dl.acm.org/citation.cfm?doid=3173574.3173929

Modeling crowdsourcing as collective problem solving - models productivity for crowdsourcing tactics https://www.nature.com/articles/srep16557

Research on knowledge gaps, esp. Recommendation systems https://research.wikimedia.org/knowledge-gaps.html

New editors research - highlights the needs and challenges of upcoming editors on established projects https://www.mediawiki.org/wiki/New_Editor_Experiences

Displays that knowledge is more effective when actively sought

Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics https://onlinelibrary.wiley.com/doi/full/10.1111/jcc4.12185

An op-ed detailing the increasing predominance of multimedia over written content https://www.nytimes.com/interactive/2018/02/09/technology/the-rise-of-a-visual-internet.html

The Substantial Interdependence of Wikipedia and Google: A Case Study on the Relationship Between Peer Production Communities and Information Technologies - shows that Google's clickthrough rates drop significantly when not surfacing Wikipedia content, but that Wikipedia's pageviews drop when Google displays such content https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15623

A study of usage of CC non-commercial licenses
Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3234435/

Usage of voice search - increases over time. Looks at whether by 2020, 50% of the world will be using voice search. https://econsultancy.com/why-we-need-to-stop-repeating-the-50-by-2020-voice-search-prediction/

https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017/Direction/Appendix#pattern3

Audiences x 3 / 5 Year Planning 2018

tools

The following are draft posi
tion papers exploring the
need for tool development
based on input from Comm
nity and WMF experts, as w
as synthesis of secondary
research.

01
For Developers

02
For Organizers

03
For Moderators

Tools Overview 00

In early 2006, a large-scale vandalism attack happened on the English Wikipedia with thousands of articles wiped out. The editors were defenseless and the site was vulnerable. Then suddenly, four Wikipedians came together to write TawkerBot, the first anti-vandalism bot[1] for Wikipedia. This bot proved to be a life-saver for the site. Today more than 300 bots work round the clock on English Wikipedia to ensure the smooth functioning of the site. Tools like Twinkle[2] (tool library that helps editors perform wiki maintenance), Huggle[3] (a diff browser for rapidly reverting vandalism), HotCat[4] (allows a user to easily add and remove categories to pages), AutoWikiBrowser[5] (semi-automated MediaWiki editor), etc. drive many of the tasks power editors do on English Wikipedia every day.[6]

At the same time, smaller language wikis like Hindi Wikipedia[7] have problems coping with vandalism and keeping up with content moderation needs. Unlike English Wikipedia, they don't have the corps of volunteer developers able to write tools to defend and curate the site's content. It's a lot harder for those communities to grow their content or their editor base, because the active contributors are stuck doing manual drudge work that the bigger wikis automated years ago.[8][9]

Tools for Developers

Empowering our volunteer developers to write better code that can work across wikis is going to be a key factor in helping us gather the sum of all knowledge. Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, keep track of the quality of incoming edits, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors able to write code for the site. This creates disparity in the resources available to wikis. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis.

Tools for Organizers

Movement organizers are another key audience for tool development in the coming years. The Foundation's 2018-19 annual plan recognizes organizers as "fundamental implementers" and a "core asset" of the free-knowledge movement. But tools that support organizers' efforts are frequently ad-hoc, poorly documented and available only on certain wikis. Access problems can be particularly acute in smaller communities, where the technical skills required to set up and run bots, scripts and other technologies are often scarce.

Organizers' needs fall into a three main areas. "Community-building" tools are required to help organizers inform, engage and manage the work of their communities. "Outreach and promotion" tools will help

organizers advertise their activities and recruit new members. "Event-management" tools are necessary to more efficiently carry out tasks like event signup and conference scheduling.

Finally, two overarching meta-problems are key areas of interest among organizers. One is the need for better guidance about best practices and the tools that do exist. As one organizer put it, "There are a lot of tools we don't know about or know what they can do for us. We need someone to help us understand what we are missing, and what to do and how to do it." The second is the need for a mechanism that can replace or augment categories, so that organizers will be able to classify content effectively and more efficiently tap volunteers' subject interests—a primary motivator, especially of new editors.

Tools for Moderators

A critical, but often overlooked, aspect of the workflows that make our projects successful are the tools and processes used to review and moderate our content. For the most part, the Wikimedia Foundation has taken a hands-off approach to content moderation tools and let the community develop their own solutions (with a few exceptions such as Recent Changes filtering). As one would expect, the systems built by the community utilize the building blocks at their disposal: templates, gadgets, user scripts, and wikitext. These tools and processes suffer from several significant problems, many of which have already been mentioned above: lack of portability, limited capabilities, lack of automated testing and code review, lack of localization support, etc.

Another major problem with these tools and processes, especially those created for content moderation, is their high learning curve. As one example, on English Wikipedia there is a system for submitting, reviewing, and publishing draft articles called Articles for Creation (AfC). In order to participate as a reviewer in AfC, you have to install a special user script, be granted a special right though a unique vetting process, and use several obscure templates and categories. The complexity of this process limits the number of people who are able and willing to participate, which in turn leads to a less diverse pool of reviewers. This lack of diversity may contribute to problems of systemic bias in our content.[10][11] The small number of reviewers also makes the review process slow, often taking a week or longer to review a submitted draft. This is likely to contribute to the overall effect of the process, decreasing newcomer productivity.[12] Unless we make these moderation tools work for less technical users, it is unlikely that the pool of moderators will grow or diversify.

Similar examples can be found throughout the moderation processes for our projects, including workflows for article assessment, deletion, and problem triaging; workflows for reviewing edits; workflows for reviewing and organizing multimedia contributions; workflows for proofreading Wikisource transcriptions; and more. While the Wikimedia Foundation has historically focused on building software for reading and editing content, the other critical pieces of the wiki ecosystem have been largely neglected, leading to volunteers feeling overwhelmed and unsupported. In a 2015 survey of experienced editors across 10 projects, only 35% said that the Foundation was mostly or completely meeting the community's technical needs around identifying and surfacing content problems.[13] Clearly, there is a lot of work for us to do in this area as we have only scratched the surface thus far. If we want to increase the capacity of our communities to efficiently and effectively moderate content,

it is time for the Foundation to begin investing seriously in this area.

Unfortunately, the Foundation's hands-off approach has resulted in a lack of credibility in this area. To build our credibility, we should first focus on the areas where there is a clear need for better tools, such as fighting vandalism and sock-puppetry. We should also investigate how editors transition into becoming moderators so that we can better facilitate that transition. Once we have proven our capacity to understand their motivations and work with moderators to build effective tools, we will then have the mutual trust needed to tackle more difficult workflows such as article deletion and conflict mediation.

Examples

HotCat

Huggle

Twinkle

AutoWikiBrowser

Programs and Events Dashboard

Wikimedia Cloud Services

CentralNotice

GeoNotice

Areas of Impact

Templates

Gadgets

Bots

Editing and Administration APIs

Discussion systems

Messaging systems

Contributor Analytics

Developer Advocacy and Outreach

Translation and Localization Infrastructure

API and Tool Documentation

1. Here TawkerBot is given the first ever approval to run on English Wikipedia:

References 00

https://en.wikipedia.org/wiki/Wikipedia:Bots/Requests_for_approval/Archive_1#Tawkerbot

2. Wikipedia:Twinkle
3. Wikipedia:Huggle
4. Wikipedia:HotCat
5. Wikipedia:AutoWikiBrowser
6. Geiger, R. S. (2014) “Bots, bespoke, code and the materiality of software platforms. Information”, Communication & Society, 17(3), 342-356.
7. https://hi.wikipedia.org/wiki/विकिपीडिया:चौपाल/पुरालेख_1#vandalism
8. https://hi.wikipedia.org/wiki/विकिपीडिया:चौपाल/पुरालेख_19#बर्बरता
9. https://hi.wikipedia.org/wiki/विकिपीडिया:चौपाल/पुरालेख_33#गंगाधर_नेहरू_पृष्ठ_में_बर्बरता
10. Lam, Shyong K., et al. (2011). “WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance”, WikiSym ’11.
11. Purtill, Corinne; Schlanger, Zoë (October 2, 2018). “Wikipedia rejected an entry on a Nobel Prize winner because she wasn’t famous enough”, Quartz.
12. Schneider, Jodi, et al. (2014). “Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review”, OpenSym ‘14.
13. https://meta.wikimedia.org/wiki/Research:Tech_support_satisfaction_poll

tools for developers

Empowering our volunteer developers to write better code that can work across wikis is going to be a key factor in helping us gather the sum of all knowledge. Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, keep track of quality of incoming edits, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors able to write code for the site. This creates disparity in the resources available to wikis. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis.

Sections

Tools 02

Pan-wiki Tools Platform

Empowering our volunteer developers to write better code that can work across wikis is going to be a key factor in helping us gather the sum of all knowledge. Wikis need code contributors as much as they need content contributors. Templates, gadgets, and bots act as superpowers in making editors more efficient at their tasks. Experienced editors use these tools to create and maintain backlogs, keep track of quality of incoming edits, perform mass actions on pages, ward off vandalism and more. However this superpower is limited to wikis which have contributors able to write code for the site. This creates disparity in the resources available to wikis. Bringing these important resources to all wikis is fundamental to bridging the equity gap across all language wikis.

This paper advocates for building a platform that can support tools which work on all our wikis seamlessly. Right now a lot of developer code lives on the wikis (gadgets, Lua modules, templates) where it really isn't possible to do any type of testing, code reviews or debugging; nor is there any straightforward way to add localization or RTL support. This often leads to issues like security vulnerabilities,[1] conflicts with MediaWiki deployed extensions,[2] and bugs due to lack of maintenance. Also, in its current state, having code hosted on the wikis (in a per-project fashion) makes it hard to get in the mindset of having the code work across wikis. It's easy to get sucked into customization and forget to think about things like RTL rendering or localization.

One aspect of the future platform depends on services being available to the developers and communities which they can use for building better tools. These may be services which can be used to do better copyright violation detection, vandalism detection, and image recognition, and provide access to better statistics, and so on. Part of the growth of services involves better partnerships with companies like Google, Turnitin and others providing such services. Another very important aspect of the platform is for Wikimedia engineering to collaborate with our volunteer developer communities to come up with documentation and best practices for creating new tools.

An example of this can be tutorials and guidelines on how gadgets can make use of OOUI [3] to standardize our interfaces and make them more accessible for everyone. Tools that facilitate communication among engineers and volunteer developer communities is key to achieving this goal.

Bots : Potential and Risks

[not edited] Bots are an important part of a wiki's infrastructure. They perform a lot of repetitive tasks that need much effort from the editors.

Bots can can have a big influence in shaping a wiki's community practices and the encyclopedia itself. This can be dangerous because most bots are programmed by individual programmers without a lot of community input.

Tools 03

Gadgets & Gadget Usage

[not edited] 16 gadgets are enabled by default for all users on English WP [4]

8 other gadgets are manually enabled by more than 30,000 active users

The numbers are comparatively low on other projects like German[5], Hebrew,[6] Tamil,[7] Italian [8] etc.

Several prominent gadgets have been adopted by several projects. Like HotCat, Navigation popups, UTCLiveClock, WikEd etc. have been made available on several different projects by volunteers who are familiar with programming.

There is no good way to test gadgets before deploying them. This means that gadgets can break things quite easily for users while being hard to detect as the source of the breakage. They also conflict with MediaWiki extensions at times, causing broken behavior. [9] [10] [11]

Most importantly, gadgets do not go through any sort of review process before being deployed which means they are a potential vector for abuse.

Conclusions & Implications

[not edited] Gadgets are very heavily used on most Wikipedias

Bigger projects (with more language speakers) like English have a lot more gadgets at their disposal than smaller projects.

Knowledge of some basic programming seems is an inhibiting factor for porting gadgets to new wikis. This is because very often gadgets are written in a way that is custom to the wiki they are written for. Internationalizing these gadgets is hard and is seldom paid attention to. Similarly, the UI is also commonly customized to the wiki it's written for and does not work well for RTL/LTR wikis.

We don't currently have any instrumentation in place to gauge how often each individual gadget is actually used in practice. This is important for two reasons:

Such instrumentation is a great way to figure out which tools are desired by the communities which are not available as part of MediaWiki and the extensions. This will allow us to fill that gap by creating tools that work across all our projects well.

Instrumentation is also important because the default enabled gadgets are loaded for every single registered user. This can amount to a significant JavaScript load for every user that needs some investigation to gauge performance impact. [12]

Wishlist & Technical Requests

[not edited] One of the big realizations we have had working on the Community Tech team is that a big chunk of wishlist survey wishes every year already have an existing solution created by the community. Here's a few examples:

- Global cross-wiki watchlists asked for in the 2015 Wishlist survey were

Tools 04

worked on by a volunteer dev who made a tool called crosswatch.[13] It stopped working because of lack of maintenance.

- Pageview stats tool also asked for in the 2015 wishlist survey was already worked on by a volunteer.
- GlobalPreferences extension was built by Legoktm in his volunteer capacity in 2013, much before it got nominated for the wishlist.[14]
- TemplateWizard[15] has several local wiki equivalents - on Hebrew [16] and German [17] wiki, developed by volunteer developers.
- Blame tool has a few existing solutions built by the community already.[18] [19]
- English Wikipedia has a page to request bots for specific tasks.[20] Similar pages exist on other wikis. At a quick glance, one can see that most of the uncontroversial requests are pretty quickly fulfilled by volunteer developers.

Conclusions & Implications

- Volunteer developers are extremely good at helping our communities when they can.
- Often wikis wish something which is already implemented on another wiki as a gadget/bot but there is a communication gap between different wikis as each wiki has its own ecosystem of tools it uses. There is no single place to surface tools used by different wikis. Also when wikis do discover these tools, there is lack of documentation to guide non-technical folks on importing and enabling these tools on their own wikis.
- They have the potential to act as an extended arm of WMF, in helping our communities grow and become better.

Tools 05

Notes

[1] https://lists.wikimedia.org/pipermail/wikitech-l/2018-March/089636.html
[2] https://phabricator.wikimedia.org/T170896
[3] A UI component library created by WMF, aimed at providing a consistent UI experience that works well for all languages. https://www.mediawiki.org/wiki/OOUI
[4] https://en.wikipedia.org/wiki/Special:GadgetUsage
[5] https://de.wikipedia.org/wiki/Spezial:GadgetUsage
[6] https://he.wikipedia.org/wiki/מיוחד:GadgetUsage
[7] https://ta.wikipedia.org/wiki/ �ப. ட:GadgetUsage
[8] https://it.wikipedia.org/wiki/Speciale:GadgetUsage
[9] https://phabricator.wikimedia.org/T178348
[10] https://phabricator.wikimedia.org/T170896
[11] https://phabricator.wikimedia.org/T22134
[12] https://phabricator.wikimedia.org/T142461
[13] https://tools.wmflabs.org/crosswatch/welcome
[14] https://www.mediawiki.org/w/index.php?title=Extension:GlobalPreferences&oldid=838650
[15] https://www.mediawiki.org/wiki/Help:Extension:TemplateWizard
[16] https://he.wikipedia.org/wiki/מדיה_ויקי:Gadget-TemplateParamWizard.js
[17] https://de.wikipedia.org/wiki/Wikipedia:Technik/Skin/Gadgets/Vorlagenmeister
[18] https://github.com/wikiwho/WhoColor
[19] http://wikipedia.ramselehof.de/wikiblame.php
[20] https://en.wikipedia.org/wiki/Wikipedia:Bot_requests

Sources

N. Kohli, R. Kaldari and J. Matazzoni : Research and Insights, Other contributors[1]: A. Amaroni, A. Baso, B. Davis, A. Halfaker, J.Hare, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, N. Wilson,

Critical Point of View: A Wikipedia Reader - The Lives of
Bots by R. Stuart Geiger https://en.wikipedia.org/wiki/Special:GadgetUsage

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Tools 01

tools for moderators

The tools and processes used for reviewing and moderating content are critical to the growth of the projects, but have historically been overlooked. Communities have, for the most part, been left to develop their own moderation workflows, tools and solutions – which means that community-built systems tend only to utilize the building blocks at their disposal: templates, gadgets, user scripts, and wikitext. This approach results in solutions with limited capabilities, that tend not to be portable, lack automated testing and code review, and don't support localization.

Many of these tools and processes also have steep learning curves and require considerable technical acumen to use.[1] The more complex the process is, the more it limits the number of people who are able or willing to participate. This, in turn, leads to a less diverse pool of reviewers and, we suspect, contribute to problems of systemic bias in our content.[2][3] And, fewer moderators makes the review process slow, impacting newcomer productivity.[4] In order to grow and diversify the pool of moderators, moderation tools must be made to work for less technical users, and the Foundation must drive this investment in a platform which enables community developers to create more portable, localizable tools, and in resources to support better practices in code development.[5]

Sections

Tools 02

Productivity

[not edited] Since Wikipedia's dramatic rise in popularly around 2009, it has faced an endless stream of spammers, vandals, PR firms, and "clueless" newbies.

- Due to this onslaught, English Wikipedia began directing new users to create new articles in the Draft namespace and seek review before publication.
- This new process, Articles for Creation (AfC), ensures that new articles measure up to Wikipedia standards for notability and verifiability.[6]
- AfC can, however, be a frustrating and discouraging process for newcomers.
  - "New article creation is a battlezone where socializing newcomers seems to take a back seat to ensuring quality control."
- Since AfC was introduced, the proportion of articles of each new page creator that survive at least 30 days has clearly declined.
- The paper suggests that one reason for this is that draft articles are hidden from potential collaborators.
- It also suggests that the review process is often slow and confusing.
  - 29.4% of AFC drafts are never submitted for review.
  - 11% of reviews take longer than a week.

Implications

- We should create better tools for the community to review new articles with.
- We should try to increase the percentage of drafts that are reviewed.
- We should try to decrease the number of reviews that take a week or longer to happen.
- By improving these metrics, we will hopefully improve newcomer productivity.

What Experts Say

[not edited]

- Wikipedia has a very strong need for better tools to deal with vandalism and sock-puppeting.
- Blocks are easy to evade, especially for dedicated vandals who know how to use proxies.
  - Current proxy blocking tools aren't very effective, especially in combating zombie proxies. [7]
- Community-built tools are important for moderators, but have lots of problems.
  - Many are abandoned and end up breaking.
  - Often they are specific to one wiki (e.g. Twinkle).
  - Many are complicated to install and use.
- Discussion-based workflows are often overly complicated and would benefit from software to streamline the process.
  - Examples: Articles for Deletion, Bot Request for Approval, Article Peer Review
- Tools must be built in close collaboration with the community to be successful.
  - PMs should provide weekly updates for the community, post

Tools 03

screenshots, ask questions, respond to feedback.

Implications

- This is a bullet
- Our top priority for tools for moderators should be counter-vandalism tools.
  - We should build tools to automatically block proxies.
- Our next priority should be tools to facilitate discussion-based workflows like Articles for Deletion.

We should establish clear guidelines for collaborating with the community on tool development to make sure we aren't building things in isolation.determined to be different enough to warrant a shift.

Decision Quality

[not edited] [8]

- Disagreements and conflicts on Wikipedia are frequent and inevitable. Effective decision-making and conflict resolution processes are essential to community health.
- The Articles for Deletion (AfD) process is one of the core moderation processes on English Wikipedia.
- AfD works through a consensus process rather than voting.
- Many factors affect the quality of AfD decisions, i.e. how likely they are to later be reversed.
  - Larger groups of participants make better decisions.
  - Newcomer participation yields worse decisions, as they often lack understanding of Wikipedia norms and processes.
  - Decision quality is generally unaffected by recruitment, unless the recruitment is heavily biased.
  - Diverse groups may make better decisions.
  - Biased admins make worse decisions than impartial admins.

Implications

We should built tools to facilitate more people participating in moderation processes such as Articles for Deletion, since this will likely improve the quality of the decisions.

These tools should provide onboarding support and education for newcomers so that they can participate constructively.

Tools 04

Notes

[1] Articles for Creation (AfC), just one example from English Wikipedia, is a system for submitting, reviewing, and publishing draft articles. In order to participate as a reviewer in AfC, you have to install a special user script, be granted a special right though a unique vetting process, and use several obscure templates and categories.

[2] Lam, Shyong K., et al. (2011). “WP:Clubhouse? An Exploration of Wikipedia’s Gender Imbalance”, WikiSym ’11.

[3] Purtill, Corinne; Schlanger, Zoë (October 2, 2018). “Wikipedia rejected an entry on a Nobel Prize winner because she wasn’t famous enough”, Quartz.

[4] Schneider, Jodi, et al. (2014). “Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review”, OpenSym ‘14.

[5] Code review procedures, documentation standards, etc.

[6] https://en.wikipedia.org/wiki/Wikipedia:Articles_for_creation

[7] A zombie proxy is a proxy server being run on a computer that has been compromised by hackers.

[8] The Effects of Group Composition on Decision Quality in a Social Production Community https://dl.acm.org/citation.cfm?id=1880083

Sources

R.Kaldari : Research and Insights, Other contributors[1]: D. Horn, J. Katz, N. Kohli, J. Matazzoni, J. Minor, T. Negrin, M. Novotny, N. Pangarkar, L. Zeimba

Accept, decline, postpone: How newcomer productivity is reduced in English Wikipedia by pre-publication review https://dl.acm.org/citation.cfm?doid=2641580.2641614

https://en.wikipedia.org/wiki/Wikipedia:Articles_for_creation

Interview with Leon Ziemba Leon is a long-time volunteer Wikipedia editor, administrator, and tool author. He was hired by the WMF Community Tech team in 2016.

https://dl.acm.org/citation.cfm?id=1880083

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Tools 01

tools for organizers

The Foundation's 2018-19 annual plan recognizes organizers as "fundamental implementers" and a "core asset" of the free-knowledge movement. But tools that support organizers' efforts are frequently ad-hoc, poorly documented and not universally available—particularly to smaller communities. As the movement puts an increasing emphasis on knowledge equity, the need to understand and support movement organizers is more vital than ever. This white paper is an early effort to analyze and document organizers' main areas of need. This examination will be followed and deepened soon by the annual plan-mandated Movement Organizer Study.

Sections

Tools 02

Who are organizers and what do they contribute?

For the purposes of this analysis, a "movement organizer" is a anyone who purposely seeks to motivate, attract and/or direct volunteer attention. Some organizers work independently, possibly having been trained by more formal groups. But most operate within the framework of various structures that support their efforts, to wit:

- Chapters are legal, nonprofit entities incorporated primarily on a per-country basis, with a few exceptions (e.g., New York and DC chapters). There are about 40.
- User groups are not legal entities. They may have a regional or thematic focus (e.g., Community User Group of Greece, Wiki Medicine). There are about 100.
- WikiProjects are online groups who organize unofficially around either subject areas or types of task (e.g., copyediting). They can exist anywhere, and go by different names on different wikis (e.g., "Portals" or just "Projects"), but are concentrated on en.wiki, where about 300 are currently active.
- Campaigns are unaffiliated working groups, usually in part supported by organizers at existing affiliates, which run widespread activities, such as Art+Feminism, Wikipedia Asian Month, or Wiki Loves Monuments
- Independent Organizations a number of individual organizations exist on the edge of recognized groups within the movement.

Typical Activities

- Content creation: E.g., content drives, editathons, photo expeditions, writing competitions
- Knowledge dissemination: E.g., training events, conferences, communications campaigns
- Process improvement: E.g., standardization of sources, procedures, style guides, templates
- Lobbying, partnerships, outreach (esp. GLAMs, governments, NGOs, etc): The province of the chapters and some affiliates, who work to improve laws, negotiate content donations, conduct professional outreach and training, etc.

Who is NOT an organizer for our purposes?

Wiki "functionaries" such as stewards, admins, and bureaucrats, whose activities generally don't encompass "attracting and directing volunteer attention," to use our earlier definition.

Problems and Needs

Tool Discovery, Documentation, Ease of Use

I've listed this meta-problem first because it applies to every section below. Tools exist to at least partially address virtually all problems this document describes, but their existence is often unknown to organizers. They also often require technical knowledge to set up and operate—yet are poorly documented. As one organizer put it, "There are a lot of

tools we don't know about or know what they can do for us. We need someone to help us understand what we are missing, and what to do and how to do it."

- Lack of a standard organizer workflow: There is no step-by-step process that organizers can follow to create a new project or campaign and make sure that it will be successful. Efforts to document existing processes have been made, but guides were created by volunteers or just for one wiki and are not easy to find or kept up to date. As one staffer put it, "We need to give people the scaffolding of how to be successful."

- Lack of end-user documentation: A lack of end-user (as opposed to technical) documentation often makes the tools organizers do locate unusable. As one staffer put it, "We need someone's time to document this stuff for mortals."

- Need for technical skills: Most organizers do not have technical backgrounds. So the bots, scripts, Wikidata tools and other technologies that benefit some groups enormously are not available to all. This problem can be particularly acute in smaller communities, where it is less easy to find people with the required skills.

Community-Building Tools

"Community-building" refers to a nexus of functions that organizers require to inform, engage and motivate their communities—to build, as someone said, "a movement not an event." Performing these functions currently requires a patchwork of tools, a high level of technical sophistication and much manual effort. Key community-building functions include:

- Group conversations: Groups need better ways to share and discuss information among themselves. Organizers need easy ways to make announcements and invite large numbers of people to participate in a discussion. Group members need to be able to subscribe to to discussions at various levels of both granularity and volume: they need to subscribe to individual topics, instead of just overall pages; they also want more control over just how they follow topics—getting notified, for example, about every development in select areas but only about entirely new topics in areas of less interest. Safety is also a concern, particularly for groups that organize around sensitive topics, such as women's issues. The ability to declare some discussions invitation-only would be one way to manage these concerns.

- Group events: Organizers want an easy way to announce an event and find out who is going to come.

- Task management: Setting out work for a community to accomplish is a core organizer role, yet the wikis lack even the most basic functions of task-management software. There is no way to subscribe to a task, in order to follow its progress; to classify a task, so as to provide meaningful ways for users to choose one that suits them (e.g, no way to classify by level of effort); no

Tools

dedicated discussion thread attached to a task; no way to claim a task, in order to avoid conflicts. Moreover tasks that exist outside the wiki are not easily tracked (e.g., getting decentralized support on a

communications plan, solicit support for event organization, or other "jobs" that don't require on-wiki activities).

- Relationship management: Who in the Foundation has successfully worked with an important partner organization? With whom did they work? What is our contact at that partner's area of expertise? The answers to these and a host of other questions represent valuable business information. Yet we have no organized way of preserving or sharing this data.

Outreach and Promotion Tools

The Community-Building tools above speak to a need for more effective communication within a group. But organizers also need to reach editors, readers and others who are not yet in their circle of contacts.

Event and group promotion: There is currently no simple or effective way for organizers to promote their groups or events to wiki readers or editors who are likely to be interested (e.g., based on geography + demonstrated subject interest). This clearly hampers these groups' ability to grow. A founder of one of the largest and most active user groups recently named their inability to reach out beyond a circle of existing, very active users as one of his biggest issues with current tools. Tools like CentralNotice and its various subspecies exist, but lack important features. Geonotice, for example, can target by location but reaches only registered users and only on their Watchlists. CentralNotice and Sitenotice reach readers, but have no targeting features. Such tools are also subject to many restrictions and layers of approval.

- Recruiting new members and matching needs with skills: Finding new members is slow and often accomplished through personal networks. As one staffer and organizer put it, "You have to find out how to find the people. Then find the people. Then invite the people one by one—it's a ton of work." Ways are needed to connect groups with volunteers and potential volunteers who share their passions. Ways are also needed to match groups with volunteers who possess needed skills, such as bot writing or conference planning (Connect is one model of how this can be done).

- Beyond email and talk pages: Email is the only message-delivery mechanism outside talk pages that our system currently supports. But the world we live in now requires that organizers broadcast and stay in touch with members on multiple social-media platforms at the same time. Feeding these multiple platforms manually is labor intensive.

- Challenges in this area: Meeting organizers' promotional needs may require us to reconsider some longstanding ideas and prohibitions. We may, for example, wish to experiment with limited, noncommercial, movement-focused advertising to wiki readers. Given the wikis' enormous

Tools

traffic, even narrowly focused banners might prove effective. But stepping up activity in this area would require at the same time a new type of (possibly technical) oversight, to make sure the level of promotion is not excessive, and a streamlining of existing community processes. There may also be ways for us to respect editors' privacy while still targeting them with messages based on geography or demonstrated subject

interest (or editors might be requested to voluntarily submit such information).

A better method for tapping subject interest

Research and experience tell us that an abstract interest in the movement per se is not what motivates volunteers at the beginning of their wiki journey. They come to us initially to share their knowledge and passion for some subject. A glance at the Wikiproject directory demonstrates the wide range of interests organizers and volunteers pursue formally, from folklore to pharmacology, football to firearms. Such projects are seriously hampered, however, by the fundamental weaknesses of the wiki category system [1], our primary means for classifying content by subject.

A world of applications: We need to be able to more effectively exploit metadata about article topics and quality and about wiki tasks. The possible applications for organizers of such a system are almost limitless. Broadly speaking, organizers need automated ways to classify (and therefore assemble and search for) articles and tasks by subject, and to reach out to potential participants based on demonstrated subject interest.

Event-Management Tools

Organizers of ediathons, training sessions, photo walks and similar in-person events have needs related to event management. Event organizers' existing workflows are rife with manual processes, workarounds and duplicated effort (partially documented here). The Event Metrics project currently in development will aid event organizers with better data about their contributions, but event management is out of scope for that project. An opportunity exists for synergy between Event Metrics and future event-management tools, since both make use of similar input data about the event and it's participants.

- Participant signup and sign-in: No system exists for acquiring and storing advance event-registration data or for checking-in users on the day of the event. Privacy issues will complicate any solution (because participants must supply email); not solving the issue, however, forces organizers to employ third-party tools (e.g., EventBrite), subjecting participants to commercial privacy practices.
- Wiki account-creation, day of: This is an urgent problem that should be addressed. For security reasons, the wikis allow only a limited number of accounts to be created from one IP during a given timeframe. This creates significant issues for event organizers on the day of, since it's common for participants to show up with no wiki account. There are workarounds, but they are not universal or well known. And even when experienced organizers follow all the best practices, participants still get blocked. (Potential solutions are discussed in this ticket.)
- Conference Tools: According to Program staff, we rebuild the infrastructure every year for conference calendaring, signup, payment, scheduling etc. More directly linked to on-wiki activities may be the tasks related to sessions, including proposal submission, scoring of submissions, scheduling, and presenting scheduled programs on wiki.

Tools

Recommendations

The broad survey of organizer needs above will be useful, I hope, for product teams,

who are largely unfamiliar with organizers' activities, having done little work on this area up to now. But we obviously can't move on all the fronts described. The Movement Organizer Study, soon to commence, will give us much more data on which to base product decisions. In the meanwhile, however, as a starting point, I offer the list of recommended priorities below, based on my limited investigation.

These problems are those whose solutions would, it seems to me, most deeply impact organizers' effectiveness. As befits goals for a 3- to 5-year horizon, the fixes to most of these problems will not be trivial. But, for the most part, neither will the benefits of those fixes be limited to organizers. "Subject interest" and "Group conversations", in particular, are fundamental technologies whose potential applications are widespread.

- Group conversations: Talk pages may not be the only way to address organizers' communication needs. But "fixing talk pages" is on our organizational agenda already. In the coming months, the annual-plan mandated consultation about "fixing talk pages" will initiate a conversation about the shortcomings of this core wiki communication platform. If we make an effort during that process to understand the particular needs of organizers, I'm confident solutions can be incorporated into our plans that will help organizers keep their groups engaged and informed. An important area for discussion will be whether we should incorporate social-media channels into our notifications system.

- Event and group promotion: Better outreach and promotion tools would enable organizers to more effectively reach desired audiences in order to spread the movement. Overcoming the problems associated with using the wikis for mission-focused promotion will require efforts in both the technical and social arenas. But, given the enormous traffic the wikis command, failing to tap the communication potential of our platforms would be an enormous opportunity missed.

- A better method for tapping subject interest: Subject interest is a key motivator of wiki activities. Providing simpler and better ways for organizers to automatically assemble tasks and perform outreach based on subject will pay dividends in increased efficiency

Tools

and effectiveness across a wide range of activities. By most accounts, revising and reforming the category system itself is unlikely to be the fix to this problem. Future solutions may come from experiments currently ongoing with structured data, AI projects like ORES Draft topic model, or from a system based on link analysis, like the one that powers Recommendation API.

- Tool Discovery, Documentation, Ease of Use: Our tools are hard to use and install, yet they are poorly documented—a bad combination. They are also hard to discover. The Technology team already employs one staffer whose job is documentation. Maybe it's time to consider more such positions, with staffers acting to mobilize, organize and standardize community writing about best practices and model workflows. As to discoverability, various efforts exist on this front already. The Toolhub project is an attempt to survey these past events and create a new, model tool-discovery platform. Among other improvements, it adds much more data about key

questions users have, like whether the tool is broken and what wikis a tool works on.

- Wiki account creation, day-of: This is something we should address in the near term. This ticket includes a good discussion of solutions. The favored fix involves a change to the Event Coordinator right that would probably require community approval. In addition, that right is not available on all wikis, so some thought should go into how to make it more universal.

Tools 08

Tools 08

Notes

[1] A few of these weaknesses: Categories are monolingual, ad-hoc and completely nonstandard, so developing universal tools based on them is challenging. Items in sub-categories don't inherit from parents, which is why searching broad categories typically yields few article results, contrary to user expectations. This non-inheritance also leads to hyper-specificity and bloat, to the point of absurdity. Technically speaking, the category system is not a "taxonomy"—a logical structure where all child categories are more specific classifications of the parents and wholly contained by them (e.g., Musical Instruments > String Instruments). It is, instead, a "category network," where children have some relationship with parents, but the relationship is unpredictable and therefore less useful (Musical Instruments > Orchestras). Categories can even be circular, with one category being both a parent and a child of itself.

Sources

Joe Matazzoni : Research and Insights Other contributors[1]: A. Bartov, T. Bolliger, M. Cruz, J. Hare, D. Horn, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pargarkar, J. Seddon, A. Stinson, N. Wilson, L. Zia

Research about editathons and other in-person events

- Eight organizer interviews: As part of the research into the ongoing Event Metrics project, I conducted about 12 hours of interviews with seven event organizers suggested by program staff. (I also received one written interview.) These interviews, for which I have extensive notes, cover the full workflow of event creation, management and reporting. I did not ask subjects for the right to publish but could request if desirable.
- Extensive talk page discussion: Also as part of Event Metrics, I've engaged extensively with organizer on the project talk page. (The discussion is organized by subject. E.g., here on the problems associated with Account Creation.)

Research about movement organizers generally

- Movement Organizers: Initial findings from Wikimania Interviews, is a slide deck from Alex Stinson and Abbey Ripstra (checking to see if I'm allowed to link to this yet).
- Staff interviews: with program and other team members: Alex Stinson, Maria Cruz., Asaf Bartov, James Hare

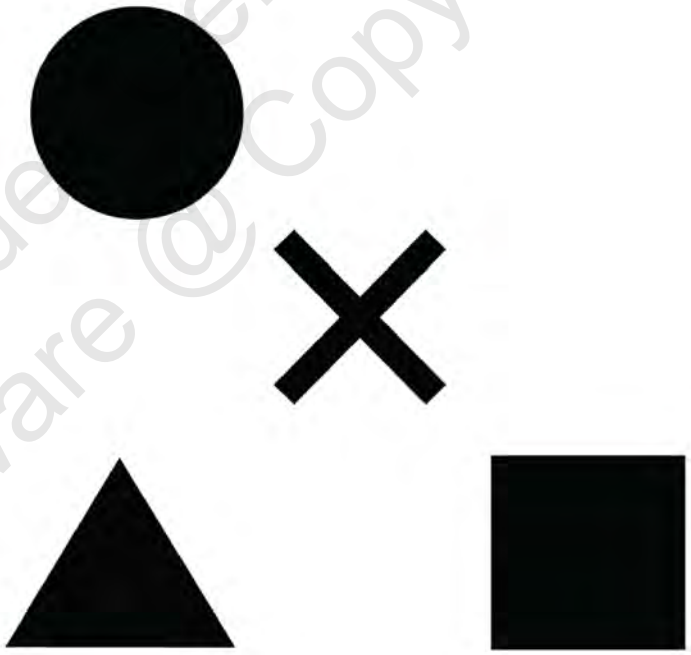
Exploring the Relationship between Reading Habits and Aesthetic Preferences in Different Cultural Contexts and Design Practices

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

Audiences x 3 / 5 Year Planning **2018**

culture

The following are draft position papers exploring aspects of movement culture based on insights from Community and WMF experts, as well as synthessios of secondary research.



01
Inclusion

02
Language

03
Content Gaps

Culture Overview 00

In many ways, the theme of Culture synthesizes the movement's strategic directions of Service and Equity. Knowledge as a service means enlisting allies and partners to make knowledge available outside the confines of Wikimedia - and that must include institutions and form factors in which the majority of the world's knowledge still resides, un-digitized and unavailable to a public who might not even know it exists. This means we have to think outside the existing Wiki projects and begin acting as a platform for knowledge organization, dissemination and exchange, and as a catalyst for other organizations and institutions engaged in knowledge dissemination. This will require new cultural as well as technical competencies.

Knowledge equity means removing the barriers preventing anyone from sharing what they know, and encouraging "respectful collaboration" between people. However, it is not only historical "structures of power and privilege" that have prevented people from sharing knowledge via Wikimedia. Focusing on specific forms of diversity that come from mainly an American context is likely to hinder rather than help the movement's global inclusiveness. Rather, we need to meet people where they are - in both the literal and figurative senses. A coherent culture strategy must start with awareness, for both consumers and contributors. Research has indicated that awareness of Wikipedia is low in emerging markets,[1]and that the fact that it is editable might actually work against some of our other priorities in the short-term, such as engendering trust.[2] We need, therefore, to find an approach in underserved communities that works - not assuming that people in these communities share our values or that the movement, as it is, can smoothly be applied to new cultural contexts. It is not certain that new users from emerging markets will necessarily share in the movement's goals right away, if ever. If we really mean all the world's knowledge, that means stepping outside our comfort zone.

In addition to fulfilling our mission of making all the world's knowledge available to everyone on our own projects, we also have the opportunity to make a lasting positive impact on the world's free culture by pursuing (or enabling others to pursue) other societal goals, such as the preservation of endangered languages.[3] Where we can identify areas our movement goals overlap with other free culture efforts, we should partner with the people and institutions pursuing them.

Culture Overview 00

Examples

Content Translation

GLAM outreach and strategy

Structured Data on Commons

Areas of Impact

Most Wiki projects (but especially Wikipedia, Commons, Wiktionary, and Wikisource)

Community Relations

Communications

Mobile (both apps and mobile web)

Community policies and guidelines

Research

Disabled or Disenfranchised Communities

Key External Factors

Between 53%[4] and 71%[5] of the world's population will be online by 2030, with the growth mainly driven by emerging markets. The overall growth rate is slowing[6] due to saturation in developed markets.

References 00

1. Movement Strategy Executive Summary of awareness
2. Finding from New Readers research in Nigeria, India and Mexico "Trust in Wikipedia is shaken when people find out anyone can edit pages."https://meta.wikimedia.org/wiki/New_Readers/Findings/India#Theme:_Using_Wikipedia
3. https://wikitongues.org/
4. https://blog.euromonitor.com/2015/04/half-the-worlds-population-will-be-online-by-2030.html
5. https://www.gsma.com/mobileeconomy/wp-content/uploads/2018/02/The-Mobile-Economy-Global-2018.pdf
6. USA Today: Smartphone Sales Have Hit a Wall

Sources

UNESCO: Fostering inclusive knowledge societies

UNESCO linguistic diversity on the internet project

Geographies of the world's knowledge

Indonesia and Brazil "new voices" research

Increasing brand awareness

Patterns: tremendous potential for growth in underserved regions

GapFinder

Knowledge Gaps @ Wiki Research

The state of the article expansion recommendation system (Research showcase)

New Voices synthesis: attracting new users

Linguistic Genocide in Education by Tove Skutnabb-Kangas

Culture 01

inclusion

What do “diversity” and “inclusion” mean in the context of product?
Diversity in this context is in reference to having a variety of social stratifications within a collective, namely class, race, sexual orientation, age, disability and gender. An important aspect of diversity to consider in the context of the movement which traverse across these various groups are differences of ideological viewpoints which informs the types of knowledge considered for inclusion.

Inclusion is recognizing the need for greater diversity across the following groups in Wikimedia:

- Contributors of content
- Beneficiaries of content (i.e., Readers and Consumers of content)
- Movement Organisers and Representatives (i.e. at an Institutional level, in policy-making & decision-making)

Ultimately, the drive for inclusion of different contributors, beneficiaries and movement leaders is to ensure diversity of knowledge. That is, broadening perspectives of the various forms of content that we consider as valid and valuable part of the knowledge space.

Sections

Culture 02

Why diversity matters

Besides being one of the core values of the Wikimedia Foundation, [1] diversity helps us to reach towards the goal of being able to share "all" knowledge, as well as ensuring greater quality of the content that is shared. Direct examples of diversity directly improving knowledge is seen in two recent studies from 2017, the first study saw that Political, Social Issues, and Science articles whose editors were comprised of more diverse political viewpoints (across Democratic & Conservative) were of higher quality than those with politically homogeneous editor groups.[2] Similarly, a separate second study noted that language-specific topics on Wikipedia are generally better quality on relevant language-editions of Wikipedia, and could be leveraged to improve the knowledge base across language editions. [3]

What's holding us back?

Conceptual barriers/issues

- Ambiguity and Subjectivity of Inclusion and Diversity – There is no end game but only an ongoing pursuit of greater inclusion and diversity. Resource constraints and biases as identified above means that there is continuing work to help recognize marginalized groups. But it is hard to reach consensus as to which groups are of higher or lower importance/need and what targets should be set (for example, the number of articles of a particular language).
- Individual Biases (Conscious and Unconscious/Implicit) – a common example is the notion of notability of BLP articles as being assessed based on criteria that often leads to underrepresentation of certain groups.
- Systemic bias – processes and organizational structures set up in a certain social-cultural groups leads to decisions that under-represent other groups
- Lack of Awareness – that certain communities exist (I don't know who I don't know), and vice versa (Community X doesn't know how or why WMF is of value to them).
- Lack of understanding and attitudinal differences between groups – Group X feels unwelcome, excluded or harassed by Group Y.

Internal organizational/resource constraints

- Gaps in Accessibility (at a high level, web content is considered Accessible when it is Perceivable, Operable, Usable, and Robust) [4]
- Providing resources so content that meets Web Accessibility Guidelines for content users
- Providing tooling for Contributors who are have Accessibility needs
- Gaps in Content – Languages, Topic areas, Media, etc
- Human Capital – people are also a limited resource. There are trade-offs to be made to promote greater diversity and inclusivity without diminishing/diluting overall impact.

External factors

- Access to resources: Internet and technology is inequitable depending on geography, socioeconomic background, etc. This is consistently reported in

Culture 03

multiple studies, including a recent report from the UN noting that there is still less than 50% of the world's population who are online as of 2018, and moreover there is a trending decline in internet growth.[5]

- Differing interests/traditions: Certain communities may not want to participate, share, or be "served" for their own individual reasons and interests. Some Wikimania 2018 sessions touch upon some of these reasons for disengagement from communities - from differences in communication norms (e.g., communities which have a stronger oral tradition[6][7]), distrust due to a history of being marginalized,[8] to fundamental differences in cultural conceptions of what is an appropriate channel for knowledge-sharing.[9]

Reducing Barriers to Inclusion

How can we identify and correct for biases that exist within the Foundation or community? OR: Ideas to reduce barriers to inclusion.

Prioritize gaps in representation at higher leadership and decision-making levels within the Movement.

It's important to have more diversity at the leadership level as it not only helps reduce systemic and unconscious bias, but also encourages greater participation from top down.

Investment in more research and socializing commonly agreed upon definitions.

We need to establish some commonly agreed upon definitions or baseline for measurement. For example, assessing our success in making knowledge accessible to every person is dependent on what we define as knowledge, diverse, and what it means to be accessible. [10]

Broadening measurement and research means both including currently identified underrepresented voices [11] and understanding what ways they feel excluded from participation, but also continuing to research where there are still existing unidentified gaps in representation.

Open and influence positive channels of communication across communities

Besides representation, one of the ways to reduce the conceptual biases of Communities is for the Wikimedia Foundation to facilitate better lines of communication in general across groups. Part of this involves continuing research to show benefits of diversity in advancing knowledge quality as well as equity (see above "Why diversity matters" section).

A second part is to continue investing in tools and programs to that encourage civil discussions and fostering more positive relationships in our communities. This relate to our work in anti-harassment tools, as well as initiatives that strive to provide better help and support for new members (e.g, New Readers program, and the Growth Team).

Cultivate effective Partnerships

Recognizing that Wikimedia has limited resources to provide and some factors are outside of our control, we should explore

Culture 04

strategic partnerships that optimize our reach.

Technology and tools

As detailed in the next section, we can utilize technology to both attract more voices as well as helping to fill content gaps.

How do we attract and retain contributors from those marginalised areas? OR: Ideas to promote participation from marginalized groups

Representation with welcoming and diverse voices

Adding representation in Movement is not only important in reducing barriers to inclusion but also as a way to be seen as welcoming to newcomers.

Establishing more strategic partnerships

- Embedded community groups - Cultivate partnerships with local experts and embedded members of a particular community who have more access to content, people and support.
- GLAM institutions – help external organizations in preserving collections without taking on the burden of maintenance.
- Content distribution platforms - working to be a source of knowledge with content distribution platforms where more and more users are seeking content. This includes Search platforms (eg., Google), News media organizations (e.g., NYTimes), and Social Media sites (e.g., Twitter, Facebook)
- Education groups - working with MOOCs and other institutions in adapting Wikimedia content to newer learning courses tailored to helping improve digital literacy in marginalized communities

Invest in new technology and tools that help attract new voices for CONSUMPTION

Content tailored to newer methods and behaviors of consumption, sharing and learning will enable more people in low reach areas to become aware of and start accessing our knowledge. As noted in "Brand awareness, attitudes, and usage"[12], awareness and familiarity with Wikipedia is a main area of concern. Some areas where we can invest to increase awareness of the overall Wikipedia brand that can broaden reach in low awareness groups (eg. those with limited internet access, Gen-Z users) include:

- Better mobile content design (improved readability, more interactive/visual content and overall design improvements can increase its shareability) [13]
- Optimized mobile content (delivering content more accessible to newer users in areas where data, internet and technology access are restricted)
- Better sharing of Wikimedia content on Social Media (for example, bringing content to Facebook in Cambodia, where 30% of users access their information) [14]
- Better sharing of content on other online communication platforms (for example, in countries where messaging apps like Whatsapp, WeChat, Telegram etc are vital ways to share information)

- Enabling news media partners to reference and cite content
- Enabling machine-translation and creation of articles in languages with a smaller content pool

Invest in new technology and tools that help attract new voices for CONTRIBUTION

A second part of awareness is know that Wikipedia as a participatory tool. Another recent paper "The Pipeline of Online Participation Inequalities: The Case of Wikipedia Editing" highlights participation is limited earlier in the 'pipeline' of possible editors based on income and racial biases, then by technical knowledge, and finally a gender gap in awareness of its participatory nature. [15]

With that in mind, some contribution tools and aids that can help attract newer voices include:

- Translation tools to help multilingual users
- Tools that recommend articles for creation/expansion for new editors
- Mobile contribution tools to help groups in places with limited access to internet data & storage
- Ensure tools conform to Accessibility guidelines (so, for example, vision impaired users are not excluded from reading and contributing)
- "Micro-contributions [16][17] that reduce the technical expertise needed to start editing (as it has been identified that many underrepresented groups struggle to participate due to such barriers.) [18]

Programmatic Initiatives

Per above, another way we can continue to increase contribution is to specifically targeting marginalized or underrepresented communities with Outreach efforts such as in-person, off-wiki awareness events to recruit newcomers, and Edit-a-thons to expand content in identified underserved topic areas.

Culture 06

Notes

[1] https://meta.wikimedia.org/wiki/Values/2008#Wikimedia_Foundation_values
[2] h. ps://arxiv.org/abs/1712.06414
[3] https://www.mdpi.com/2227-9709/4/4/43
[4] WCAG "POUR" Accessibility principles - https://developers.google.com/web/fundamentals/accessibility/
[5] Sample, I. (2018-10-18). "Exclusive: drama□c slowdown in global growth of internet access". the Guardian. Retrieved 2018-10-21.
[6] Wikimania 2018 presenta□on: "Wikipedia and Bhutan can learn from each other"
[7] Wikimania 2018: presenta□on: "The quota□on of oral sources in a decoloniza□on context"
[8] Wikimania 2018 Panel: "Centering Knowledge from the Margins: A Whose Knowledge? discussion"
[9] Wikimania 2018 presenta□on: "Wikipedia for Indigenous Communi□es"
[10] UNESCO's Fostering inclusive knowledge socie□es report par□cularly focuses on the importance of defini□ons and measurements for what it means to be inclusive, have access to knowledge, etc; and stresses the need for establishing common defini□ons so their members states can aim for common goals.
[11] See Brand awareness, at udes, and usage - Execu□ve Summary
[12] See Brand awareness, at udes, and usage - Execu□ve Summary
[13] "In 2016 Internet/Facebook became the most important channel through which Cambodians access informa□on (30%) — surpassing TV (29%) and almost doubling radio (15%)" – from "Mobile Phones and Internet Use in Cambodia 2016" http://www.open.org.kh/research/phones_2016.pdf
[14] Already iden□fied underrepresented groups are those discussed in the references, including but not limited to: groups in other Western-European regions (per "Geographies of the world's knowledge"), Women and Non-binary (per "Gender equity report 2018", and those in areas with limited internet access (per "UNESCO's Fostering inclusive knowledge socie□es report").
[15] See Brand awareness, at udes, and usage - Execu□ve Summary
[16] See Brand awareness, at udes, and usage - Execu□ve Summary
[17] "In 2016 Internet/Facebook became the most important channel through which Cambodians access informa□on (30%) — surpassing TV (29%) and almost doubling radio (15%)" – from "Mobile Phones and Internet Use in Cambodia 2016" http://www.open.org.kh/research/phones_2016.pdf
[18] Hargittai, E., & Shaw, A. (2018). "The Pipeline of Online Par□cipa□on Inequali□es: The Case of Wikipedia Edi□ng". *Journal of Communication*, Vol. 68, Iss. 1, 1-Feb-2018, pp143–168, doi.org/10.1093/joc/jqx003
[19] https://www.mediawiki.org/wiki/Mobile_design/Micro_contribu□ons
[20] https://www.mediawiki.org/wiki/Reading/Readers_contribu□ons_via_Android
[21] Hargittai, E., & Shaw, A. (2015). "Mind the skills gap: The role of Internet know-how and gender in differen□ated contribu□ons to Wikipedia". *Information, Communication & Society*, 18(4), 424–442. doi: 10.1080/1369118X.2014.957711

Sources

Rita Ho Research and Insights

Culture 01

language

This document is an exclusionary act. Conceived and written in English by highly literate speakers of English, the world's most dominant and well-represented system of knowledge. But Wikimedia's vision aims to be *for all people,* in *their* language, calling for us to break down the exclusionary systems of knowledge creation and access. To do this, our systems, software and governance will have to evolve away from an English first, others maybe model, to a system that not only supports all languages but empowers them. For 60% of the world [1] using multiple languages, for different purposes, is a the normal way to live. Pidgins and inter-languages produce knowledge too, so how must the current one-project-one-language model adapt to support evolving languages such as Hinglish and Spanglish? And while capturing and growing knowledge in all languages is an uncontroversial goal, it is at these fuzzy edges that we can see, for example, the unintended exclusionary potential of technologies such as machine translation. As isolated cultural "space" collapses, a tension arises between support for cultural uniqueness and support for the experience of living between languages. And finally, language acquisition and cultural adaptation will be a huge factor in the foundation's success – what new capabilities does this demand of our organization and of the movement as a whole?

Sections

Culture 02

Privilege

[not edited] **This section is about the need to see and address our English first model.** How does culture (both Foundation/Wiki community culture and the cultures of the underserved language communities) play into what languages and/or topics are "marginalised"?

- Isolated cultural spaces are collapsing thus more languages are going to co-exist in the same space
- Diversity of cultures needs to be better handled.
- There will be predominance of some languages due to various factors like demographic majority, regional prevalence, social prestige, or academic needs
- We will need to find common ground that can act as the acceptable connecting points to diversify from.
- This will need more case based exploration. Language may not be the connection in all cases, but can be a catalyst.

Where to Focus?

[not edited] **How Do We Determine The Scope Of Influence Related To Languages That We Need To Consider?**

Language is a vast source of information and influence. Small and large impacts arising from languages are always happening.

Specialized tracking of all kinds of changes related to languages will be a heavy investment, without many clear benefits.

For the strategy plans, we need to isolate a few areas of impact that we expect to spill over into our work that we can build upon for technical and social intervention.

As language is as good as the living entities that define them, we need to track some of these defining factors and the changes that matter to them.

Priorities

[not edited]**How do language interventions get prioritized? What should drive this?**

What Are The Global Factors That May Influence Language Trends For The Strategy Planning Period?

Since Languages are influenced by factors affecting the people themselves, what are some of the sharp changes we are already seeing around us:

- Migrations: Forced migrations from conflict zones, Economic migrations for subsistence and specialized jobs
- Cultural preservation
  Increased attempts especially at places experiencing unusual amounts of sudden cultural intermingling
- Increased purchasing power in emerging markets around the world
  Improvement in lifestyle
- Better access to education
- More freedom in choice
- Lower barriers+cost to adoption of technology

Culture 03

- Cheaper devices and services
- Increased digital literacy
- Ease of use
- Need for increased connectivity

Tracking cultural shifts due to economic migration and other current global factors is important to understand the adaptive linguistic uses that can impact the technology we operate with.

Are Individual Languages Important In The High Level Strategy Process?

Yes. When we talk about languages in general, we also talk about the regions and people that we connect with them. By that extension, if we try tracking through factors like:

- emerging markets for increased technology adoption
- languages with need of additional content
- languages with need of special support: this could also include the cases for additional multilingual support

The actual set of languages may not be restrictive, even if special focus is put on them, as resolutions for one or more languages may extend to others, and can be used for quick gains.

Interoperability

Multilingual use patterns

[not edited] This section is about how the platform must support context switching between languages (according to context of use).

(How) Can we create content that is consumable in multiple languages, or doesn't require language at all to enjoy and use?

The model of language around which the projects are built, with parallel encyclopedias and dictionaries for each language, and some set of shared content that is available across them all, understates the complex way people use languages. For 60% of the world (http://ilanguages.org/bilingual.php) using multiple languages, for different purposes, is a the normal way to live.

Code switch, pidgins and inter-languages produce knowledge too. What does our model of 1 project:1 language adapt to a world of Hinglish and Spanglish?

Language and culture are intimately related, in multilingual societies languages come to serve specialized purposes, and some types of knowledge are associated with a particular language. For example it is not uncommon for there to be a language for science, a language for government and a day-to-day spoken language, in which mass media and popular culture are conducted. The idea that each language will have a complete vocabulary and source material to build an encyclopedia that looks like English's doesn't stand scrutiny.

New formats needed?

Culture 04

Support

[not edited] How do we provide not just content, but also support in marginalised or underserved languages? What happens when a contributor who only speaks e.g., Yoruba or Tibetan has a policy question? Participation demands support, especially on the long tail.

- Languages new to the internet, we need to continue pushing forward Unicode font support and basic text entry (reference-- Minh, Odia, etc), to ensure the most basic forms of access for all languages

Opportunities & Threats

[not edited] This section is about threats to culture posed by machine learning and language extinction. How can we avoid reinforcing language inequalities if we use technology (machine translation, for example) to fill our content gaps?

Content translation is crucial, but insufficient. Language and culture are deeply bound up with categorization, notability and content expectations. Simply translating all articles between all pairs of languages would not make a relevant or even comprehensible experience for most humans.

Machine translation is making the boundaries between languages softer. Wikidata and Commons doing the same. How do we preserve the cultural values and self-determination of each speaker community in a world where these communities and languages are becoming more porous and mixing more than ever?

Advocacy

Otherness

[not edited] This section is about how or if we take on the role of extracting and cataloguing knowledge from languages that don't have a writing system? Our intervention could be at the level of platform - a DIY kit for capture? The hawaiian Oli.[x]

"Since the beginning of time, every culture has developed means of passing on important information to its people. For Hawaiians, there was no written language per se until the 1820s. The missionaries introduced the alphabet which made it possible to represent Hawaiian language in the written form. Until then, all information was passed orally through the use of songs, chants, and poems.

Hawaiians devised various methods of recording information for the purpose of passing it on from one generation to the next. The oli was one such method. Elaborate chants were composed to record important information, e.g. births, deaths, triumphs, losses, good times and bad."

Preservation

This section is about the extent to which the WMF should advocate for or orchestrate the preservation of endangered languages. Is partnership the way to go here? Developing educational materials in languages that are spoken by

Culture 05

many people and aren't immediately endangered, but are rarely used for education and business

Societal Goals

How can the Wiki movement contribute to larger societal goals? Should it?
Facilitating language teaching and learning

Notes

[1] (http://ilanguages.org/bilingual.php)

[2] Hawaiian Oli
https://apps.ksbe.edu/olelo/learning-place/performance-indicators/chant/foundational/significance-oli-chants-in-hawaiian-society

Sources

R. Bhattacharjee and C. Gauthier Research and Insights, Other contributors[1]: A. Amaroni, D. Horn, R. Isler, J. Katz, J. Minor, T. Negrin, M. Novotny, N. Pangarkar

https://annual.wikimedia.org/2016/fact-1.html

Google-KPMG report titled 'Indian Languages - defining India's internet
https://assets.kpmg.com/content/dam/kpmg/in/pdf/2017/04/Indian-languages-Defining-Indias-Internet.pdf

The Rise and Rise of Hinglish in India
https://theconversation.com/the-rise-and-rise-of-hinglish-in-india-53476

- ➔ Article coverage by language
  - Research: Increasing article coverage
  - Growing Wikipedia across languages

[1] If your name was left off the list by mistake please contact JMinor or MNovotny

- Content translation recommendation API
- Expanding articles cross-language @ Meta

➔ New Voices synthesis: cultural issues hindering content creation

➔ Language Extinction:

- Wikitongues
- Foundation for Endangered Languages
- Rosetta Project
- Living Tongues Institute
- UNESCO linguistic diversity on the internet project
  - And an associated report
- BBC: Why we must save dying languages
- Book: Linguistic Genocide in Education - Tove Skutnabb-Kangas

Culture 01

content gaps

Are we there yet? Making the sum of the world's knowledge accessible to all the world's people presumes that the extent of knowledge can be described,[1] measured,[2] and tracked.[3] It also implies that content gaps can be identified and filled with context-relevant takes on the missing knowledge – this may be the most difficult part of the problem. For our purposes, gap knowledge can be said to fall into one of these general categories: potentially portable knowledge,[4] contextually nuanced knowledge[5] or contextually anchored knowledge.[6] The three types require different intervention strategies, and may have different relative priority for a given a regional community. For this reason, intervention strategies (product features, global tools, community mobilization) must be developed in parallel. Solutions for filling content gaps also surface new challenges around notability and verifiability[7] – how must current standards for authority need to evolve in order to encourage the growth of each type of gap knowledge? And finally, assuming that it is possible to close the content gap through various means, what support mechanisms will be required to maintain this body of knowledge? It will be necessary to have tools that monitor and signal the freshness of content to the community best suited to maintain it. In sum, making all knowledge available to all the world's people demands the organization and participation of all the world's peoples, and interventions deployed at pace with the rate of growth of the overall pool of world knowledge. [3]

Sections

Culture 02

Identifying Gaps

How Do We Identify Content Gaps And The Appropriate Ways To Fill Them?

- What is "the sum of all knowledge" - is that a thing? Does it exist?
    - Filled by everyone contributing rather than by circumscribing the pie?
    - How/whether to measure our progress towards the "Grand Total"?
- Content gaps may exist for a number of reasons that go beyond the well-known reasons of institutional bias or lack of internet access. Knowledge may be held by a group or groups of people who:
    - Lack awareness of Wikipedia's existence entirely, or lack awareness that it is editable
    - Are aware Wikipedia exists, but don't see the relevance to them
    - Have specific cultural, religious or other taboos about sharing specific knowledge (such as about religious rites)
    - Have ingrained cultural or personal beliefs that community contributions are somehow "less than" institution-created knowledge
- Filling in content gaps goes beyond identification of the gap and the creation of initial content. It also needs to include:
    - Finding an audience for the material - why would contributors want to create content that's not going to get read?
    - Finding contributors who want to maintain and expand the initial content - otherwise it's likely to get stale, and then we will lose that audience.
    - Prioritisation of the gaps - we cannot fill all gaps immediately, and there are some gaps we may never be able (or want) to fill. Prioritisation therefore is necessary based on (among others):
    - Potential audience size
        - Language
        - Internet penetration
        - Awareness
    - Potential contributor community size
- Notability of the topic in that particular language

Care should be taken with finding automated ways to fill the gaps, so that we do not inadvertently reinforce inequalities (see Language section, for example)

How Do We Bring Knowledge Out Of Legacy Media And Institutions And Onto The Web?

- There is a vast quantity of knowledge locked away in legacy media that has never been digitised. These heritage materials are not available online for reasons including that:
    - No one realises they exist
    - They are copyrighted

- The material is considered too "specialist" or "esoteric" to warrant digitisation
- It is difficult or impossible to successfully digitise with current technologies, or difficult/impossible to retrieve without specialist knowledge
- The knowledge might not be tangible or digitisable in its current form

- Rather than acquiring the rights and/or digitising the content ourselves, we should consider acting as a catalyst for other organisations or individuals who are doing this:
  - Libraries
  - Archives
  - Museums
  - Online projects (Gutenberg, Internet Archive, etc)
- The catalyzation could take many forms, including:
  - Acting as a technology partner
  - Contributing money or resources to the project
  - Consulting or contributing expertise
- It's also important to realise that legacy media is not the only place knowledge lives. Institutions (universities, governmental bodies, etc) also have huge amounts of knowledge that is not available on the web.
  - Partnering with academics to bring their research to the public (such as by adding citations of published material to articles)
  - Advocating for digitisation of government materials where governments are not already doing this
- Meta-knowledge: knowledge of how to get the knowledge

Evolving Notability Standards

(How) Do We Manage Types Of Knowledge That Don't Conform To Our Current Standards Of Notability Or Published Authority?

- There are types of knowledge (such as oral history) that by their very nature do not conform to Wikipedia's standards of notability or published authority
- If we wish to make Wiki the repository of ALL the world's knowledge, we need to consider how various alternative forms of knowledge can fit in with existing or new Wiki projects beyond Wikipedia.
- We might also consider acting as a catalyst for organisations that are gathering and digitising these forms of knowledge, rather than attempting to reinvent the wheel - especially when those organisations already have an existing body of work and useful contacts that would take us a long time to build up.

Culture 04

Incorporating Other Formats

(How) Do We Move Beyond Primarily Text-based Content And Into Other Formats Such As Video, Audio, Images, Etc?

- This assumes that users want this type of content, and that current and/or new contributors will wish to branch out into providing this sort of content.
- We also need to define the purpose of this type of content:
  - To add to existing Wikipedia articles?
  - To encourage social media sharing?
  - As a destination in and of itself?
- Learn from other sites (newspapers, etc) that have embraced video: Short-form video content is popular on social media, but rarely repays its costs
- New content types bring new considerations:
- Tooling changes
  - Not only for us but also for our contributors
- Process changes
  - How to review and approve this content?
  - How to maintain and update this content?

Supporting New Platforms

How Can We Make The Tools For Filling In Content Gaps Available On The Devices People Are Likely To Need Them On?

- Many of the most urgent content gaps are in languages and about topics relevant to places outside Europe and North America.
  - Mobile-friendly tools for content creation
  - Translation tools
  - Less data-intensive tools for rich media creation
- Internet penetration is growing in these countries, though the majority of users access the web on mobile devices.
- Addressing content gaps therefore means thinking outside the desktop editing paradigm
-

Culture 05

Notes

[1] T. Negrin: “The Static Knowledge Pie”: using the sum of all topics currently covered on all wikis as a measure of completeness, the content gap is measured as topics existing on some projects and absent on others. Requires an exhaustive but extensible topic model.

[2] T. Negrin: “The Potential Knowledge Pie”: The Static Knowledge Pie with the addition of known gaps across all wikis (e.g.estimated number of articles there should be about female scientists).

[3] T. Negrin: “The Dynamic Knowledge Pie”: the diameter of the pie increases over time making the content gap a dynamic function of the relationship between rate of knowledge created in the world and knowledge captured on the system.

[4] Potentially Portability reflects the degree to which certain classes of knowledge have the potential to be “ported” or translated with low loss in fidelity. Scientific terminology and topics are an example of PPK. Certain types of legacy media (e.g. illuminated manuscripts) that can be ported from one medium to another could also be considered PPK. Transferring Potentially Portable topics and formats is challenging, but methods for tackling them are known unknowns.

[5] Contextually Nuanced knowledge is a category of knowledge that requires a higher degree of human input to transfer (e.g. subject matter expertise, topical knowledge and interpretation). The history of the Korean War written from a Japanese, or American or Korean point of view would be an example of CNK.

[6] Contextually Anchored knowledge is a category of knowledge that is utilized by a specific community in a specific way. This knowledge can be *described* in other contexts, but not utilized in a comparable way. The Hawaiian Oli chant is an example of CAK.

[7] Decline of editor retention in mature language communities on Wiki being (partially) caused by lack of welcome for newcomers

Sources

Charlotte Gauthier Research and Insights

K. Sweet and K. Wirth : One-to-One Personalization in the Age of Machine Learning

H. Lee and E. Chang : Consumer Attitudes Toward Online Mass Customization: An Application of Extended Technology Acceptance Model

S. S. Sundar and S. S. Marathe Personalization versus Customization: the Importance of Agency, Privacy, and Power Usage

Exploring the Relationship between Reading Habits and Aesthetic Preferences in Different Cultural Contexts and Design Practices

Index

P

R

S

T