Wikipedia as a socio-technical system

by Aaron Halfaker
About me

Hi, I'm Aaron Halfaker. I'm a computer scientist. I've been a Wikipedian since 2008. See my publications. In 2011, I started working with the Wikimedia Foundation as a researcher. See my volunteer account, User:EpochFail, for my non-staff work.

My work

I use the scientific method to explore the structure and function of Wikipedia. My goal is to identify problems/opportunities and to use this knowledge to improve/extend the software. See The Rise and Decline for an example of my research and Snuggle for an example of a tool for Wikipedians that I'm actively developing.

Contact me

- E-mail: ahalfaker@wikimedia.org
- Website: http://halfaker.info
- Twitter: http://twitter.com/halfak
- On IRC, I go by "halfak"
Aaron Halfaker
Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

computer scientist. I've been a Wikipedian since 2008. See my publications. In 2011, I started working with the Wikimedia Foundation as a researcher. See my volunteer account, User:EpochFail, for my non-staff work.

the structure and function of Wikipedia. My goal is to identify problems/opportunities and to use this knowledge to improve/extend the software. See The Rise and Decline for an example of my research and Snuggle for an example of a tool for Wikipedians that I'm actively developing.

@wikimedia.org

- Website: http://halfaker.info
- Twitter: http://twitter.com/halfak
- On IRC, I go by "halfak".
Aaron Halfaker
Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

computer scientist. I've been a Wikipedian since 2008. See my publications. In 2011, I started working with the Wikimedia Foundation as a researcher. See my volunteer account, User:EpochFail, for my non-staff work.

the structure and function of Wikipedia. My goal is to identify problems/opportunities and to use this knowledge to improve/extend the software. See The Rise and Decline for an example of my research and Snuggle for an example of a tool for Wikipedians that I'm actively developing.

@wikimedia.org
- Website: http://halfaker.info
- Twitter: http://twitter.com/halfak
- On IRC, I go by "halfak".

Usually...
Aaron Halfaker
Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

I'm a computer scientist. I've been a Wikipedian since 2008. See my publications. In 2011, I started working with the Wikimedia Foundation as a researcher. See my volunteer account, User:EpochFail, for my non-staff work.

The structure and function of Wikipedia. My goal is to identify problems/opportunities and to use this knowledge to improve/extend the software. See The Rise and Decline for an example of my research and Snuggle for an example of a tool for Wikipedians that I'm actively developing.

@wikimedia.org
- Website: http://halfaker.info
- Twitter: http://twitter.com/halfak
- On IRC, I go by "halfak".
DISCLAIMER

● These are my opinions.
● While they are well informed, they do not necessary reflect that of the Wikimedia Foundation or anyone else.
● I am not a lawyer, a prophet or Jimbo.

● No one asked me to put this disclaimer here.
Outline

1. The “-” in socio-technical

2. Open problems in open collaboration
   (for collaboration)

3. What’s next?
Social
Technical
Socio-technical
Technologist
Techno-biologist?
Bacterium

Paramecium
socio - technical
System with specialized sub-systems ~ System with specialized sub-systems
Part 2

Wikipedia’s specialized sub-systems
Work allocation

- Identify, prioritize and assign tasks
Work allocation

- Largely for free due to Linus’ Law
Linus’s Law?
Linus’s Law?

Eric Raymond
Linus’s Law?

“given enough eyeballs, all bugs are shallow”
Linus’s Law?

“given enough eyeballs, all bugs are shallow”

yessss…

bugssss….
Linus’s Law?

“given enough eyeballs, all bugs are shallow”
Linus’s Law?

“given enough eyeballs, all bugs are shallow”
Linus’s Law?

visibility is critical to open collaboration
A corollary

Given that enough people see an incomplete article, all potential contributions to that article will be easy for someone.


User:SuggestBot

From Wikipedia, the free encyclopedia

This user account is a bot operated by Nettrom (talk). It is a legitimate alternative account, used to make repetitive automated or semi-automated edits that would be extremely tedious to do manually. The bot is approved and currently active – the relevant request for approval can be seen here.

SuggestBot is a program that attempts to help Wikipedia users find pages to edit based on their past contributions. If you want to try it out and get some personalized recommendations we have documentation explaining how to do that. You can also sign up to receive suggestions regularly. If you're wondering what the bot's recommendations looks like, you can find examples here.

SuggestBot uses a variety of algorithms, including standard information retrieval and collaborative filtering techniques, to make suggestions. Most of it runs at the GroupLens Research Lab, some parts also run on Wikimedia Tool Labs and the Wikimedia Toolserver. The recommendations are usually done based on downloading a list of a user's contributions (to avoid recommending articles they've recently edited). The bot was originally developed by ForteTuba.

If SuggestBot made personalized recommendations for you, please leave feedback on whether they were useful and how to make them better. Comments on recommendations, as well as general comments, suggestions, or complaints, are best left on our talk page. Comments are welcome and valuable, as they will help us make SuggestBot do a better job of helping Wikipedia.
SuggestBot is a program that attempts to help you to identify areas of Wikipedia that could use improvement. It actively looks for pages that are referenced by articles but are not linked from them, and pages that are linked from articles but not referenced by them. It also looks for pages in the categories "People" and "Music" that are not linked from other pages. You can sign up to receive suggestions as they are made.

SuggestBot uses a variety of algorithms, including: make suggestions. Most of it runs at the GroupLens. The recommendations are wikipedia toolserver. The recommendations are based on a user's recent editing history. If SuggestBot made personalized recommendations, how to make them better. Comments on recommendations are best left on the talk page. Comments are welcome.
Work allocation
- Identify, prioritize and assign tasks

Regulation of behavior
- Norm formation, propagation and enforcement
A *norm* is a group-held belief about how members should behave in a given context. [citation needed]
Coordination and beyond: social functions of groups in open content production
A Forte, N Kittur, V Larco, H Zhu, A Bruckman, ... - Proceedings of the ..., 2012 - dl.acm.org
... In our first round of interviews, we sought to understand governance in Wikipedia as a site-wide phenomenon. As such, we interviewed eleven individuals who had been involved in the site for many years, in many different capacities ...
Cited by 13 Related articles All 5 versions Cite Save

Making restitution work: the challenge of building sustainable governance and institutional structures in public administration
L Schoeman, DJ Fourie - Journal of Public Administration; ..., 2008 - reference.sabinet.co.za
Cited by 3 Related articles All 5 versions Cite Save

Changing governance of local economies: Responses of European local production systems
C Crouch, P Le Galès, C Trigilia, H Voelzkow - OUP Catalogue, 2004 - ideas.repec.org
... specialized manufacturing among small and medium-sized enterprises (SMEs) in France, Germany, Italy, and the United Kingdom, the authors find patterns of economic governance far more ... This item is not listed on Wikipedia, on a reading list or among the top items on IDEAS. ...
Cited by 157 Related articles All 5 versions Cite Save More

Foundations of service science management and business
H Katzan - Journal of Service Science (JSS), 2011 - journals.cluteonline.com
... The development of a service context involves the asking of tough questions to examine the strategic goal and objectives of a service organization in order 9 Adapted from the Wikipedia articles on strategy and corporate governance. Page 9. ...
Cited by 6 Related articles All 5 versions Cite Save

Building semantic kernels for text classification using wikipedia
P Wang, C Domeniconi - Proceedings of the 14th ACM SIGKDD ..., 2008 - dl.acm.org
... Chief executive officer Shareholder Fiduciary Corporate governance Corporation Boards of directors Build Thesaurus from Wikipedia Ambiguous Concepts: Puma Puma (Car) "Felidae" Category "Puma" "Cougar" "Mountain Lion" "Ford Vehicles" Category "Puma (Car)" ...
Cited by 121 Related articles All 5 versions Cite Save
Prescriptive
Idea for a Norm

Descriptive
Observation of consistent behavior
Prescriptive
Idea for a Norm

Descriptive
Observation of consistent behavior
Prescriptive
Idea for a Norm

Descriptive
Observation of consistent behavior

ESSAY

Requests for Comments
Growth of regulations
Figure 8: Policy vector similarities for admin and registered users are converging to 1, indicating that the two user groups are becoming similar in their policy citations. The convergence of similarities of policy vectors 4 weeks apart within each population indicates stability within user classes.

Figure 9: On average, unique users who cite policy for the first time account for over 10% of all unique users who cite a policy in a week. The high percentage indicates that participation in Wikipedia’s governance structure is inclusionary.

Work allocation
- Identify, prioritize and assign tasks

Regulation of behavior
- Norm formation, propagation and enforcement

Quality control
- Identify and remove damage
Quality control
Quality control

Fully automated
- Fast (~ 5 seconds)[1]
- No human effort
- Only obvious vandalism

Semi automated
- Still pretty fast (~ 30 seconds)[1]
- Minimizes human effort
- Humans catch most vandalism at a glance

Quality control

Fully automated

Semi automated

Banning (Admins)
Quality control

Fully automated

Semi automated

Banning (Admins)

Innate
  - Fast
  - General
  - Pervasive

Adaptive
  - Slow
  - Specific
  - Pervasive


Work allocation
- Identify, prioritize and assign tasks

Regulation of behavior
- Norm formation, propagation and enforcement

Quality control
- Identify and remove damage

Community management
- Newcomer socialization, dispute mediation & training
Newcomer socialization

Research: The Rise and Decline

Key Personnel

Contact + Co-investigators and:
- Maryana Pinchuk
- Steven Walling
- Oliver Keyes

Project Overview

Open collaboration systems like Wikipedia need to maintain a pool of volunteer contributors in order to remain relevant. Despite historically garnering massive amounts of contributions, recent research has shown that the number of active contributors in Wikipedia has entered a steady decline and suggests that declining retention of newcomers is the cause. Through a data-driven approach, this paper presents evidence that several changes that the Wikipedia community made to manage quality and consistency in the face of a massive growth in participation have led to a more restrictive environment for newcomers. Specifically, the restrictiveness of the encyclopedia’s primary quality control mechanism and the algorithmic tools used to reject contributions is implicated as a cause of decreased newcomer retention. Also, the community’s formal mechanisms for norm articulation are shown to have calcified against changes – especially for newcomers. Recommendations are offered for Wikipedia’s community organizers and the Wikimedia Foundation.


https://meta.wikimedia.org/wiki/Research:The_Rise_and_Decline
TL;DR:

Fully automated

Semi automated

Baby

Bath Water
TL;DR:

Fully automated

Semi automated

Wikipedia

Quality control

The Teahouse

https://snuggle-en.wmflabs.org/
Work allocation
- Identify, prioritize and assign tasks

Regulation of behavior
- Form, propagate and enforce norms

Quality control
- Identify and remove damage

Community management
- Socialize & train newcomers; mediate disputes

Reflection (Adaptation)
- Where are we going? Where do we want to go? How do we want to get there?
Reflection (Adaptation)

Where *are* we going?

Where *do* we want to go?

*How* do we get there?
Reflection (Adaptation)

- Open model
- Formulated policies
- WP:BLP
- Automated counter-vandalism
- Teahouse
- Snuggle

Graph:
- Active editors
- Calendar year
- Early, Growth, Decline
Reflection (Adaptation)

WHERE’S ALL THE ADAPTATION?

- Open model
- Formalized policies
- WP:BLP
- Automated counter-vandalism

Graph showing the number of active editors over calendar years.
Reflection (Adaptation)

- What enables massive distributed communities the ability to adapt?
- Does the ability to adapt *necessarily* decrease over time? Can it be jump-started?
Work allocation

Regulation of behavior

Quality control

Community management

Reflection (Adaptation)
Part 3

What’s next?
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

1:10000

Monthly Wikipedia Readers (500 million)

Monthly Wikipedia Editors (113,304)
Work allocation

WikiGrok

Monthly Wikipedia Readers (500 million)

1:10000

Monthly Wikipedia Editors (113,304)
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Efficiency (overall)

https://meta.wikimedia.org/wiki/Research:HHVM_newcomer_engagement_experiment
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Efficiency (overall)

https://meta.wikimedia.org/wiki/Research:HHVM_newcomer_engagement_experiment
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Reflection

https://meta.wikimedia.org/wiki/Research:WikiCredit

Efficiency (overall)

https://meta.wikimedia.org/wiki/Research:HHVM_newcomer_engagement_experiment
Work allocation

WikiGrok

https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Efficiency (overall)

Reflection

https://meta.wikimedia.org/wiki/Research:WikiCredit

<table>
<thead>
<tr>
<th>Revisions</th>
<th>PWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Apples are red.</td>
<td>6</td>
</tr>
<tr>
<td>2: Apples are blue.</td>
<td>0</td>
</tr>
<tr>
<td>3: Apples are red.</td>
<td>0</td>
</tr>
<tr>
<td>4: Apples are tasty and red.</td>
<td>1</td>
</tr>
<tr>
<td>5: Apples are tasty and blue.</td>
<td>0</td>
</tr>
</tbody>
</table>
Work allocation

WikiGrok

[Image: https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok]

Efficiency (overall)

[Image: HHVM]

Reflection

Wiki Credit

[Image: https://meta.wikimedia.org/wiki/Research:WikiCredit]

Quarry βeta

[Image: Quarry beta]


[Image: Quarry beta demo]

https://meta.wikimedia.org/wiki/Research:HHVM_newcomer_engagement_experiment