Managing Open Production at Scale

Aaron Halfaker, PhD

Wikimedia Research
/who
/who

- Social technologist
who

- Social technologist
/who

- Social technologist
/who

- Social technologist
/who

- Social technologist
/who

• Social technologist

Science → Theory → New technology
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

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

Contact me
- E-mail: ahalfaker
  @wikimedia.org
- Website: http://halfaker.info
- Twitter:
Outline
Outline

1. Wikipedia is *academically interesting*
Outline

1. Wikipedia is *academically interesting*

2. My work studying the *socio-technical*
1. Wikipedia is *academically interesting*

2. My work studying the *socio-technical*

3. WMF Research *community support*
Part 1.

Wikipedia is academically interesting
What is Wikipedia?
It’s a wiki.
It’s a wiki.

- What’s a wiki?
Traditional Wiki

![Diagram showing the difference between Traditional and Wiki documents](image-url)
Traditional Wiki

Document

Audience
Traditional

You suck!

Wiki

Audience
Traditional

Wiki

Document

You suck!

Abe Lincoln was gay

Audience
Traditional

Wiki

Document

You suck!

Audience
What is Wikipedia?

- Wiki platform
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage
- Open governance model
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage

- Open governance model
  - Who is allowed to contribute? Everyone!
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage

- Open governance model
  - Who is allowed to contribute? Everyone!
  - Who owns the content? Everyone?
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage

- Open governance model
  - Who is allowed to contribute? Everyone!
  - Who owns the content? Everyone?
  - Who makes the rules? ...everyone?
What is Wikipedia?

- Wiki platform
  - Readers → Contributors
  - Openness → more contributions + damage

- Open governance model
  - Who is allowed to contribute? Everyone!
  - Who owns the content? Everyone?
  - Who makes the rules? ...everyone?

How’s that going to work?
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
Cited by 3  Related articles  All 5 versions  Cite  Save

Changing governance of local economies: Responses of European local production systems
... 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
Wikipedia is absurd
Wikipedia is absurd

- Anyone can edit. No registration necessary.
Wikipedia is absurd

- Anyone can edit. No registration necessary.
- No reward system. If anything, negative attention.
Wikipedia is absurd

● Anyone can edit. No registration necessary.

● No reward system. If anything, negative attention.

● Actually a wild success!
Planet Earth
(6.97 billion)
data from March 2012
Planet Earth
(6.97 billion)

1:10

Monthly Wikipedia Readers
(500 million)

data from March 2012
Planet Earth
(6.97 billion)

Monthly Wikipedia Editors
(113,304)

1:10000

Monthly Wikipedia Readers
(500 million)

1:10

data from March 2012
Monthly Wikipedia Editors (113,304)

Planet Earth (6.97 billion)

Monthly Wikipedia Readers (500 million)

data from March 2012
Powerlaw of Contribution

Number of contributors vs. Amount of contribution
Powerlaw of Contribution

Number of contributors

Amount of contribution

User: Koavf
(> 1 million)
Powerlaw of Contribution

Amount of contribution

Number of contributors

Me

(~ 5000 edits)

User: Koavf

(> 1 million)
Powerlaw of Contribution

Amount of contribution

Number of contributors

Drive-by typo fixers

Me
(~ 5000 edits)

User:Koavf
(> 1 million)
Powerlaws everywhere

- Wikipedia
- Bugzilla
- Digg
- Essemble

Powerlaws everywhere

- Wikipedia
- Bugzilla
- Digg
- Essembly (???)
Essembly

From Wikipedia, the free encyclopedia

Essembly is a non-partisan political social networking website that allows its users to connect with one another based on political opinions, participate in unmoderated discussion, and organize for political action. It was founded in 2005 by Joe Green, a Harvard graduate and college roommate of Mark Zuckerberg, the founder of the collegiate social-networking site Facebook. Essembly differentiates itself from other social-networking sites by actively promoting intelligent discussion and debate while remaining (as the site’s “About Us” page claims) “fiercely non-partisan.” A July 2006 Newsweek article put the number of registered users at over 17,000[1]

After its launch in 2005, Essembly membership was offered by invitation only, as the website was still in closed beta. In May 2006, Essembly opened membership to the public (to anyone with a valid email address), and began to run advertisements on Facebook. As of September 2006, Essembly remains in public beta, the staff of Essembly has, on the whole, moved on to a new venture called Project Agape. This has resulted in intermittent user support and server crashes, as well as some bitterness by many of the long-time users.[who?]  

On January 30, 2010 Essembly stopped working. After being alerted to the problem by a long-time Essembly user, Joe Green restored the site by Valentine's Day. However, since May, 2010 the site has been down.
How much time goes into Wikipedia?
How much time goes into Wikipedia?

Estimated session duration: 2603 seconds
How much time goes into Wikipedia?

Jimmy Wales

Estimated session duration: 2603 seconds

- Time line:
  - 00:04: Estimated session start.
  - 00:11
  - 00:20
  - 00:29
  - 00:47

- Actions:
  - Adds link to "Little Rock Nine"
  - Removes self-reference link
  - Creates redirect "American (U.S.)"
  - Creates article: "Virgil Blossom"
How much time goes into Wikipedia?

Jimmy Wales

- **English Wikipedia**: 41,018,804 hours

March, 2012
How much time goes into Wikipedia?

- **English Wikipedia:** 41,018,804 hours
- **Empire State Building:** ~7,000,000
How much time goes into Wikipedia?

- **English Wikipedia**: 41,018,804 hours
- **Empire State Building**: ~7,000,000
- **Channel tunnel**: ~170,000,000
How much time goes into Wikipedia?

Jimmy Wales

- **English Wikipedia**: 41,018,804 hours
- **Empire State Building**: ~7,000,000
- **Channel tunnel**: ~170,000,000
- **1st Encyclopedia Britannica**: ~12,000

https://meta.wikimedia.org/wiki/Research:Edit_session
How did Wikipedia succeed?
How did Wikipedia succeed?

- It’s so easy to vandalize!
Elephant

This article is about the living species. For extinct relatives also known as elephants, see Elephantidae. For other uses, see Elephant (disambiguation).

Elephants are large mammals of the family Elephantidae and the order Proboscidea. Traditionally, two species are recognised, the African elephant (Loxodonta africana) and the Asian elephant (Elephas maximus), although some evidence suggests that African bush elephants and African forest elephants are separate species (L. africana and L. cyclotis respectively). Elephants are scattered throughout sub-Saharan Africa, South Asia, and Southeast Asia. Elephantidae are the only surviving family of the order proboscidea; other, now extinct, families of the order include mammoths and mastodons. The largest living terrestrial animals, male African elephants can reach a height of 4 m (13 ft) and weigh 7,000 kg (15,000 lb). All elephants have several distinctive features the most notable of which is a long proboscis or trunk used for many purposes, particularly for grasping objects. In addition to their trunks, their incisors grow into tusks, which serve as tools for moving objects and digging and as weapons for fighting. The elephant's large ear flaps help to control the temperature of its body. African elephants have larger ears and concave backs while Asian elephants have smaller ears and convex or level backs.

Elephants are herbivorous and can be found in different habitats including savannas, forests, deserts and marshes. They prefer to stay near water. They are considered to be keystone species due to their impact on their environments. Other animals tend to keep their distance, and predators such as lions, tigers, hyenas and wild dogs usually target only the young elephants (or "calves"). Females ("cows") tend to live in family groups, which can consist of one female with her calves or several related females with offspring. The groups are led by an individual known as the matriarch, often the oldest cow. Elephants have a fission-fusion society in which multiple family groups come together to socialise. Males ("bulls") leave their family groups when they reach puberty, and may live alone or with other males. Adult bulls mostly interact with family groups when looking for a mate and enter a state of increased testosterone and aggression known as musth, which helps them gain dominance and
''Elephants'' are STUPID mammals of the family [[Elephantidae]] and the order [[Proboscidea]]. Traditionally, two species are recognised, the [[African elephant]] (''Loxodonta africana'') and the [[Asian elephant]] (''Elephas maximus''), although some evidence suggests that [[African bush elephant]]s and [[African forest elephant]]s are separate species (''L.&nbsp;africana'' and ''L.&nbsp;cyclotis'' respectively). Elephants are scattered throughout [[sub-Saharan Africa]], [[South Asia]], and [[Southeast Asia]]. [[Elephantidae]] are the only surviving family of the order proboscidea; other, now extinct, families of the order include [[mammoth]]s and [[mastodon]]s. The largest living [[terrestrial animal]]s, male African elephants can reach a height of {{convert|4|m|ft|0|abbr-on}} and weigh {{convert|7000|kg|lb|-3|abbr-on}}. All elephants have several distinctive features the most notable of which is a long [[proboscis]] or trunk used for many purposes, particularly for grasping objects. In addition to their trunks, their [[incisor]]s grow into tusks, which serve as tools for moving objects and digging and as weapons for fighting. The elephant's large ear flaps help to control the temperature of its body. African elephants have larger ears and concave backs while Asian elephants have smaller ears and convex or level backs.
Wikipedia works in practice.
Wikipedia works in practice.

But it doesn’t work in Theory.
Wikipedia works in *practice*.

But it doesn’t work in *Theory*.
Part 2.

Towards a theory of wikis
How to get some Theory
How to get some Theory

1. Observe unexplained phenomena
How to get some Theory

1. Observe unexplained phenomena

2. Hypothesize explanations
How to get some Theory

1. Observe unexplained phenomena
2. Hypothesize explanations
3. Look for evidence
How to get some Theory

1. Observe unexplained phenomena
2. Hypothesize explanations
3. Look for evidence
4. Repeat
WHY?
A long time ago -- before it was obvious that Wikipedia worked at all...
- Small
- Few editors
- Fewer vandals
Little time needed for:
- Vetting new articles
- Removing vandalism
- Socializing newcomers
Popularity $\rightarrow$ More readers $\rightarrow$ More editors
Popularity → More readers → More editors

More editors → More better, right?
The London Beer Flood happened on 16 October 1814 in the parish of St. Giles, London, England. At the Meux and Company Brewery[^1][^2] on Tottenham Court Road[^1] a huge vat containing over 135,000 imperial gallons (610,000 L) of beer ruptured, causing other vats in the same building to succumb in a domino effect. As a result, more than 323,000 imperial gallons (1,470,000 L) of beer burst out and gushed into the streets. The wave of beer destroyed two homes and crumbled the wall of the Tavistock Arms Pub, trapping teenage employee Eleanor Cooper under the rubble.[^3] Within minutes
the former brewery. In 2012, a local tavern, the "Holborn Whippet", has started to mark this event with a specially created vat of porter brewed especially for the day.

**Known drowning fatalities**  [edit]

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann Saville</td>
<td>53</td>
</tr>
<tr>
<td>Eleanor Cooper</td>
<td>15–16</td>
</tr>
<tr>
<td>Hannah Bamfield</td>
<td>4</td>
</tr>
<tr>
<td>Catherine Butler</td>
<td>63</td>
</tr>
<tr>
<td>Elizabeth Smith</td>
<td>27</td>
</tr>
<tr>
<td>Mary Mulvey</td>
<td>30</td>
</tr>
<tr>
<td>Thomas Mulvey</td>
<td>3</td>
</tr>
<tr>
<td>Sean Duggins</td>
<td>29</td>
</tr>
</tbody>
</table>

**See also**  [edit]

- Boston Molasses Disaster
- Great Fire of London
- Honolulu molasses spill

**References**  [edit]
Popularity $\rightarrow$ More readers $\rightarrow$ More editors

New editor/content firehose!
14 November 2013

20:08 Mohadevpur Upazila (2 changes | history) (.+7,205) . [Sabbire.p (2×)]
20:08 Thalia discography (diff | hist) . (.59) . 88marcus (talk | contribs) (Undid revision 581605390 by Sailormoon94 (talk) wrong charts)
20:08 Vinny Appice (diff | hist) . (.+29) . 99.99.165.175 (talk) (→ with Kill Devil Hill)
20:08 List of songs produced by Dallas Austin (6 changes | history) . (.+199) . [172.56.4.215 (6×)]
20:08 Old Second National Bank (Aurora, Illinois) (diff | hist) . (.+44) . Teemu08 (talk | contribs)
20:08 User creation log . [ChristieAnnMarchese; Erichtaylorach; Bennyjw123; BeatnikS11; AnIreddy248; Bazzfazzmillenium; Flezyb13; Frankfair; Rock23953; SeliVanisrinivasan; Plastig001; Msmsmarlow; Imawesome24; Lloyd Burgundy; Tonyjzheng]
20:08 Global Hunger Index (diff | hist) . (.+2,466) . Marianian (talk | contribs) (Reverted 1 edit by 71.255.186.70 (talk) to last revision by 98.151.215.4. (TW))
20:08 Wisden Cricketers of the Year (diff | hist) . (.+2) . 101.62.216.24 (talk) (→ List of Cricketers of the Year)
20:08 Climate (4 changes | history) . (0) . [Rhinestone K; ClueBot NG; 152.26.73.92 (2×)]
20:08 Casualties of the Syrian civil war (diff | hist) . (.+13) . EkoGraf (talk | contribs) (→ Opposition forces)
20:08 Georges Danton (diff | hist) . (.+4) . Pavel Voznenilek (talk | contribs) (→ Arrest, trial, and execution: Antoine Quentin Fouquier-Tinville linked)
20:08 St. Peter, Minnesota (2 changes | history) . (0) . [66.220.213.193 (2×)]
20:08 Daisuke (diff | hist) . (.+96) . ElizabethEmig (talk | contribs)
20:08 Treaty of Paris (1783) (diff | hist) . (0) . Illegitimate Barrister (talk | contribs) (Undid revision 581628816 by 63.116.236.22 (talk))
20:08 Lenovo Vibe Z (2 changes | history) . (.7) . [Intermittentgardener (2×)]
20:08 Piscvore (diff | hist) . (.+18) . Claire.Edelman (talk | contribs) (→ Examples of extant piscivores)
20:08 Timeline of heavy metal music (6 changes | history) . (.+72) . [65.7.223.111 (6×)]
20:08 Konkani language (2 changes | history) . (.+48) . [PaintedCarpet (2×)]
20:08 Devakottai (diff | hist) . (0) . ChristieAnnMarchese (talk | contribs) (→ History) (Tag: new editor getting started)
20:08 Twelve Olympians (diff | hist) . (.334) . Paul August (talk | contribs) (Undid revision 581664683 by Andybishop (talk))
20:08 List of The Ricky Gervais Show (animated series) episodes (3 changes | history) . (0) . [2.124.96.226 (3×)]
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Event</th>
<th>Author/Contributor</th>
<th>Revision Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 November</td>
<td>29:08</td>
<td>Mohadevpur Upazila (2 changes</td>
<td>history) (+7,205) [Sabbire.p (2x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Thalia discography (diff</td>
<td>hist) (-59) 88marcus (talk</td>
<td>contribs) (Undid revision 581605390 by Sailormoon94 (talk) wrong charts)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Vinny Appice (diff</td>
<td>hist) (+29) 99.99.167.175 (talk) (with Kill Devil Hill)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Wikipedia:Administrators' noticeboard/Incidents (diff</td>
<td>hist) (+290) Jprg1966 (talk</td>
<td>contribs) (Vandalism by User talk:Kppardett closing as resolved)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>List of songs produced by Dallas Austin (6 changes</td>
<td>history) (+199) [172.56.4.219 (6x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Old Second National Bank (Aurora, Illinois) (diff</td>
<td>hist) (+44) Teemu08 (talk</td>
<td>contribs)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>User creation log [ChristieAnnMarchese; Ercitaylorch; Bennyjw123; Beatinist511; Anilreddy248; Bazzfazzmillenium; Flezyb13; Frankfair; Rock23953; Sevaranisririvasan; Plastic003; Msmsmarow; Imawesome24; Lloyd Burgundy; Tonyzheng]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Global Hunger Index (diff</td>
<td>hist) (+2,466) Marianian (talk</td>
<td>contribs) (Reverted 1 edit by 71.255.186.70 (talk) to last revision by 98.151.215.4 (TW))</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Wisden Cricketers of the Year (diff</td>
<td>hist) (+2) 101.62.216.24 (talk) (List of Cricketers of the Year)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Climate (4 changes</td>
<td>history) (0) [Rhinestone K; ClueBot NG; 152.26.73.92 (2x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Casualties of the Syrian civil war (diff</td>
<td>hist) (+13) EkoGraf (talk</td>
<td>contribs) (Opposition forces)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Georges Danton (diff</td>
<td>hist) (+4) Pavel Vozenilek (talk</td>
<td>contribs) (Arrest, trial, and execution: Antoine Quentin Fouquier-Tinville linked)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>St. Peter, Minnesota (2 changes</td>
<td>history) (0) [66.220.213.193 (2x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Daisuke (diff</td>
<td>hist) (+96) ElizabethEmig (talk</td>
<td>contribs)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Treaty of Paris (1783) (diff</td>
<td>hist) (0) Illegitimate Barrister (talk</td>
<td>contribs) (Undid revision 581628816 by 63.116.236.2 (talk))</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Lenovo Vibe Z (2 changes</td>
<td>history) (-7) Intermittentgardener (2x)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Pischore (diff</td>
<td>hist) (+18) Claire.Edelman (talk</td>
<td>contribs) (Examples of extant piscivores)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Timeline of heavy metal music (6 changes</td>
<td>history) (+72) [65.7.223.111 (6x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Konkani language (2 changes</td>
<td>history) (+48) [PaintedCarpet (2x)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Devakottai (diff</td>
<td>hist) (0) ChristieAnnMarchese (talk</td>
<td>contribs) (History) (Tag: new editor getting started)</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>Twelve Olympians (diff</td>
<td>hist) (-334) Paul August (talk</td>
<td>contribs) (Undid revision 581664683 by Andyb浕ş at (talk))</td>
</tr>
<tr>
<td></td>
<td>29:08</td>
<td>List of The Ricky Gervais Show (animated series) episodes (3 changes</td>
<td>history) (0) [2.124.96.226 (3x)]</td>
<td></td>
</tr>
</tbody>
</table>
The Elephant Incident

Active editors

Calendar year

Early

Decline

2002

2004

2006

2008

2010

2012

32,000
Elephant

From Wikipedia, the free encyclopedia

This article is about the living species. For extinct relatives also known as elephants, see Elephantidae. For other uses, see Elephant (disambiguation).

Elephants are large mammals of the family Elephantidae and the order Proboscidea. Traditionally, two species are recognised, the African elephant (Loxodonta africana) and the Asian elephant (Elephas maximus), although some evidence suggests that African bush elephants and African forest elephants are separate species (L. africana and L. cyclotis respectively). Elephants are scattered throughout sub-Saharan Africa, South Asia, and Southeast Asia. Elephantidae are the only surviving family of the order Proboscidea; other, now extinct, families of the order include mammoths and mastodons. Male African elephants are the largest surviving terrestrial animals and can reach a height of 4 m (13 ft) and weigh 7,000 kg (15,000 lb). All elephants have several distinctive features the most notable of which is a long trunk or proboscis, used for many purposes, particularly breathing, lifting water and grasping objects. Their incisors grow into tusks, which can serve as weapons and as tools for moving objects and digging. Elephants' large ears are used to increase their hearing range and their body temperature is low, which helps them cool off. They are herbivores and feed mainly on grasses, leaves, bark, fruits and tree sap. Many species are endangered due to poaching and habitat loss.
Elephant

From Wikipedia, the free encyclopedia

This article is about the living species. For extinct relatives also known as elephants, see Elephantidae. For other uses, see Elephant (disambiguation).

Elephants are large mammals of the family Elephantidae and the order Proboscidea. Traditionally, two species are recognised, the African elephant (Loxodonta africana) and the Asian elephant (Elephas maximus), although some evidence suggests that African bush elephants and African forest elephants are separate species (L. africana and L. cyclotis respectively). Elephants are scattered throughout sub-Saharan Africa, South Asia, and Southeast Asia. Elephantidae are the only surviving family of the order Proboscidea; other, now extinct, families of the order include mammoths and mastodons. Male African elephants are the largest surviving terrestrial animals and can reach a height of 4 m (13 ft) and weigh 7,000 kg (15,000 lb). All elephants have several distinctive features the most notable of which is a long trunk or proboscis, used for many purposes, particularly breathing, lifting water and grasping objects. Their incisors grow into tusks, which can serve as weapons and as tools for moving objects and digging. Elephants' large ears help regulate body temperature.
The Elephant Incident

Wikimania 2006

32,000
WIKIMANIA '06

IDENTIFYING
AND ELIMINATING
TROLLS
Technology as Counter-vandalism
Fully automated

Machine learning

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

Machine learning

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


Human computation

- Still pretty fast (~ 30 seconds)[1]
- Minimizes human effort
- Humans catch most vandalism at a glance
What’s actually declining?
What’s actually declining?

Newcomers! Show up, but leave quickly.

What’s happening to the newbies?

- Wikipedia is done.
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*[1]

1. [https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000](https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000)
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*[1]
- Wikipedia has enough editors.

1. https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*[1]
- Wikipedia has enough editors. *It doesn’t.*[2,3]

1. [https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000](https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000)
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.* [1]
- Wikipedia has enough editors. *It doesn’t.* [2,3]
- Early adopter effect.

1. https://en.wikipedia.org/wiki/Wikipedia:Core_topics_-_1,000
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.* [1]
- Wikipedia has enough editors. *It doesn’t.* [2,3]
- Early adopter effect.

*Newbies just aren’t as good as they used to be.*
- Crotchety Wikipedians
What’s happening to the newbies?
Newcomer rejection especially by anti-vandal tools predicts the decline.
Newcomer rejection, especially by anti-vandal tools, predicts the decline.
Newcomer rejection especially by anti-vandal tools predicts the decline.
80% of first interactions automated
TL;DR:

Newcomers → Wikipedia
TL;DR:

Newcomers → Wikipedia
TL;DR:

Newcomers → Quality control → Wikipedia
TL;DR:

Newcomers

Wikipedia

Quality control
TL;DR:

What's causing Wikipedia's editor decline?


[Image: Baby vs. Bathwater Annotated.JPG by Richfife (Own work) [Public domain], via Wikimedia Commons]
What can be done?
What can be done?

- Efficient quality control is necessary
What can be done?

- Efficient quality control is necessary
- Socializing new editors is necessary
<table>
<thead>
<tr>
<th>Username</th>
<th>Interactive edit graph</th>
<th>Talk page activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinineburx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Betabot75</td>
<td>W</td>
<td>wam</td>
</tr>
<tr>
<td>Rodriguezandres999</td>
<td>vand</td>
<td>TH</td>
</tr>
<tr>
<td>PriscillaFilix</td>
<td></td>
<td>csd</td>
</tr>
<tr>
<td>RichardHC</td>
<td></td>
<td>TH</td>
</tr>
<tr>
<td>J3sseung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lostbad</td>
<td>W</td>
<td>vand</td>
</tr>
<tr>
<td>Wikij34</td>
<td></td>
<td>TH</td>
</tr>
<tr>
<td>Dswatcha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keaclamvictot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meleri Morgan</td>
<td>W</td>
<td>TH</td>
</tr>
<tr>
<td>Tomski12</td>
<td>vand</td>
<td></td>
</tr>
<tr>
<td>Nordicanuck</td>
<td></td>
<td>W</td>
</tr>
<tr>
<td>MarieMarshall</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More good

Less good
Welcome to the 
teahouse
A friendly place to help new editors become accustomed to Wikipedia culture, ask questions, and develop community relationships.

Do you have a question about editing?

Get answers

Recent questions...

I am a bit confused after trying to edit this page:
https://en.wikipedia.org/wiki/Cycling_advocacy#Copenhagenization

I wanted to add relevant supporting links but have run into a battle with another user. I wanted to link to a TED talk by one of the individuals mentioned under Bicycle Advocates - Mikael Colville-Andersen - but another user keeps undoing it, even though allowing a link to a talk by another one of the people mentioned.

Warrenkychu
On Wikipedia I wish to edit and contribute to the best of my ability. I enjoy editing Wikipedia.
Growth Experiments

Projects [edit]

The following pages hold the methodology, experimental design and results of each experiment.

Active
- Research:Task recommendations
- Research:Anonymous editor acquisition
  - Research:Asking anonymous editors to register
  - Research:Anonymous phenomena
- Donor email campaigns

Complete
- Research:Onboarding new Wikipedians
- Research:GuidedTour usage
- Research:Wikipedia article creation
- Research:Newcomer survival models
- Microsurveys, e.g. Research:Gender micro-survey in support of VisualEditor
- Research:Account creation UX
- Research:Donor engagement
- Research:Community portal redesign
- Research:Post-registration editor survey
- Research:Post-edit feedback
- Template A/B testing
- Research:Necromancy
- Research:Timestamp position modification
- Research:Editor milestoness

Proposals and ideas
- Research:Section edit modification
- Research:Improve your edit
- Research:Wikimania unconference experiment (suppress cleanup templates)
- Research:Left-hand navigation
Growth Experiments

Projects [edit]

The following pages hold the methodology, experimental design and results of each experiment.

Active
- Research: Task recommendations
- Research: Anonymous editor acquisition
  - Research: Asking anonymous editors to register
  - Research: Anonymous phenomena
- Donor email campaigns

Complete
- Research: Onboarding new Wikipedians
- Research: Guided Tour usage
- Research: Wikipedia article creation
- Research: Newcomer survival models
- Microsurveys, e.g. Research: Gender micro-survey in support of VisualEditor
- Research: Account creation UX
- Research: Donor engagement
- Research: Community portal redesign
- Research: Post-registration editor survey
- Research: Post-edit feedback
- Template A/B testing
- Research: Necromancy
- Research: Timestamp position modification
- Research: Editor milestones

Proposals and ideas
- Research: Section edit modification
- Research: Improve your edit
- Research: Wikimania unconference experiment (suppress cleanup templates)
- Research: Left-hand navigation

Pages to improve

Recommended based on your last edit

- Arctic Fox
  94 views this month
- Mount Whitney
  924 views this month
- Nova Scotia Duck Tolling Retriever Dog farts blah blah...
  2394 views this month
Growth Experiments

Projects [edit]

The following pages hold the methodology, experimental design and results of each experiment.

Active
- Research: Task recommendations
- Research: Anonymous editor acquisition
  - Research: Asking anonymous editors to register
  - Research: Anonymous phenomena
- Donor email campaigns

Complete
- Research: Onboarding new Wikipedians
- Research: Guided tour usage
- Research: Wikipedia article creation
- Research: Newcomer survival models
- Microsurveys, e.g. Research: Gender micro-survey in support of VisualEditor
- Research: Account creation UX
- Research: Donor engagement
- Research: Community portal redesign
- Research: Post-registration editor survey
- Research: Post-edit feedback
- Template A/B testing
- Research: Necromancy
- Research: Timestamp position modification
- Research: Editor milestones

Proposals and ideas
- Research: Section edit modification
- Research: Improve your edit
- Research: Wikimania unconference experiment (suppress cleanup templates)
- Research: Left-hand navigation

Pages to improve

Recommended based on your last edit

Arctic Fox
94 views this month

Mount Whitney
924 views this month

Nova Scotia Duck Tolling Retriever Dog farts blah blah...
Growth Experiments

Hypotheses  [edit]

Rationale: We suspect that the prominence of guider-based UI cues, as well as a recent concern that anonymous editors will increase the rate at which anonymous editors decide to register an account on Wikipedia are passive for anonymous editors, in the form of static links to account creation or the edit form. We suspect that these passive "create account" links (in the personal information messages mid-edit) are relatively hidden to a new users' eyes. We've observed that this increases completion rates within a funnel.[1]

Hypothesis 1: More anonymous editors in the experimental conditions will register an account.

Rationale: Further, we suspect that the pre-edit CTA will encourage more users to react between a user and their goal -- to make an edit to a page. Further, the flow of in is a common pattern in web applications and is already used by mw:Extension:Mo only functionality like watchlists.

Hypothesis 2: More anonymous editors in the pre-edit conditions will register an account.

Proposals and ideas
- [Research:Section edit modification]
- [Research:Improve your edit]
- [Research:Wikimania unconference experiment (suppress cleanup templates)]
- [Research:Left-hand navigation]
Growth Experiments

Hypotheses

Rationale: We suspect that concerns, will increase traffic to articles on Wikipedia are passive for new users, and that messages mid-edit are more likely to increase completion rates.

Hypothesis 1: More articles

Rationale: Further, we suspect that the action between a user and a page is a common pattern in which only functionality like wikitext is used.

Hypothesis 2: More articles

Methods

Measuring user behavior via random token

In order to distinguish between distinct clients as they move between IP addresses, we will be assigning users randomly-generated tokens that will identify their browser/device during the observation and experiment periods. This token is attached to all user edits, either via wikitext or VisualEditor. Due to the potential sensitivity of user behavior, this data is retained for up to 90 days (per the Data retention guidelines) and only used for analytics databases.

Note: There are limitations in the assumption that one token equals one user to one device. Possible complications include shared computers (some users have different sessions, others don’t), which have different effects, users who habitually use multiple devices and automatically clear cookies on browser close, and users who switch between browser devices.

Schemas

Schema: SignupExpPageLinkClick
Growth Experiments

Hypotheses

Rationale: We suspect that concerns, will increase the edit form. We suspect messages mid-edit) are increases completion rates.

Hypothesis 1: More accounts

Rationale: Further, we see a common pattern in only functionality like wa in.

Hypothesis 2: More accounts

Methods

Measuring user behavior

In order to distinguish between accounts, we will be assigning browser/device during the user edits, either via wikite retained for up to 90 days patterns (e.g. number of times from analytics databases).

Note: There are limitation to one device. Possible co sessions, others don’t, which automatically clear cookies.

Results

Hyp 1 recovered productivity

![Graph showing article edits and productive edits](image-url)

Proposals and ideas

- Research: Section edit mod
- Research: Improve your editor
- Research: Wikimania uncon
- Research: Left-hand navigation

Schemas

Schema: SignUpExpPage
We need **Theory** in order to effectively support Wikipedia-like projects.
Part 3

Research strategy & Community support
Critical insights
Critical insights

- Hard problems are hard
Critical insights

● Hard problems are hard

● Open collaboration is more productive
The problem space
The problem space
Openness

Access: [https://meta.wikimedia.org/wiki/Research:Index](https://meta.wikimedia.org/wiki/Research:Index)

Data:
- [http://dumps.wikimedia.org/](http://dumps.wikimedia.org/) -- General data dumps from public wikis

Source:
- [https://github.com/halfak/MediaWiki-Utilities](https://github.com/halfak/MediaWiki-Utilities) -- General data processing utilities
- [https://github.com/halfak/MediaWiki-OAuth](https://github.com/halfak/MediaWiki-OAuth) -- Authorization for wiki tools
- [https://github.com/halfak/Deltas](https://github.com/halfak/Deltas) -- Experimental difference detection
- [https://github.com/halfak/Wiki-Class](https://github.com/halfak/Wiki-Class) -- Automatic article quality classification
- [https://github.com/halfak/MediaWiki-events](https://github.com/halfak/MediaWiki-events) -- Standardized public event utilities
Outreach

Showcases
- https://www.mediawiki.org/wiki/Analytics/Research_and_Data/Showcase

Hackathons
- https://meta.wikimedia.org/wiki/Research:Labs2/Hackathons/August_6-7th,_2014

wiki-research-l@lists.wikimedia.org
- https://lists.wikimedia.org/mailman/listinfo/wiki-research-l

#wikimedia-research @ irc.freenode.net (connect)
Outreach

CSCW Workshop:
- Advancing an industry/academic partnership model for Open Collaboration research (under review)
Need data?

I want to help you get it.
Want to run an intervention?

I want to help you get access.
Want to collaborate?

I’m game.
... and if I can’t help you,

I’ll get you in touch with someone who can.
TL;DR:

We couldn’t succeed without you.

So, we’re going to put a lot of energy behind your work.
Thank you

Aaron Halfaker
ahalfaker@wikimedia.org

Send me emails about science, Wikipedia, the internet or whatever.

Ping me (halfak) in #wikimedia-research on freenode