Wikipedia

Maintaining Production Efficiency

Aaron Halfaker

User:EpochFail
User:Halfak (WMF)

ahalfaker@wikimedia.org
Outline

1. Wikipedia is Awesome
2. (English) Wikipedia’s discontents
3. What can we do?
Part 1: Wikipedia is Awesome
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, contributing is likely to bring negative attention.
Wikipedia is absurd

- Anyone can edit. No registration necessary.
- No reward system. If anything, contributing is likely to bring negative attention.
- Actually a wild success!
Planet Earth

(6.97 billion)

data from March 2012
Monthly Wikipedia Readers (1.5 billion)

Planet Earth (6.97 billion)

data from March 2012
Planet Earth (6.97 billion)

Monthly Wikipedia Readers (1.5 billion)

Monthly Wikipedia Editors (113,304)

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

Planet Earth (6.97 billion)

Monthly Wikipedia Readers (1.5 billion) 1:10000

1:5
Powerlaw of Contribution

Number of contributors vs. Amount of contribution
Powerlaw of Contribution

Number of contributors

Amount of contribution

Me
(~ 5000 edits)

User: Koavf
(> 1 million)
Powerlaw of Contribution

Number of contributors

Amount of contribution

Drive-by typo fixers

Me
(~ 5000 edits)

User:Koavf
(> 1 million)
Powerlaws everywhere

- Wikipedia
- Bugzilla
- Digg
- Essembly

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.[2]

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?

Jimmy Wales

Estimated session duration: 2603 seconds

- 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?

Estimated session duration: 2603 seconds

- 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?

- **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

From Wikipedia, the free encyclopedia ▶ 4 reader comments

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 savannahs, 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 attractiveness to potential mates. If the bull is not accepted as a member of the group, it may be chased out, or the matriarch may reduce musth.
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. 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 [[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, not in theory.

--?????
Wikipedia works in practice, not in theory.

- Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Wikipedia works in practice, not in theory.

- Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Wikipedia works in practice, not in theory.

- Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Wikipedia works in practice, not in theory.

- Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Wikipedia works in practice, not in theory.

- Why is Wikipedia so successful?
- Where are we less successful?
- How can we fix that?
Part 2: (English)
Wikipedia’s discontents
WTF?
A long time ago on an internet that had yet to take Wikipedia seriously ...
- Small
- Few editors
- Fewer vandals
Most of editors’ time was spent adding content to articles.
Popularity → More readers → More editors
Popularity → More readers → More editors

More editors → More better, right?
Popularity → More readers → More editors

= Massive new burden to maintain quality
London Beer Flood

From Wikipedia, the free encyclopedia

The London Beer Flood happened on 16 October 1814 in the parish of St. Giles, London, England. At the Meux and Company Brewery[1] on Tottenham Court Road,[1][2] 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]
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) . Teemuo8 (talk | contribs)
20:08 (User creation log) . [ChristieAnnMarchese; Erixtayloroch; Bennyjw123; BeatinK511; Anirreddy248; Bazzfazzmillenium; Flezyb13; Frankfair; Rock23953; Selvaranisrinivasan; Plastig003; Msmlarow; 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 Vozenilek (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) . Intermittenntgardener (2×)
20:08 Pischore (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>Time</th>
<th>User</th>
<th>Action</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>2 changes</td>
<td>Mohadevpur Upazila (2 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>diff</td>
<td>Thalia discography (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>29</td>
<td>Vinny Appice (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>290</td>
<td>Wikipedia:Administrators' noticeboard/Incidents (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>199</td>
<td>List of songs produced by Dallas Austin (6 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>44</td>
<td>Old Second National Bank (Aurora, Illinois) (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>283</td>
<td>(User creation log) .. [ChristieAnnMarchese; Erticaylorch; Bennyjw123; Beatinik511; Anirredy248; Bazzfazzmillenium; Flezyb13; Frankfair; Rock23953; Selvanarinsivasan; Plastig001; Msmlarow; Imawesome24; Lloyd Burgundy; Tonyzheng]</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>2466</td>
<td>Global Hunger Index (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>2</td>
<td>Wisden Cricketers of the Year (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>Climate (4 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>13</td>
<td>Casualties of the Syrian civil war (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>4</td>
<td>Georges Danton (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>St. Peter, Minnesota (2 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>96</td>
<td>Daisuke (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>Treaty of Paris (1783) (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>7</td>
<td>Lenovo Vibe Z (2 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>18</td>
<td>Pischivere (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>72</td>
<td>Timeline of heavy metal music (6 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>48</td>
<td>Konkani language (2 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>Devakottai (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>334</td>
<td>Twelve Olympians (diff</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>List of The Ricky Gervais Show (animated series) episodes (3 changes</td>
</tr>
<tr>
<td>29:08</td>
<td>Saffirio</td>
<td>0</td>
<td>Between the windows (2003 film) (diff</td>
</tr>
</tbody>
</table>
The Elephants Incident

32,000
The Elephant Incident

Wikimania 2006

Early Growth Decline

Active editors

Calendar year


0 10000 20000 30000 40000 50000

32,000
IDENTIFYING AND ELIMINATING TROLLS
Fully automated

Machine learning

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

- Machine learning
  - Fast (~ 5 seconds)
  - No human effort
  - *Only obvious vandalism*

Semi automated

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

Newcomers!

What’s happening to the newbies?
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*
- Wikipedia has enough editors. *It doesn’t.*
What’s happening to the newbies?

- Wikipedia is done. *It isn’t.*
- Wikipedia has enough editors. *It doesn’t.*
- Early adopter effect.
What’s happening to the newbies?

- Wikipedia is done.  *It isn’t.*
- Wikipedia has enough editors.  *It doesn’t.*
- Early adopter effect.

“Newcomers just aren’t as good what they used to be.”  -Crotchety Wikipedians
What’s happening to the newbies?
What’s happening to the newbies?

Thanks!
Oliver Keyes
Maryana Pinchuk
Steven Walling
Stuart Geiger

2100 newcomers
Quality newcomers

Rejection of first edits

Declining retention
Quality newcomers

Rejection of first edits

Declining retention

Proportion of newcomers

Year

Proportion of surviving newcomers

Proportion of reverted newcomers

Year
Quality newcomers

Reverts of first edits

Declining retention
Quality newcomers

Rejection of first edits

Declining retention
Quality newcomers

Rejection of first edits

Declining retention

Rejection by automated tools

ClueBot NG
Huggle
… and others
TL;DR:

Automation made quality control scale.
TL;DR:

Automation made quality control scale.

It’s also scaring the good newbies away.
What do?

- We need good newcomers to stick around
- Automated tools are necessary to preserve quality.
Part 3: What can we do?
Part 3: What can we do?

Efficient vandal detection & removal
Part 3: What can we do?

huggle

Efficient vandal detection & removal

Snuggle

Efficient good newcomer detection & support
Detecting good newcomers with vandal bots
Recent activity

<table>
<thead>
<tr>
<th>Time</th>
<th>User</th>
<th>Activity</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>12:46, 8 Aug 2013</td>
<td>Steven Zhang</td>
<td>Categorized Bluedudemi as good-faith</td>
<td></td>
</tr>
<tr>
<td>23:11, 7 Aug 2013</td>
<td>King jakob c 2</td>
<td>Categorized Arturo13 as good-faith</td>
<td></td>
</tr>
<tr>
<td>02:24, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Invited Jeliza to the teahouse</td>
<td></td>
</tr>
<tr>
<td>02:23, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Categorized Jeliza as good-faith</td>
<td></td>
</tr>
<tr>
<td>02:22, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Categorized JulesMarshall as bad-faith</td>
<td></td>
</tr>
<tr>
<td>02:21, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Invited Suvi Murugan to the teahouse</td>
<td></td>
</tr>
<tr>
<td>02:20, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Categorized Suvi Murugan as good-faith</td>
<td></td>
</tr>
<tr>
<td>02:19, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Invited Mumtaz muhammed to the teahouse</td>
<td></td>
</tr>
<tr>
<td>02:18, 6 Aug 2013</td>
<td>Jethrobot</td>
<td>Categorized Mumtaz muhammed as good-faith</td>
<td></td>
</tr>
<tr>
<td>17:46, 5 Aug 2013</td>
<td>EpochFail</td>
<td>Categorized YorktonFilm as good-faith</td>
<td></td>
</tr>
</tbody>
</table>

Welcome!

Snuggle is a user interface designed to help experienced Wikipedians find good newcomers to support and mentor in Wikipedia.

See also
- Documentation
- Work log
- Code repository
Thanks!

Props to my collaborators

- Stuart Geiger
- Jonathan Morgan
- John Riedl

ahalfaker@wikimedia.org

Send me emails about science, Wikipedia, the internet or whatever.
Ping me (halfak) in #wikimedia-research on freenode