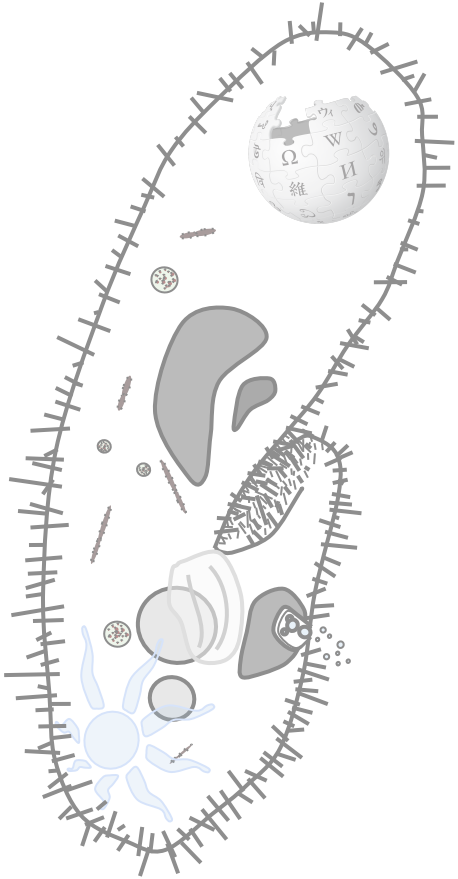
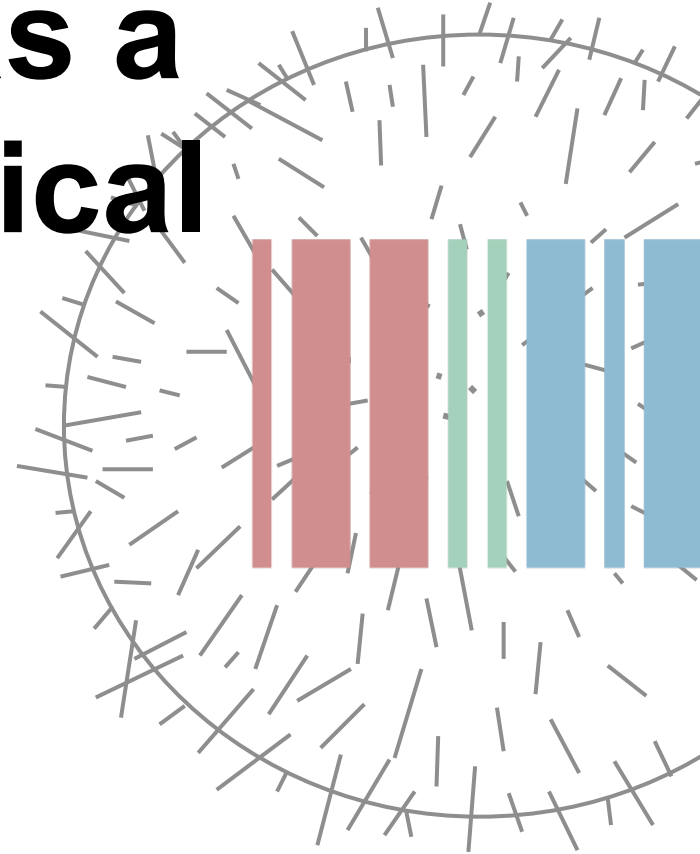


Wikipedia as a socio-technical system

by Aaron Halfaker



/who

Aaron Halfaker

Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

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".

/who

Aaron Halfaker

Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

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".

/who

Aaron Halfaker

Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

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...

/who

Aaron Halfaker

Research Scientist, Wikimedia Foundation

Think big. Measure what you can.

Think big. Measure what you can.

Today

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

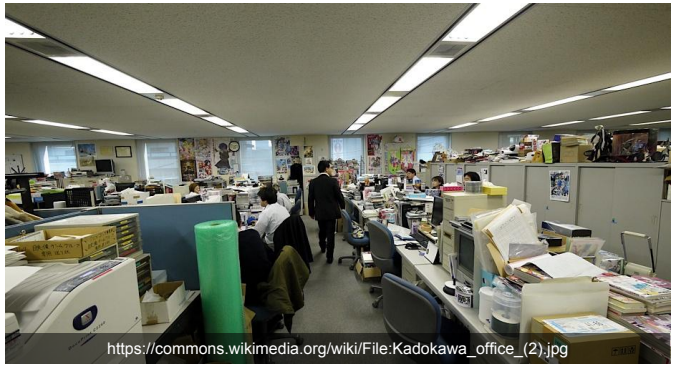
1. The “-” in socio-technical

2. Open problems in open collaboration
   (for collaboration)

3. What’s next?

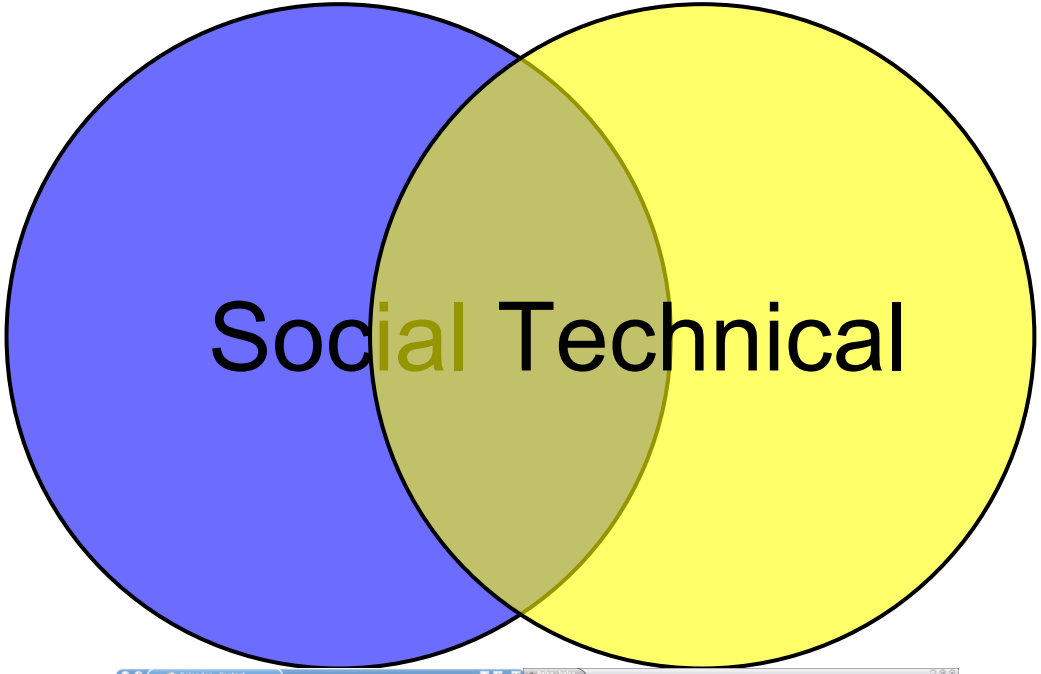
Social

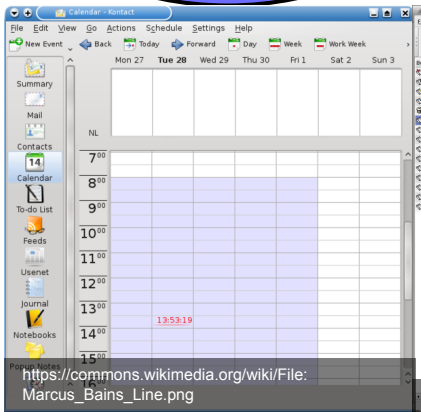
Technical



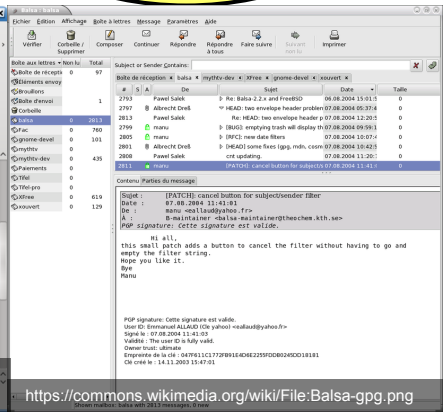
https://commons.wikimedia.org/wiki/File:Kadokawa_office_(2).jpg

https://commons.wikimedia.org/wiki/File:Commodore_PET_4016,_Google_NY_office_computer_museum.jpg

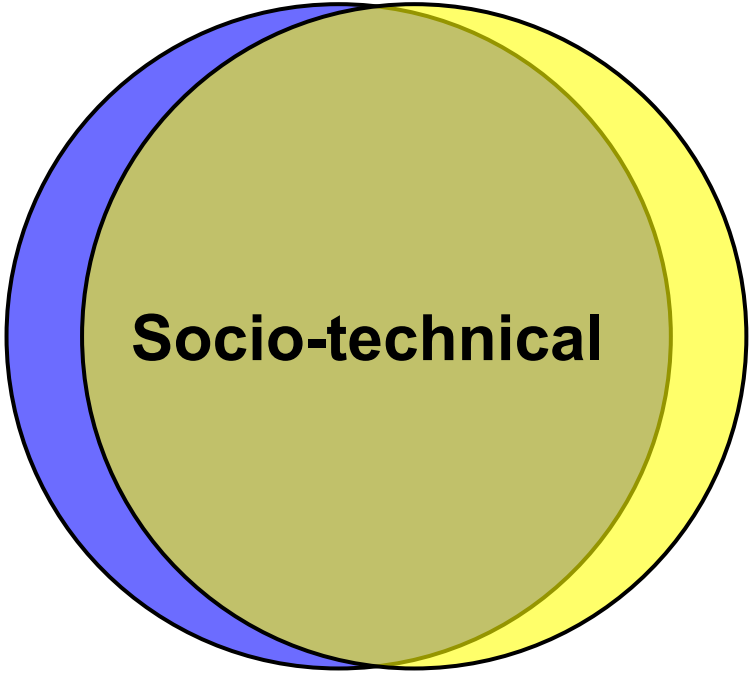


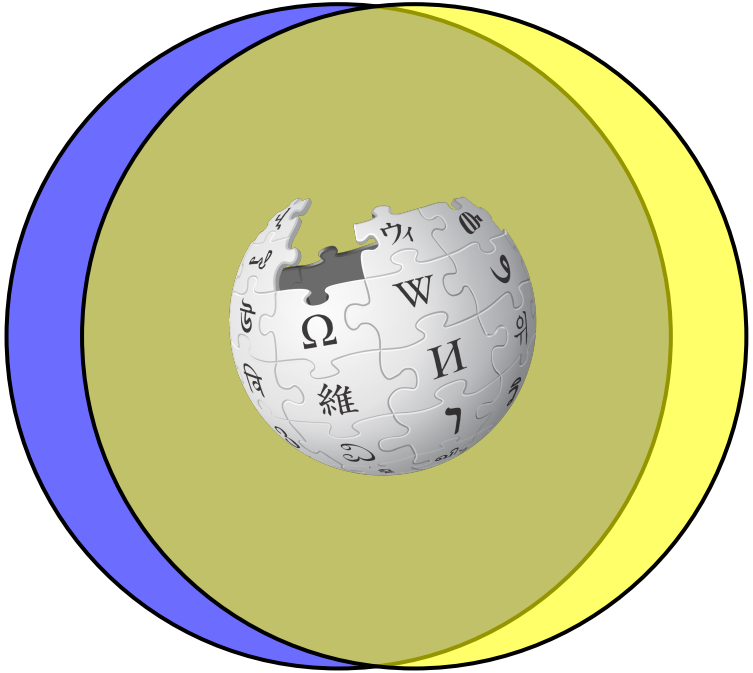


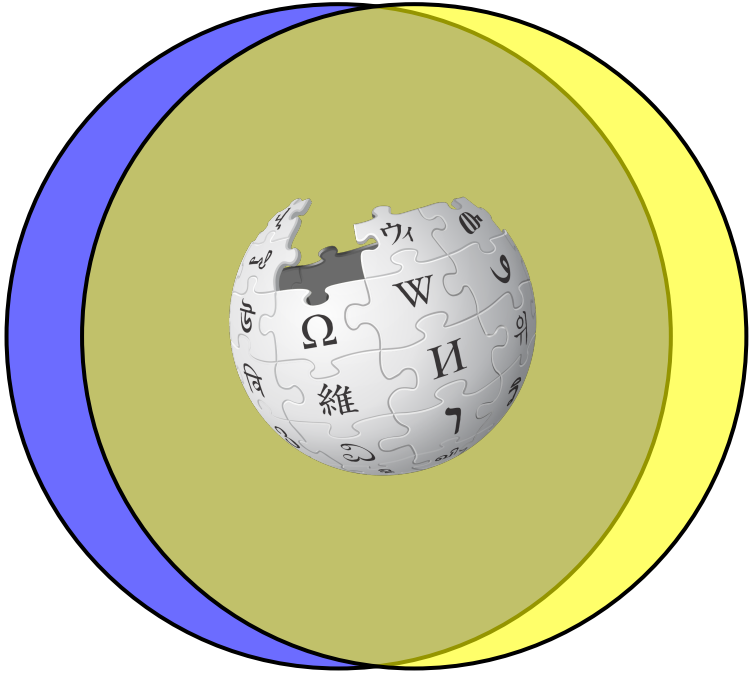
https://commons.wikimedia.org/wiki/File:Marcus_Bains_Line.png



https://commons.wikimedia.org/wiki/File:Balsa-gpg.png


Socio-technical





reddit

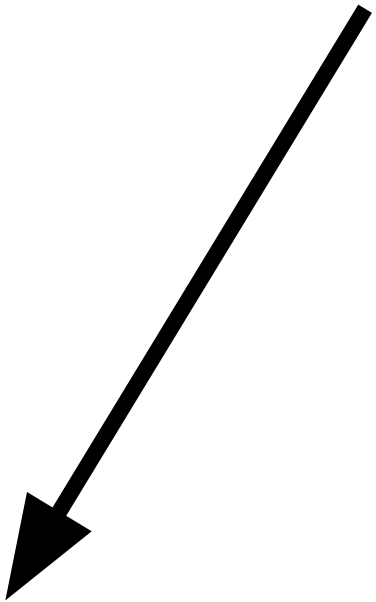
imgur

ZOONIVERSE

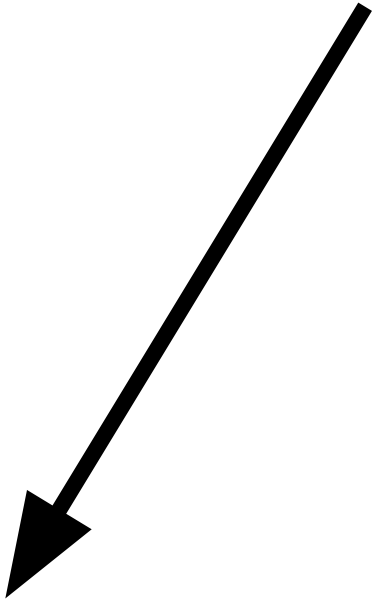
OpenIDEO

GitHub

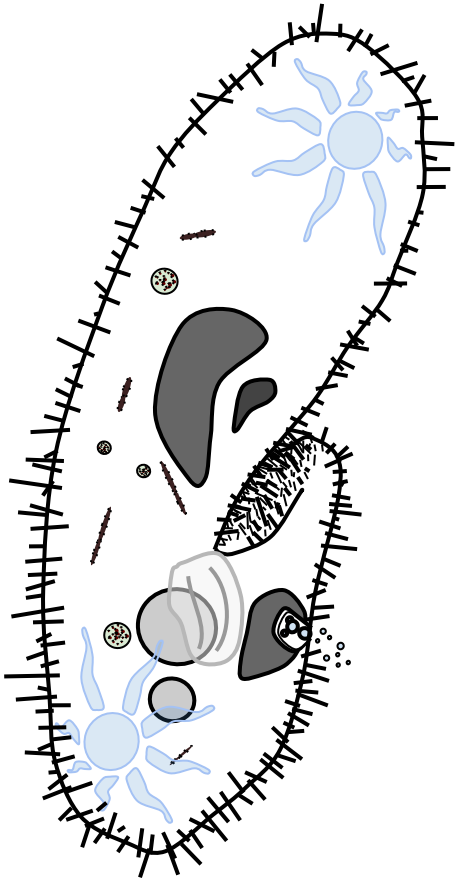
Technologist



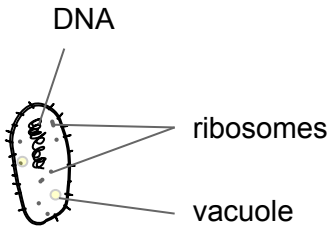
Techno-biologist?



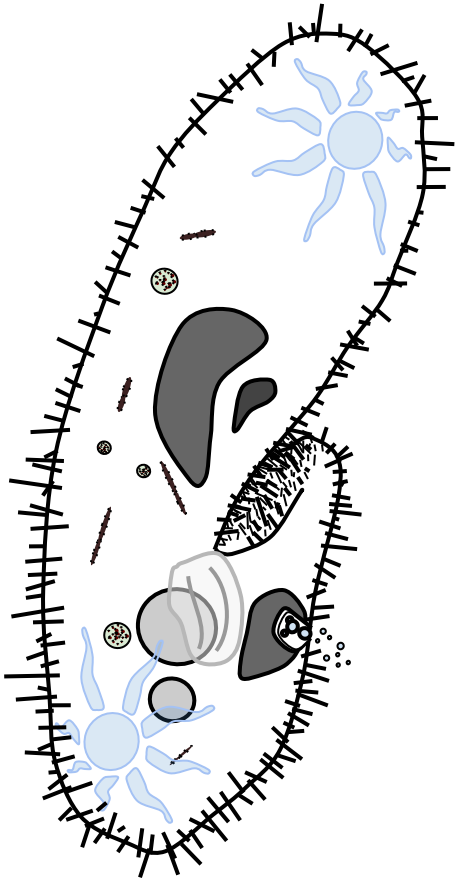
Bacterium



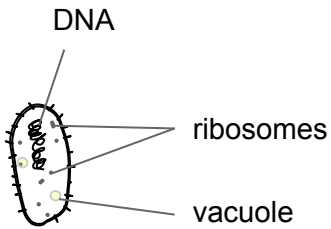
Paramecium



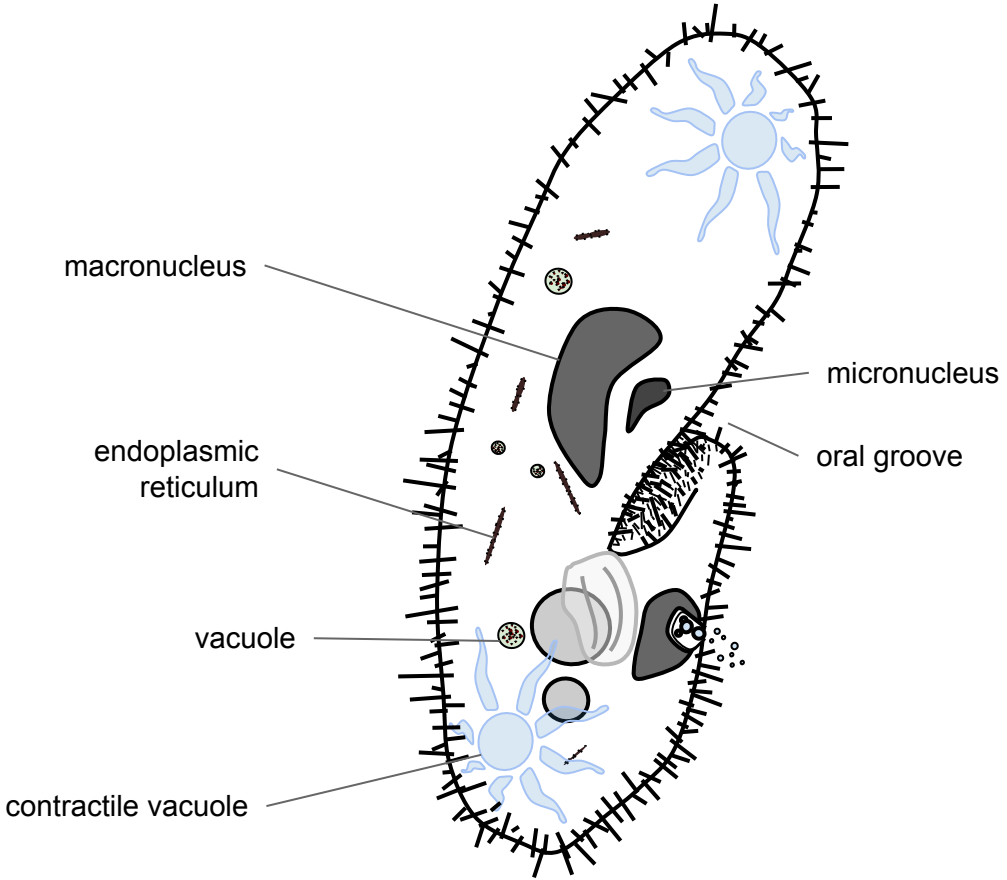
Bacterium



Paramecium



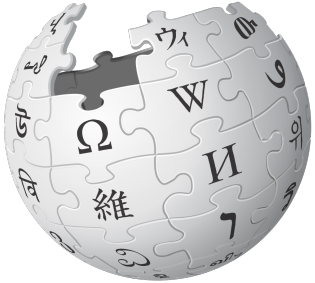
Bacterium



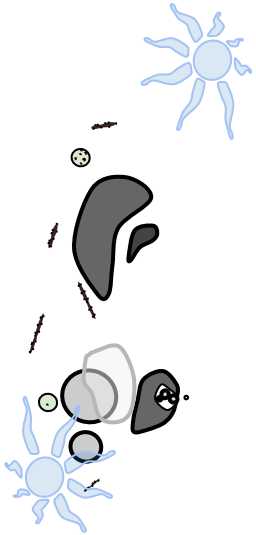
Paramecium

Bacterium

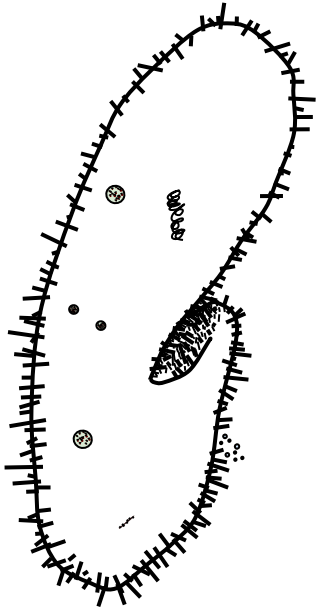
Paramecium



Bacterium



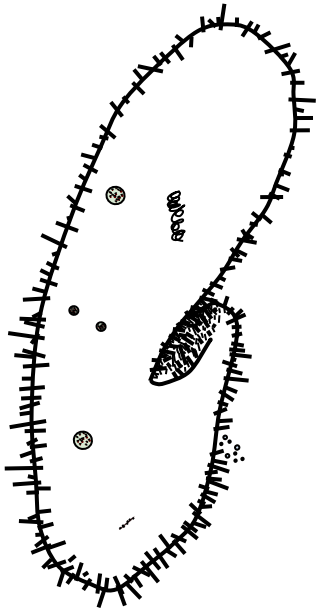
Spare parts



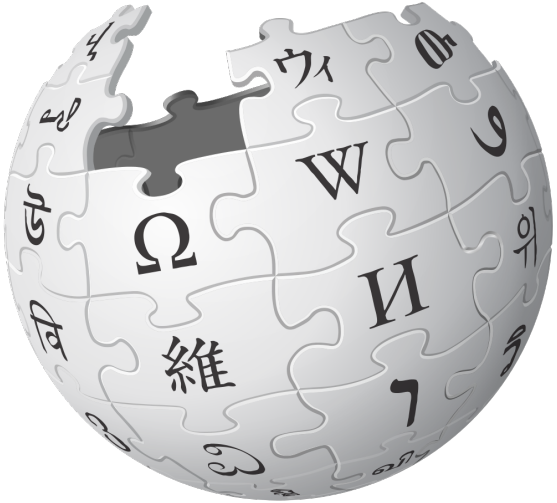
Bare-amecium

Paramecium

socio - technical



~

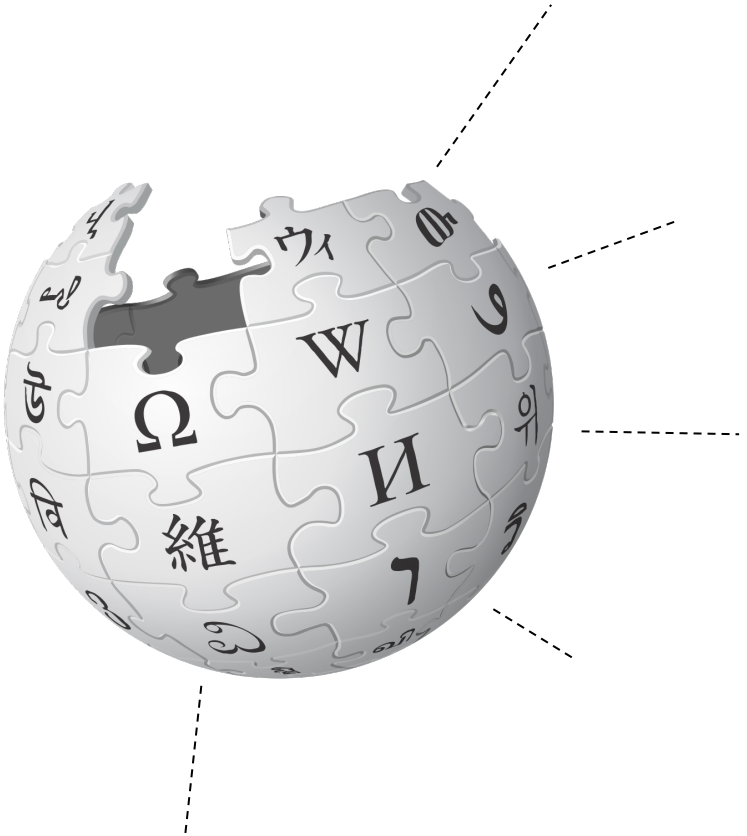


System with specialized
sub-systems

System with specialized
sub-systems

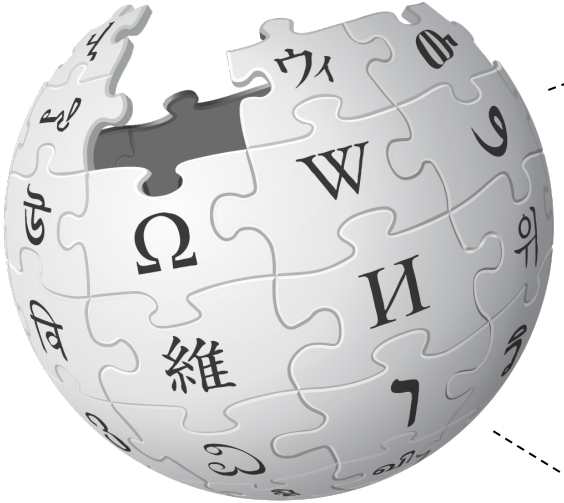
Part 2

Wikipedia’s specialized sub-systems



Work allocation

- Identify, prioritize and assign tasks

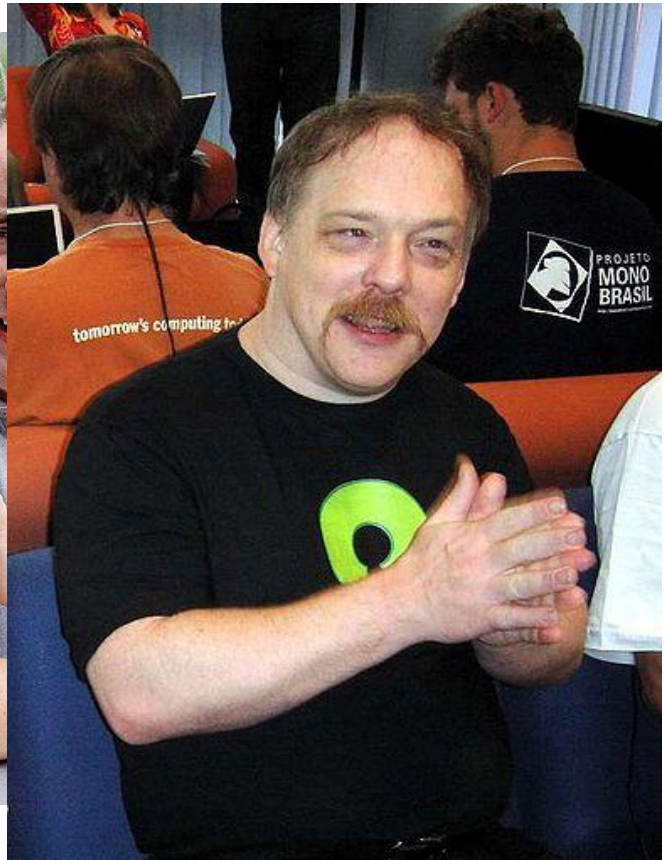


Work allocation

- 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?

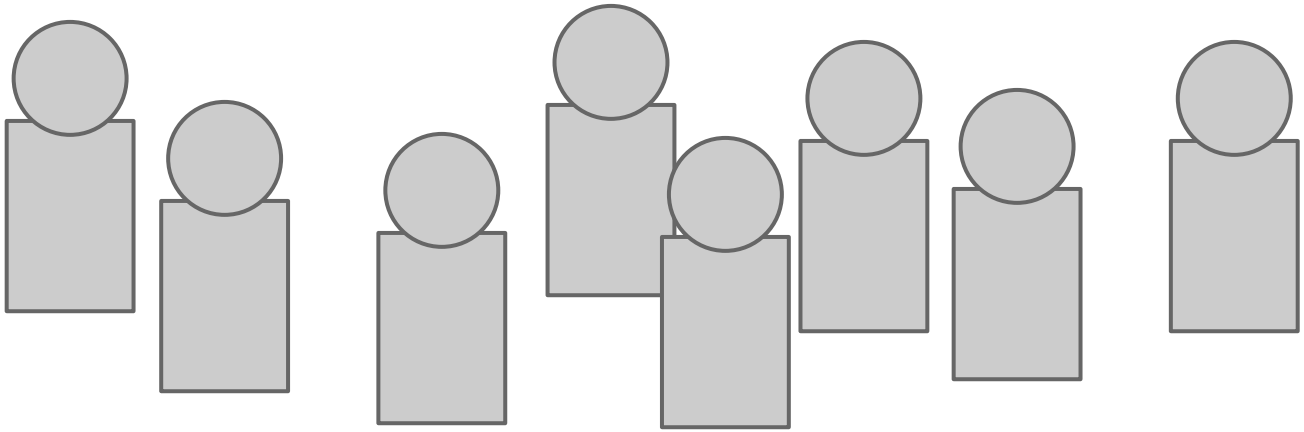
“given enough eyeballs, all bugs are shallow”

yessss…

bugssss….

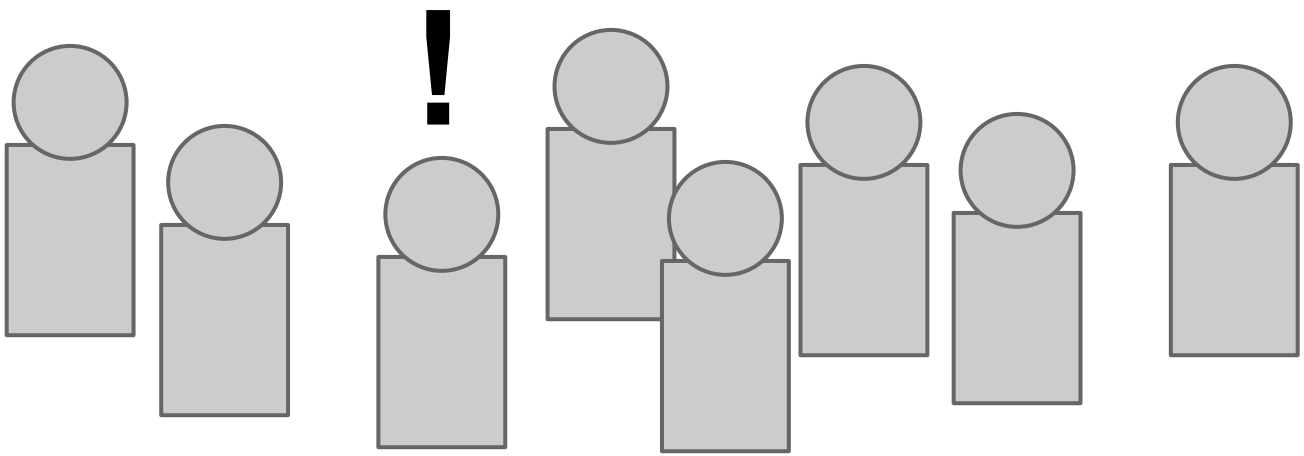
Linus's Law?

"given enough eyeballs, all bugs are shallow"



Linus's Law?

"given enough eyeballs, all bugs are shallow"

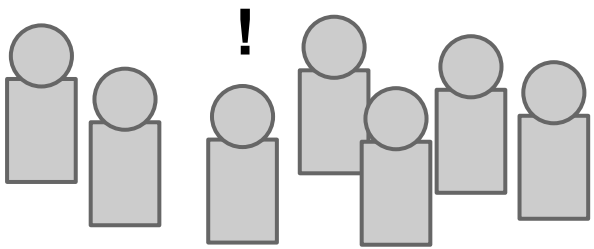


Linus’s Law?

visibility is critical to **open** collaboration

A corollary

Given that enough people see an incomplete article, all potential contributions to that article will be <u>easy for *someone*</u>.



Works in theory -- Stvilia, B., et al. (2008). Information quality work organization in Wikipedia. *JASIST*, *59*(6), 983-1001.

Part of becoming a Wikipedian -- Bryant, S. L., Forte, A., & Bruckman, A. (2005, November). Becoming Wikipedian. GROUP (pp. 1-10). ACM.

Supported with technology -- Cosley, D., et al. (2007, January). SuggestBot: using intelligent task routing to help people find work in wikipedia. IUI (pp. 32-41). ACM.

Bad things happen when we take it away -- Schneider, J., et al. (2014, August). Accept, decline, postpone. OpenSym(p. 26). ACM.

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.

User:SuggestBot

From Wikipedia, the free encyclopedia

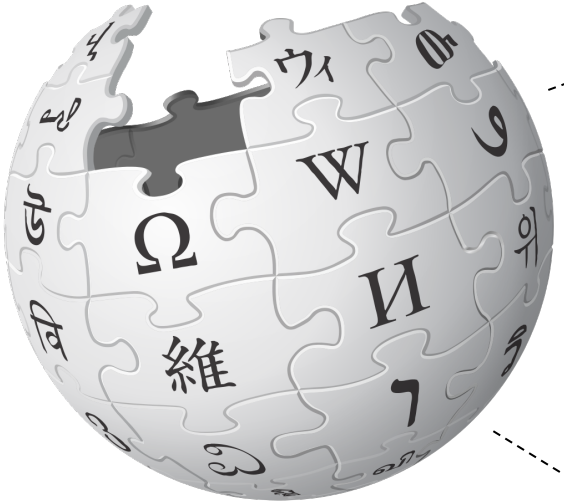
This user account is alternative account, us edits that would be ext and currently active – t

SuggestBot is a program that attempts to help Wik you want to try it out and get some personalized re You can also sign up to receive suggestions regula you can find examples here.

SuggestBot uses a variety of algorithms, including make suggestions. Most of it runs at the GroupLens Wikimedia Toolserver. The recommendations are u avoid recommending articles they've recently edite

If SuggestBot made personalized recommendation how to make them better. Comments on recommer are best left on our talk page. Comments are welco job of helping Wikipedia.

Views/Day	Quality	Title	Content	Headings	Images	Links	Sources	Tagged with...
46	★★★	Yamaha electric guitar models (talk)	X			X	X	Add sources
23	★★★	Fender Prodigy (talk)	X	X	X	X	X	Add sources
18	★★★	Guitarist (magazine) (talk)	X	X	X	X	X	Add sources
165	★★★	Fender Precision Bass (talk)	X	X			X	Add sources
10	★★★	Elite Stratocaster (talk)	X	X	X	X	X	Add sources
21	★★★	Yngwie Malmsteen Stratocaster (talk)	X	X	X		X	Add sources
172	★★★	Tim Commerford (talk)	X				X	Cleanup
204	★★★	Echinopsis pachanoi (talk)	X				X	Cleanup
23	★★★	National String Instrument Corporation (talk)	X			X	X	Cleanup
3,039	★★★	Puerto Rico (talk)						Expand
19	★★★	Compression (physics) (talk)	X		X	X	X	Expand
38	★★★	The Experiment (talk)	X		X	X	X	Expand
113	★★★	Shred guitar (talk)	X		X		X	Unencyclopaedic
189	★★★	Rickenbacker (talk)					X	Unencyclopaedic



Work allocation

- Identify, prioritize and assign tasks

Regulation of behavior

- Norm formation, propagation and enforcement

Regulation of behavior

A **norm** is a group-held belief about how members should behave in a given context.[*citation needed*]

wikipedia governance

About 37,400 results (0.07 sec)

Decentralization in **Wikipedia governance** [PDF] from andreaforte.net
A Forte, V Larco, A Bruckman - Journal of Management Information ..., 2009 - ME Sharpe
How does" self-**governance**" happen in **Wikipedia**? Through in-depth interviews with 20 individuals who have held a variety of responsibilities in the English-language **Wikipedia**, we obtained rich descriptions of how various forces produce and regulate social structures **...**
Cited by 107 Related articles All 5 versions Cite Save

[CITATION] Unpacking **Wikipedia governance**: The emergence of a bureaucracy of peers?
A Aaltonen, GF Lanzara - LAEMOS. Buenos Aires. http://www. egosnet. org/jart/ ..., 2010
Cited by 11 Related articles Cite Save

[CITATION] Is something fundamentally wrong with **Wikipedia governance** processes?
M Bauwens - ... -fundamentally-wrong-with-**wikipedia-governance**- ..., 2008
Cited by 6 Related articles Cite Save

Governance, organization, and democracy on the Internet: The iron law and the evolution of **Wikipedia**
P Konieczny - Sociological Forum, 2009 - Wiley Online Library
This study examines whether the Iron Law of Oligarchy exists in **Wikipedia** by analyzing how a key policy of the website regarding verifiability evolved into its current form. The study describes the decision-making processes of **Wikipedia** and shows that there are many **...**

Coordination and beyond: social functions of groups in open content production [PDF] from andreaforte.net
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 [PDF] from 137.215.9.22
L Schoeman, DJ Fourie - Journal of Public Administration: ..., 2008 - reference.sabinet.co.za
... **Wikipedia**. 2008. **Governance**. Encyclopaedia **Wikipedia**. Available online: http://en.**wikipedia**.org/wiki/**governance** [Accessed 25 February 2008]. **Wikipedia**. 2008. Treaty. Encyclopaedia **Wikipedia**. Available online: http://en.**wikipedia**.org/wiki/treaty [Accessed 25 February 2008]. ...
Cited by 3 Related articles All 5 versions Cite Save

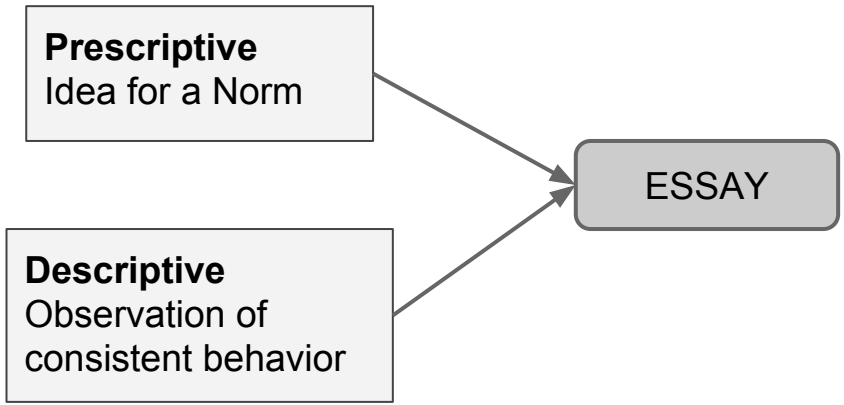
Changing **governance** of local economies: Responses of European local production systems
C Crouch, P Le Galès, C Trigilia, H Voelzkow - OUP Catalogue, 2004 - ideas.repec.org
... specialized manufacturing among small and medium-sized enterprises (SMEs) in France, Germany, Italy, and the United Kingdom, 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

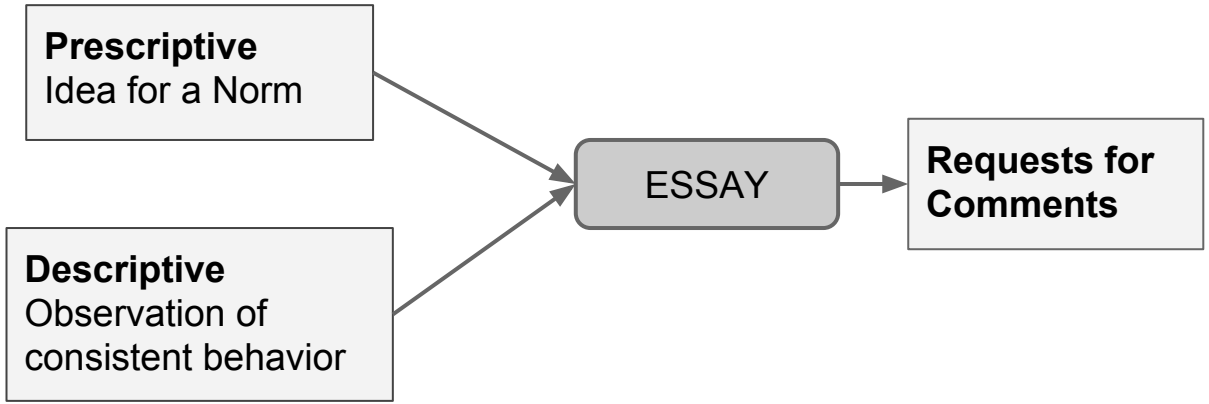
Foundations of service science management and business [PDF] from cluteonline.com
H Katzan - Journal of Service Science (JSS), 2011 - journals.cluteonline.com
... The development of a service context involves the asking of tough questions to examine the strategic goal and objectives of a service organization 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

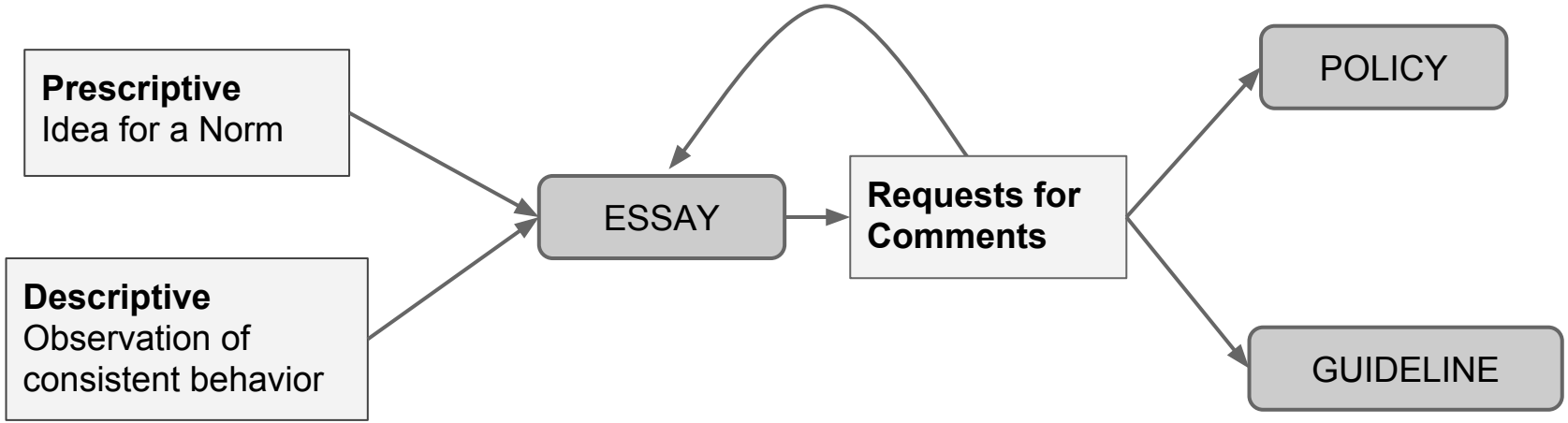
Building semantic kernels for text classification using **wikipedia** [PDF] from edudocs.ru
P Wang, C Domeniconi - Proceedings of the 14th ACM SIGKDD ..., 2008 - dl.acm.org
... Chief executive officer Shareholder Fiduciary Corporate **governance** Corporation Boards of directors Build Thesaurus from **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

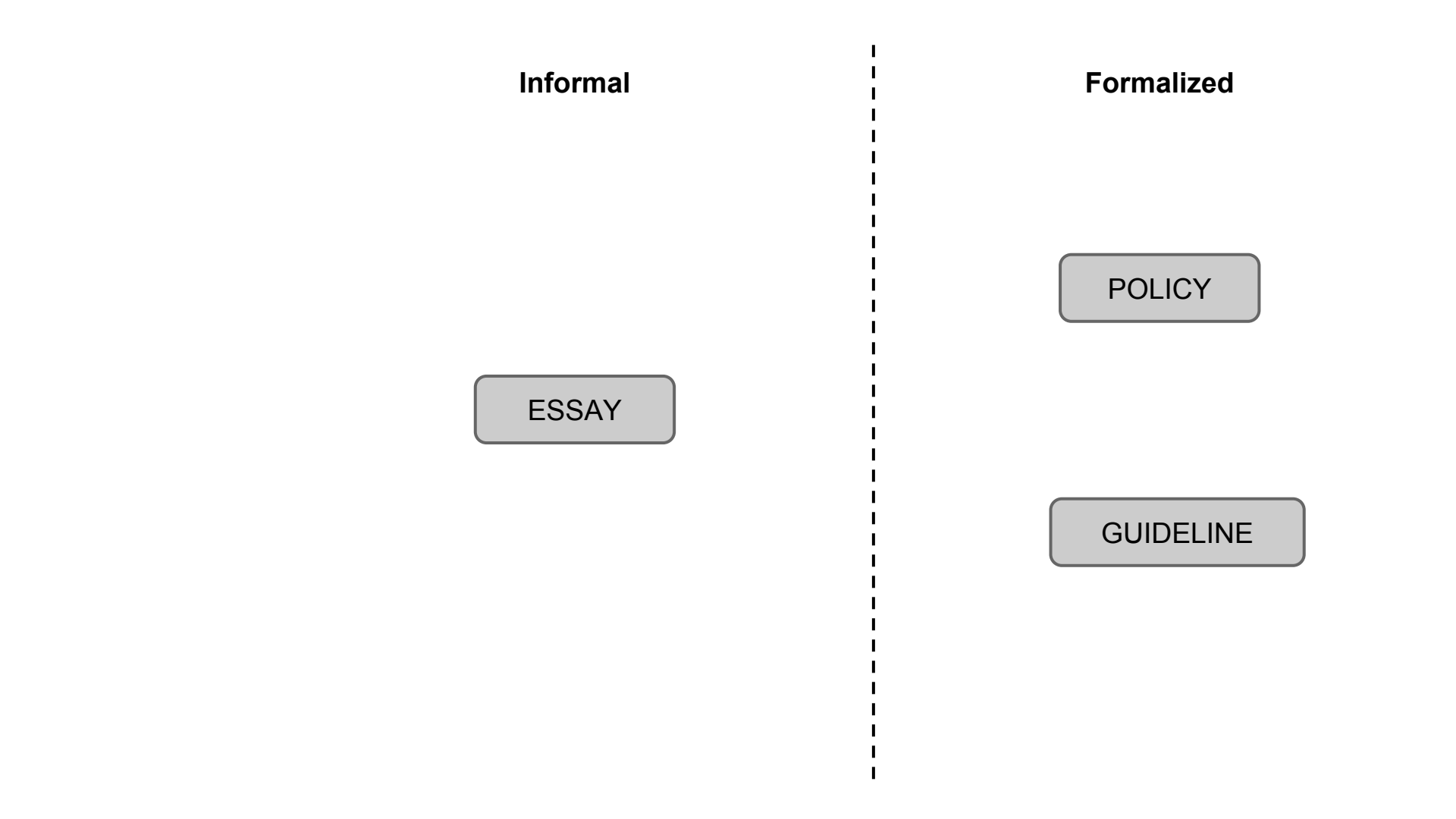
Prescriptive
Idea for a Norm

Descriptive
Observation of consistent behavior


Prescriptive
Idea for a Norm
Descriptive
Observation of consistent behavior
ESSAY


Prescriptive
Idea for a Norm
Descriptive
Observation of consistent behavior
ESSAY
Requests for Comments

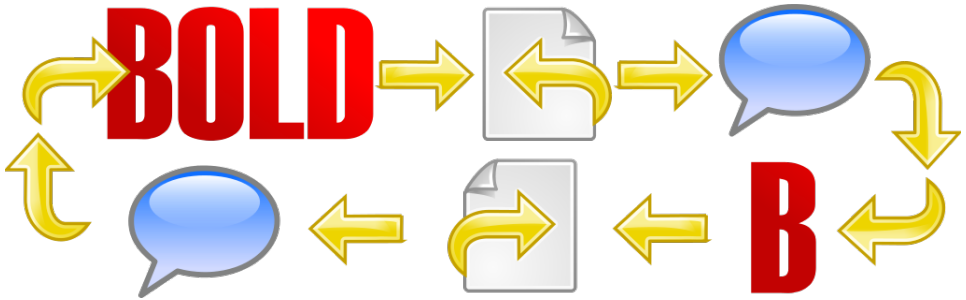

Prescriptive
Idea for a Norm
Descriptive
Observation of consistent behavior
ESSAY
Requests for Comments
POLICY
GUIDELINE


Informal
Formalized
POLICY
ESSAY
GUIDELINE

Informal

Wikipedia:No climbing the Reichstag dressed as Spider-Man

From Wikipedia, the free encyclopedia



Wikipedia:Don't stuff beans up your nose

From Wikipedia, the free encyclopedia

This essay contains the advice or opinions of one or more Wikipedia contributors on Wikipedia:Behavioral policy. Essays are *not* Wikipedia policies or guidelines. Some essays represent widespread norms; others only represent minority viewpoints.

Formalized

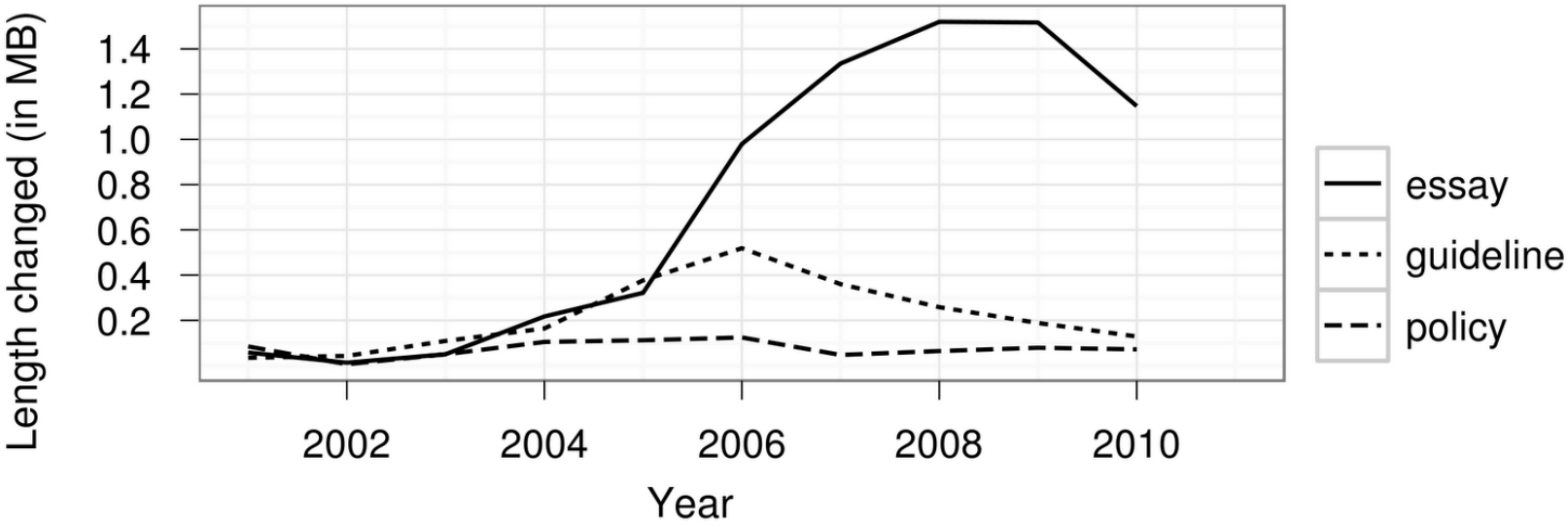
Wikipedia:Verifiability

From Wikipedia, the free encyclopedia

To discuss particular sources, see the reliable sources r
see WP:VAND. For the default Wikipedia skin, see WP:

This page documents an English Wikipedia policy, a widely accepted standard that all editors should normally follow. Changes made to it should reflect consensus.

Growth of regulations



Citation distribution

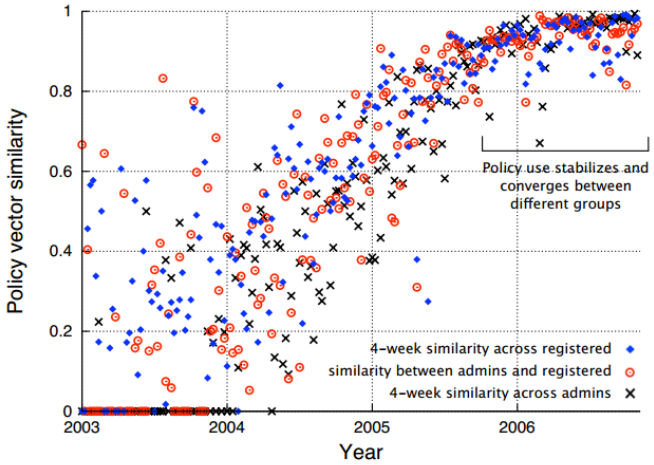


Figure 8: Policy vector similarities for admin and registered users are converging to 1, indicating that the two user groups are becoming similar in their policy citations. The convergence of similarities of policy vectors 4 weeks apart within each population indicates stability within user classes.

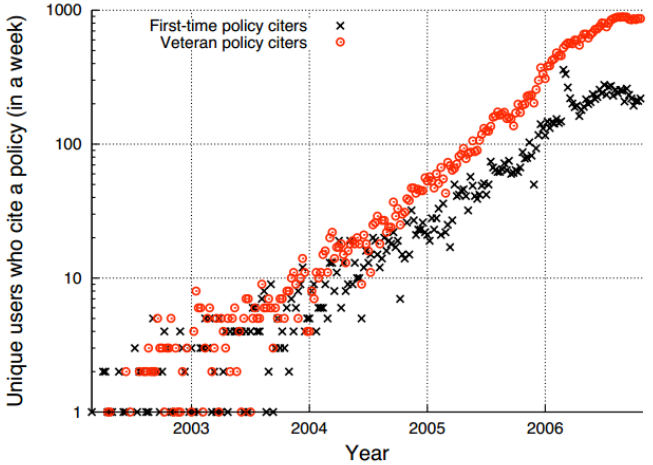
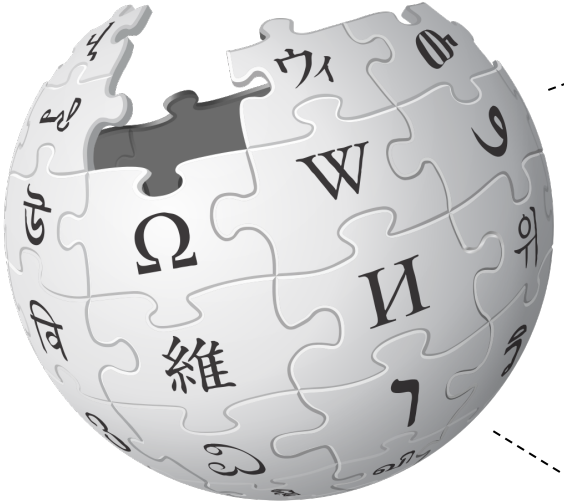


Figure 9: On average, unique users who cite policy for the first time account for over 10% of all unique users who cite a policy in a week. The high percentage indicates that participation in Wikipedia's governance structure is inclusionary.

Beschastnikh, I., Kriplean, T., & McDonald, D. W. (2008, March). **Wikipedian Self-Governance in Action**: Motivating the Policy Lens. In ICWSM.



Work allocation

- Identify, prioritize and assign tasks

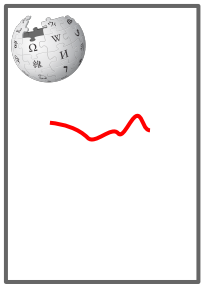
Regulation of behavior

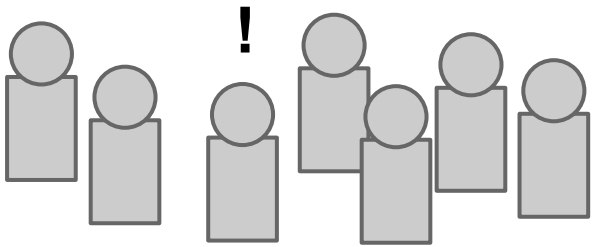
- Norm formation, propagation and enforcement

Quality control

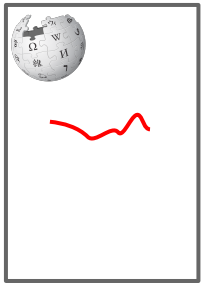
- Identify and remove damage

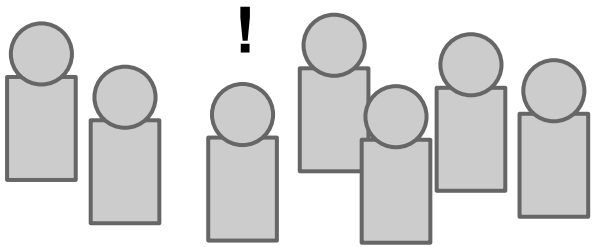
Quality control



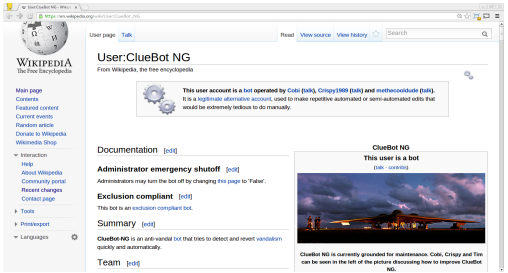


Quality control





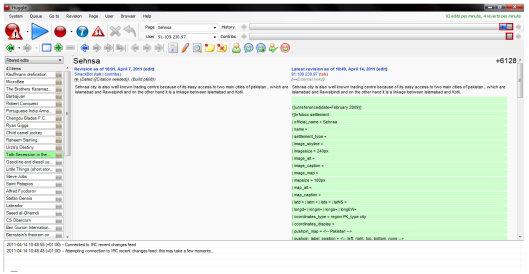
Fully automated



Machine learning

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

Semi automated



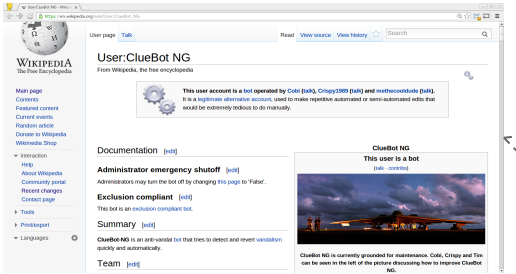
Human computation

- Still pretty fast (~ 30 seconds)[1]
- Minimizes human effort
- <u>Humans catch most vandalism at a glance</u>

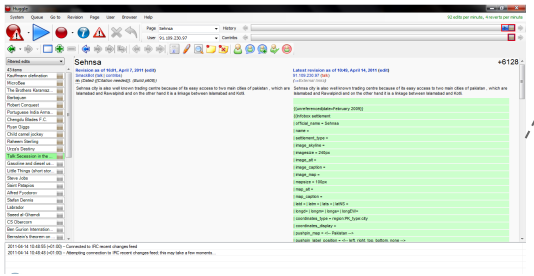
1. R. Stuart Geiger & Aaron Halfaker. **When the Levee Breaks** (2013). *WikiSym*.

Quality control

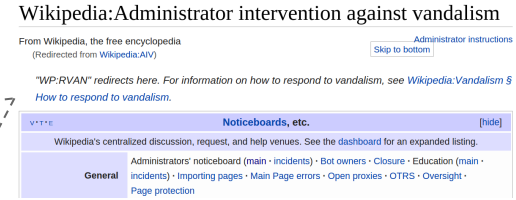
Fully automated

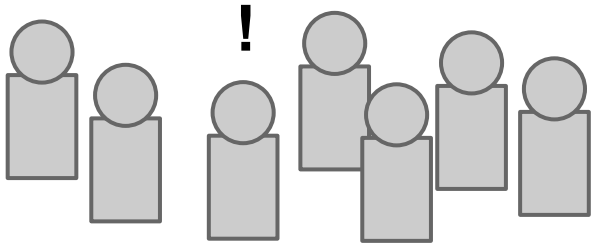


Semi automated



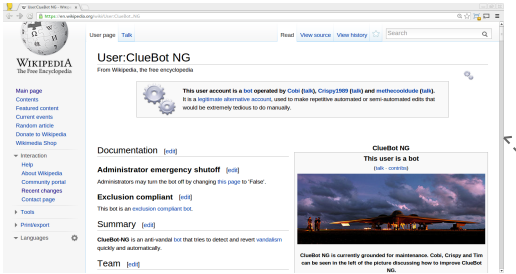
Banning (Admins)



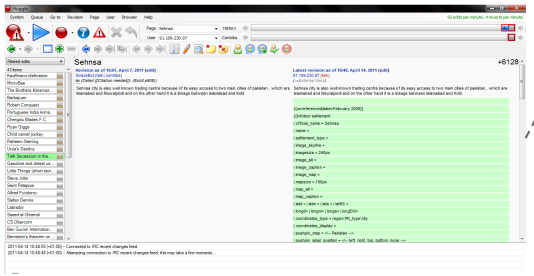


Quality control

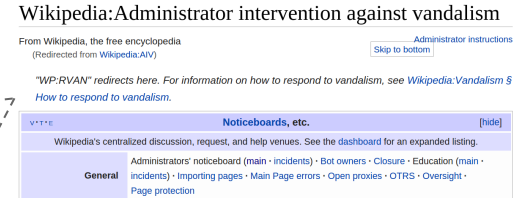
Fully automated

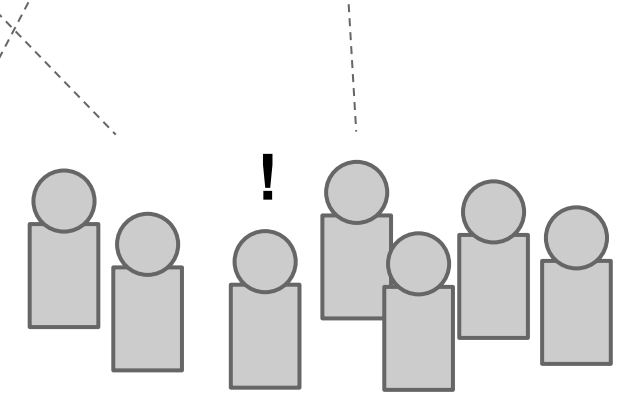


Semi automated



Banning (Admins)





Innate

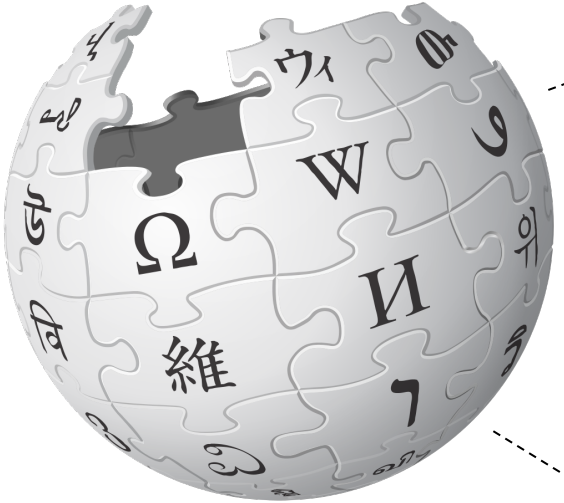
- Fast
- General
- Pervasive

Adaptive

- Slow
- Specific
- Pervasive

Halfaker, A. & Riedl, J. (2012) Bots and Cyborgs, *IEEE Computing* 45(3) (p. 79-82)

Geiger, R. S., & Ribes, D. (2010, February). The work of sustaining order in wikipedia: the banning of a vandal. CSCW (pp. 117-126). ACM.



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

Contents [show]

Key Personnel [edit]

Contact + Co-investigators and:

- Maryana Pinchuk
- Steven Walling
- Oliver Keyes

Project Overview [edit]

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.

Research project

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

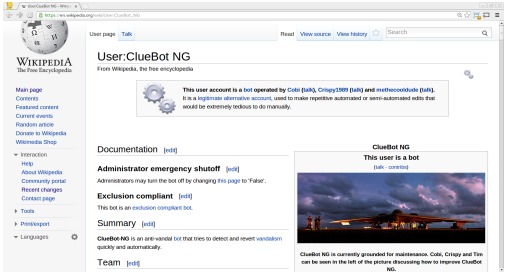
Main contact	**Aaron Halfaker** *University of Minnesota*
Co-investigators	**R Stuart Geiger** *University of California, Berkeley*
	Jonathan Morgan *University of Washington*
	John T. Riedl *University of Minnesota*
WMF contact	EpochFail (Aaron Halfaker)
Start	2011-05
End	2011-12

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

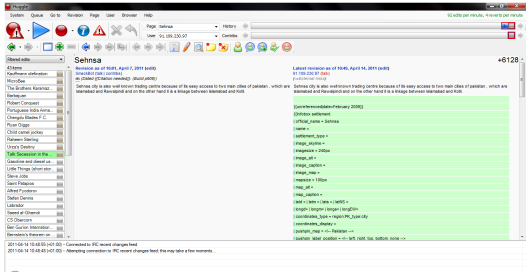
Halfaker, A., Geiger, R. S., Morgan, J. T., & Riedl, J. (2012). The rise and decline of an open collaboration system: How Wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, 0002764212469365.

TL;DR:

Fully automated

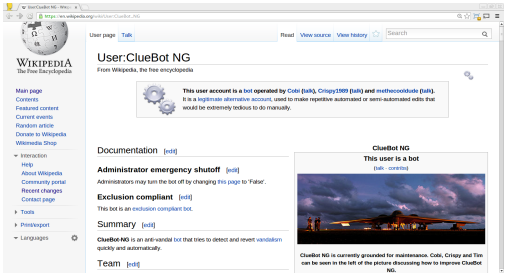


Semi automated

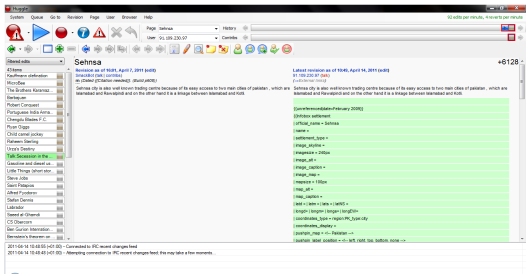


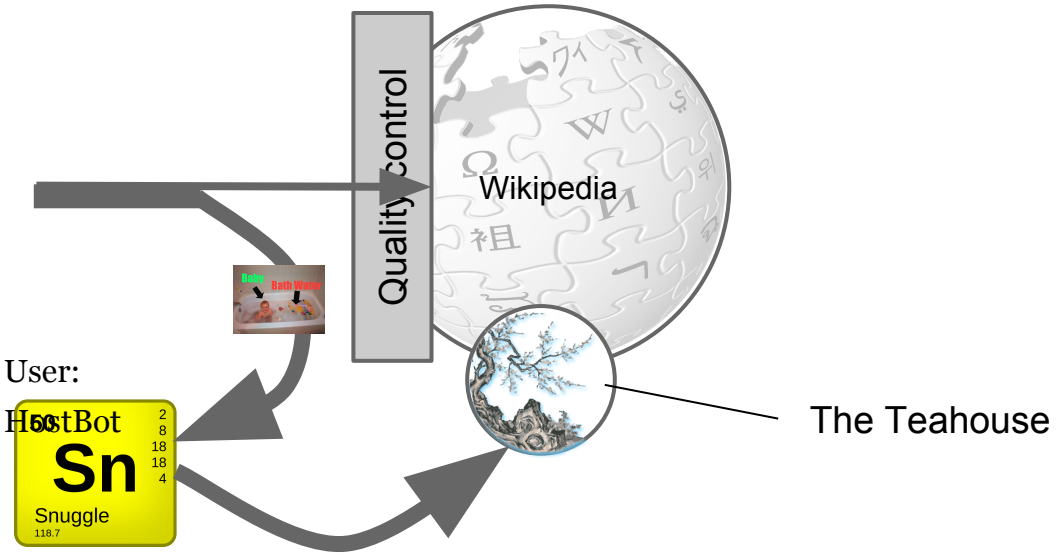
TL;DR:

Fully automated

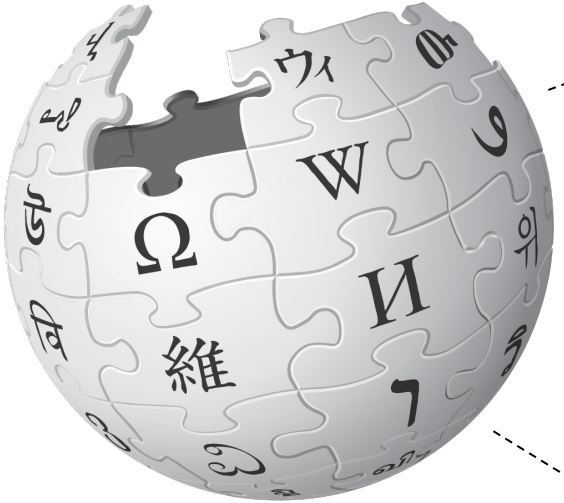


Semi automated





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)

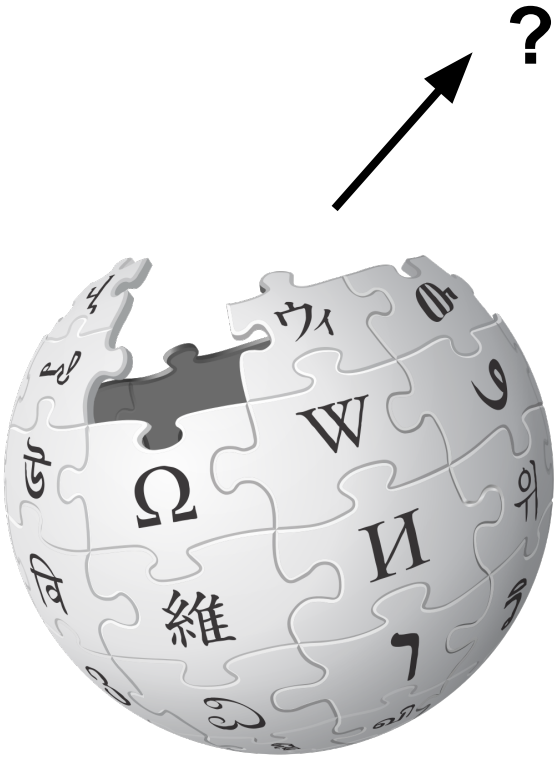
- Where are we going? Where do we want to go? How do we want to get there?

Reflection (Adaptation)

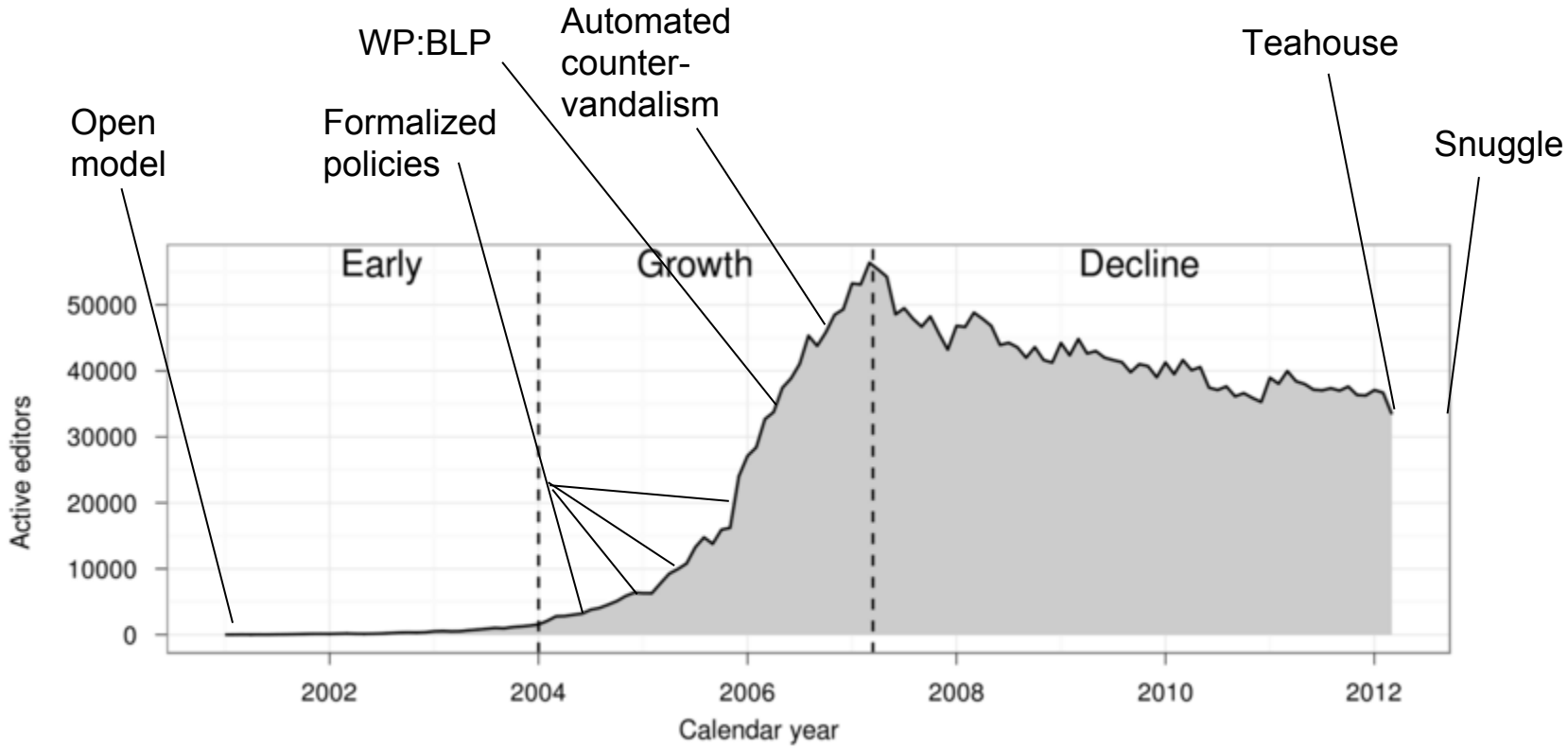
Where *are* we going?

Where *do* we want to go?

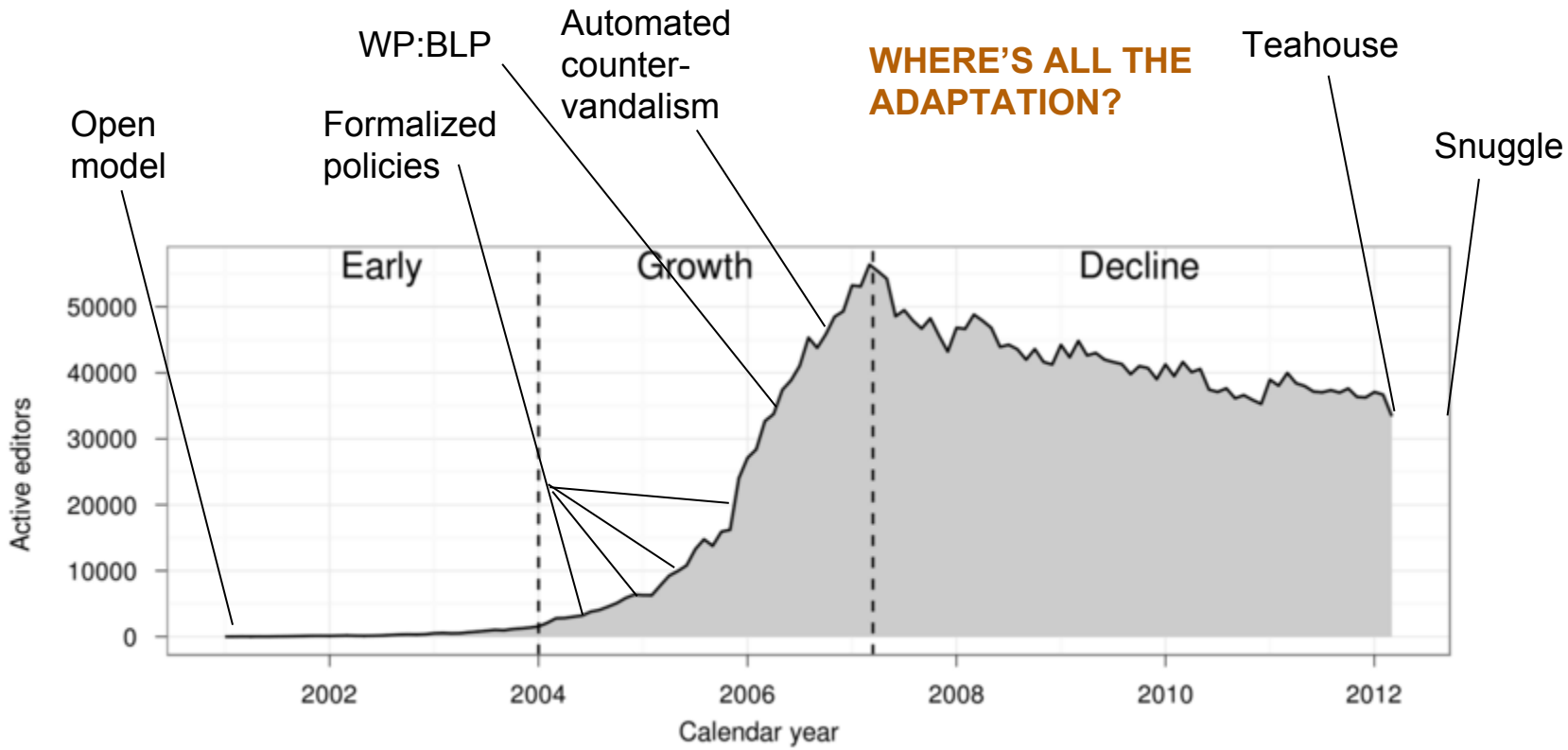
How do we get there?

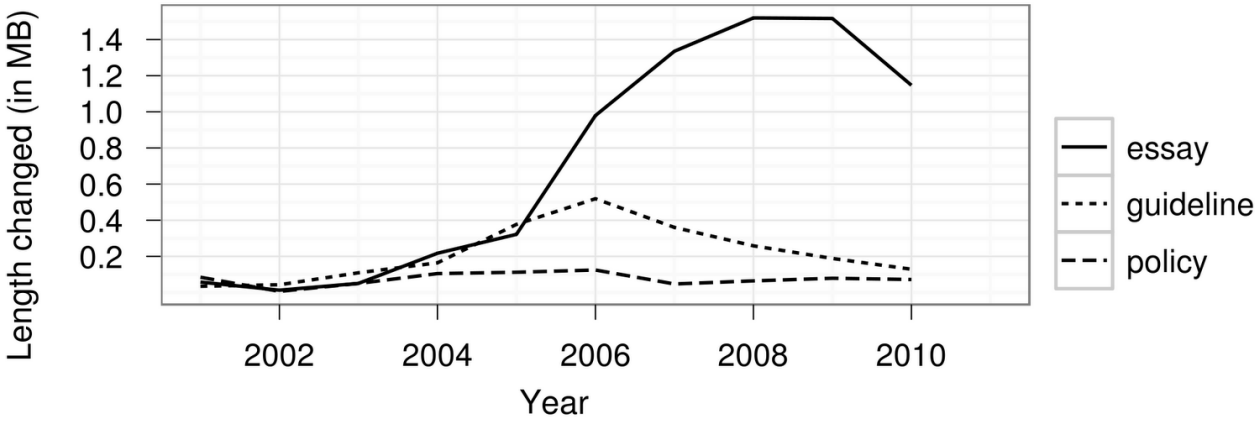


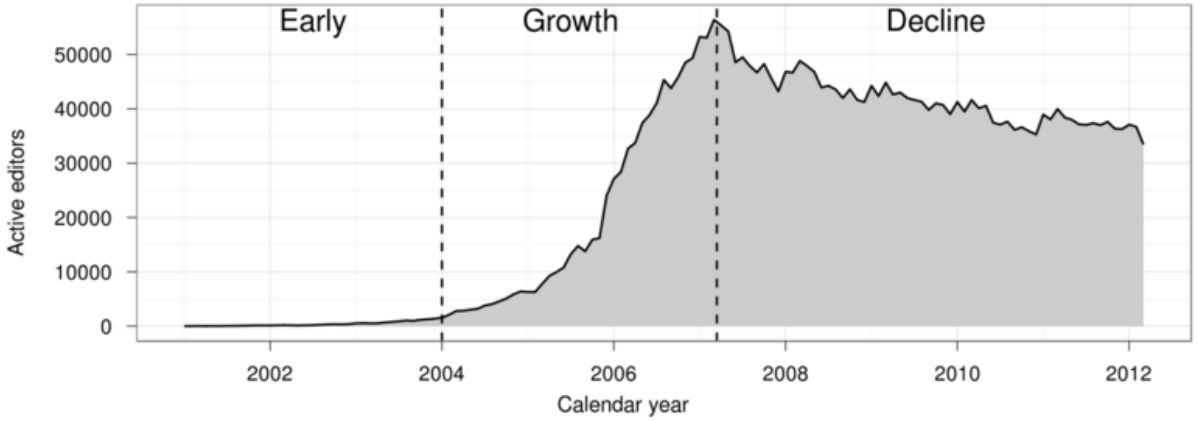
Reflection (Adaptation)

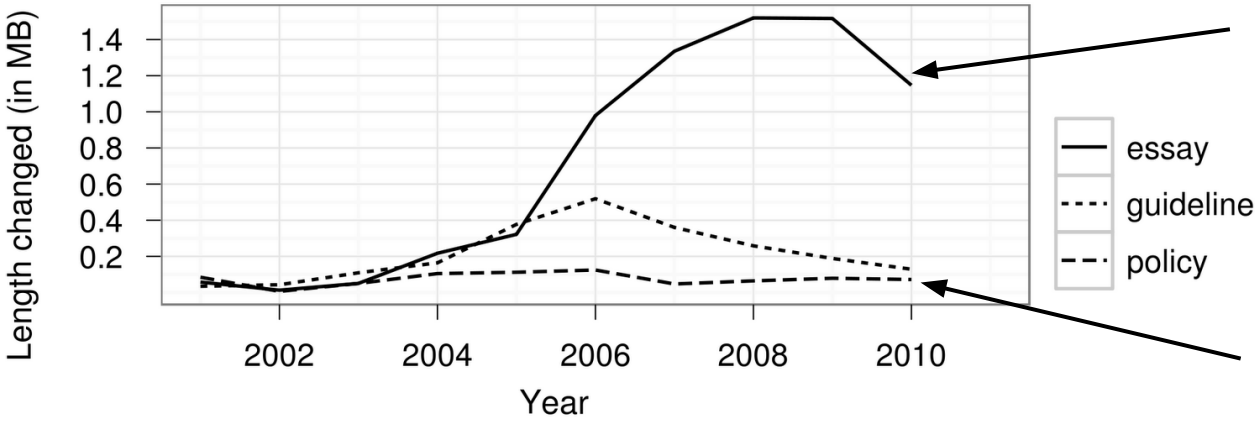


Reflection (Adaptation)



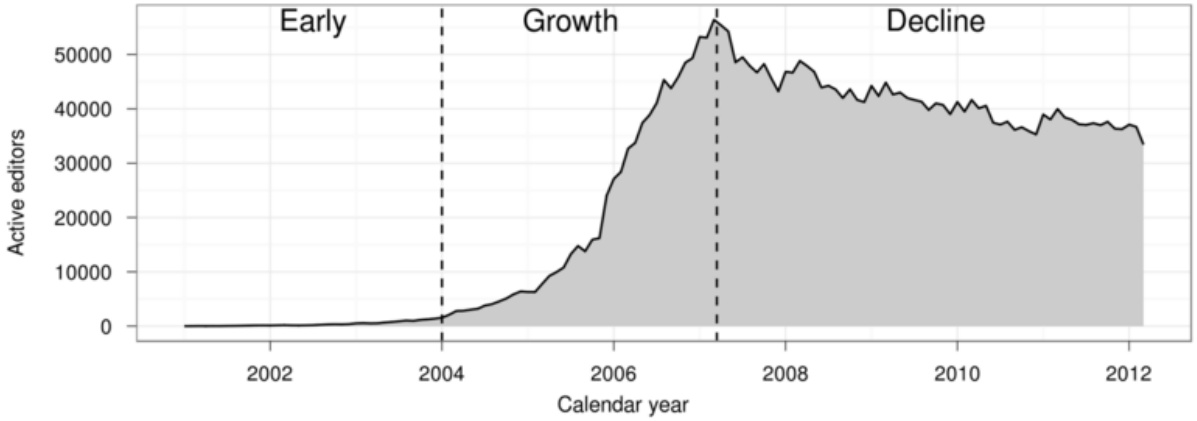

Length changed (in MB)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
2002
2004
2006
2008
2010
Year
essay
guideline
policy


Early
Growth
Decline
Active editors
50000
40000
30000
20000
10000
0
2002
2004
2006
2008
2010
2012
Calendar year



Reflection

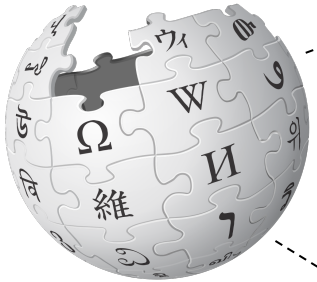
Adaptation



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?



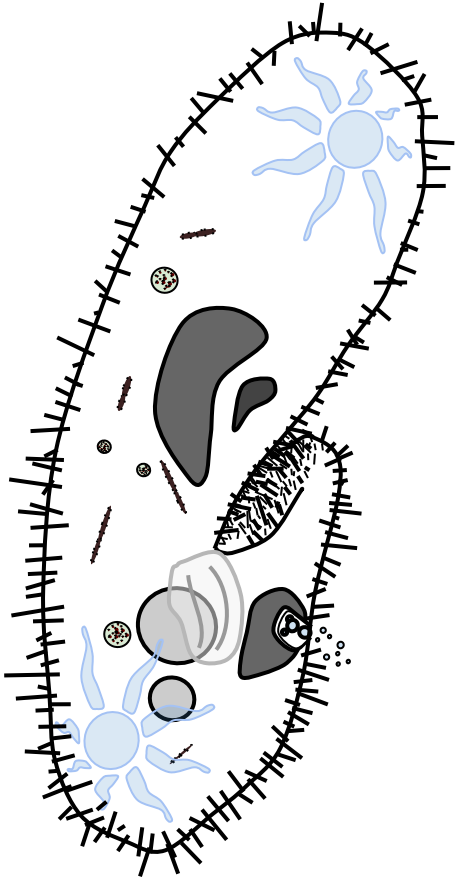
Work allocation

Regulation of behavior

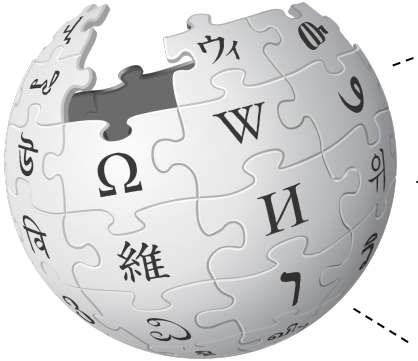
Quality control

Community management

Reflection (Adaptation)



OH YEAH!



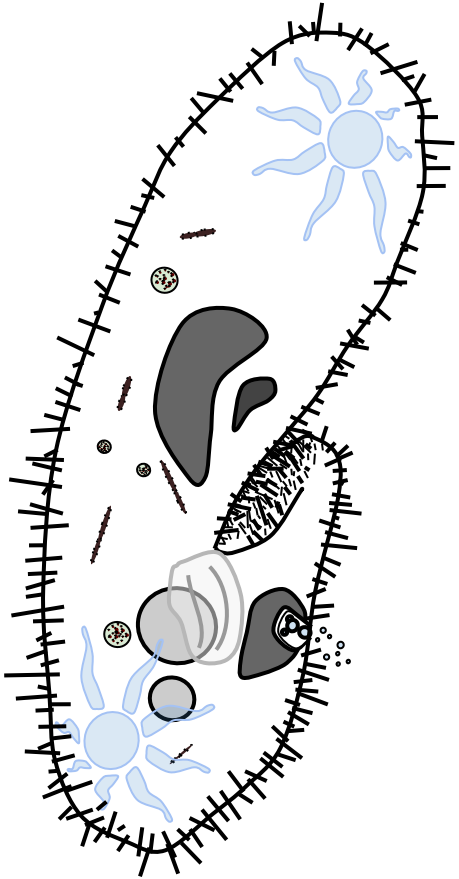
Work allocation

Regulation of behavior

Quality control

Community management

Reflection (Adaptation)



Part 3

What’s next?

Work allocation

WikiGrok

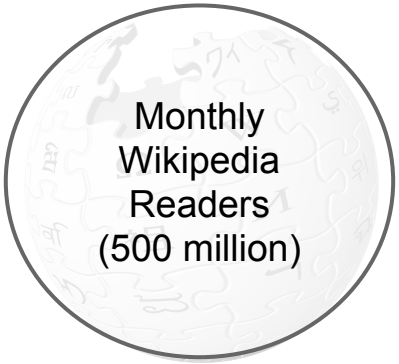
https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Work allocation

WikiGrok

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

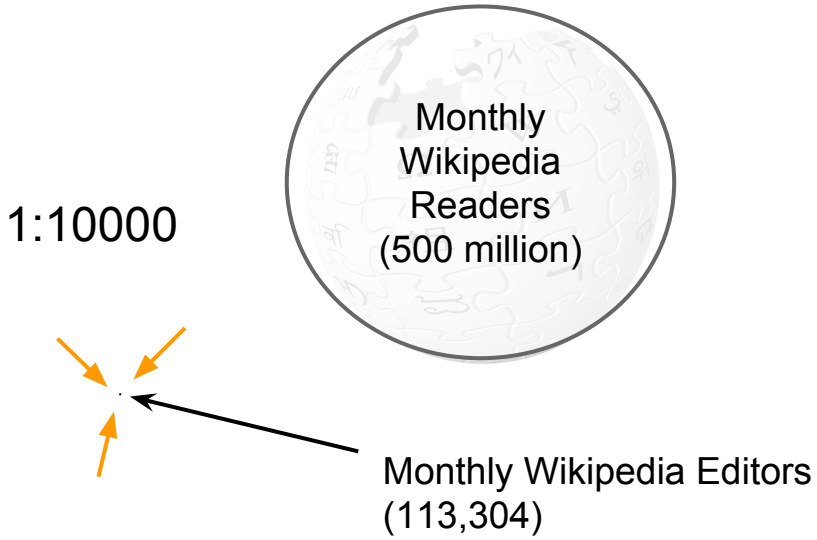
1:10000



Monthly Wikipedia Editors
(113,304)

Work allocation

WikiGrok



Work allocation

WikiGrok

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

Efficiency (overall)

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

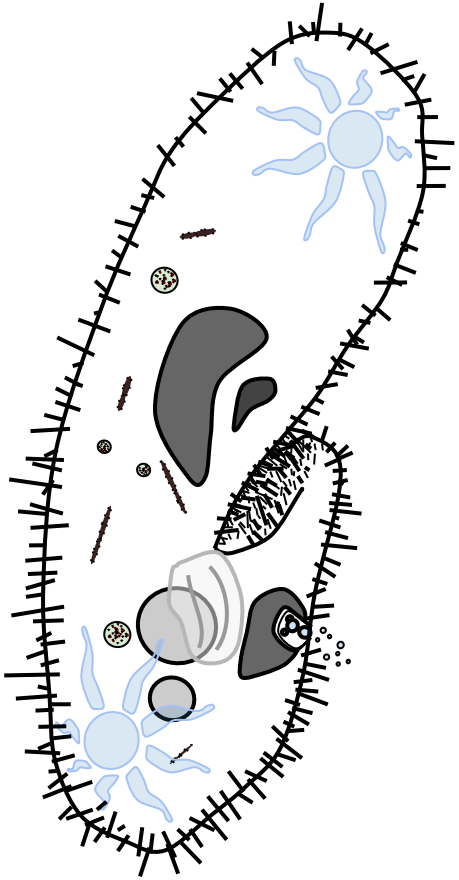
Work allocation

WikiGrok

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

Efficiency (overall)

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



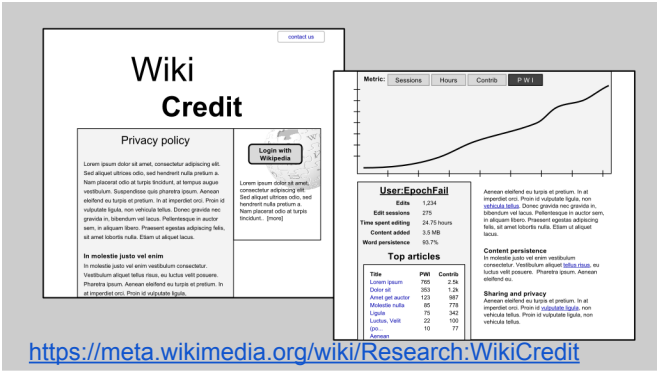
Work allocation

WikiGrok

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

Efficiency (overall)

Reflection



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

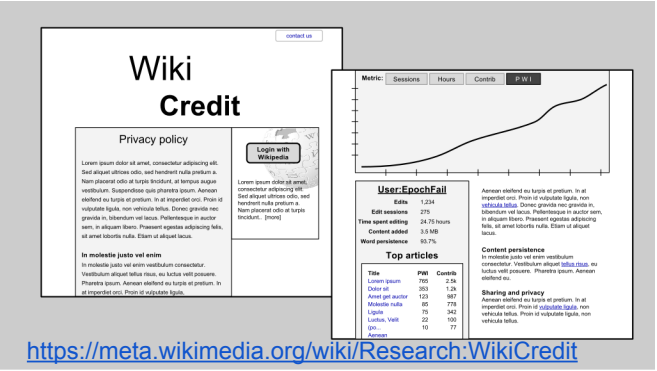
Work allocation

WikiGrok

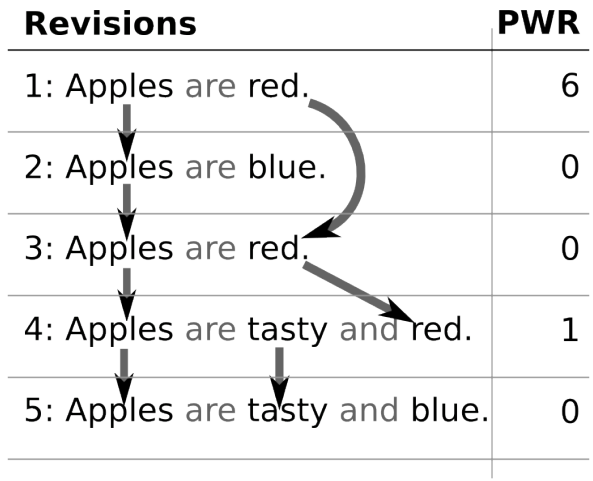
https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Efficiency (overall)

Reflection



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



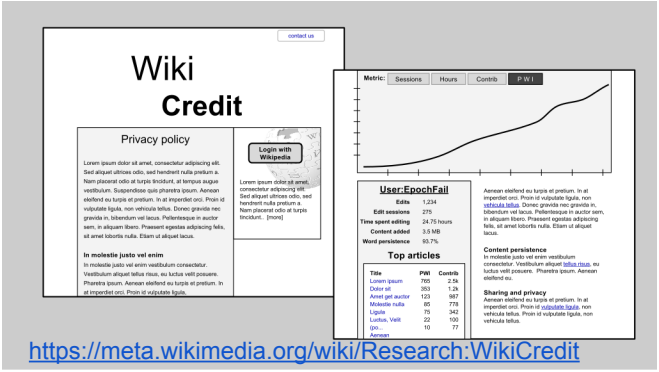
Work allocation

WikiGrok

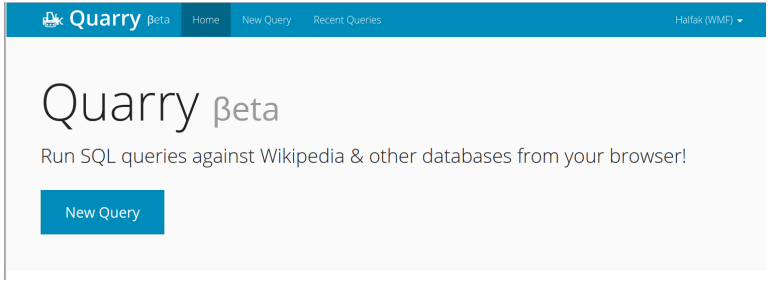
https://www.mediawiki.org/wiki/Extension:MobileFrontend/WikiGrok

Efficiency (overall)

Reflection

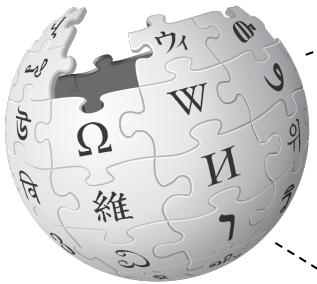


https://meta.wikimedia.org/wiki/Research:WikiCredit



Demo! http://quarry.wmflabs.org/

Thanks!



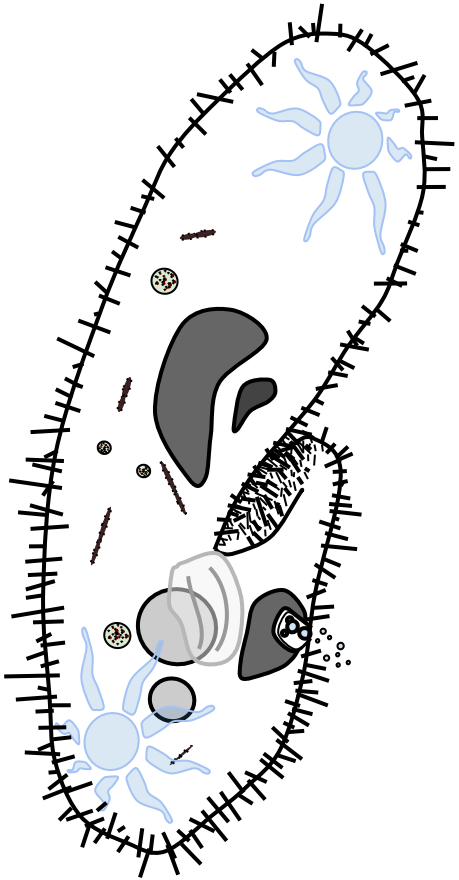
Work allocation

Regulation of behavior

Quality control

Community management

Reflection (Adaptation)



Aaron Halfaker (ahalfaker@wikimedia.org)

All original artwork licensed copyright Aaron Halfaker CC-BY-SA 3.0