Automoderator and Edit Patrol

Make fighting vandalism more easier

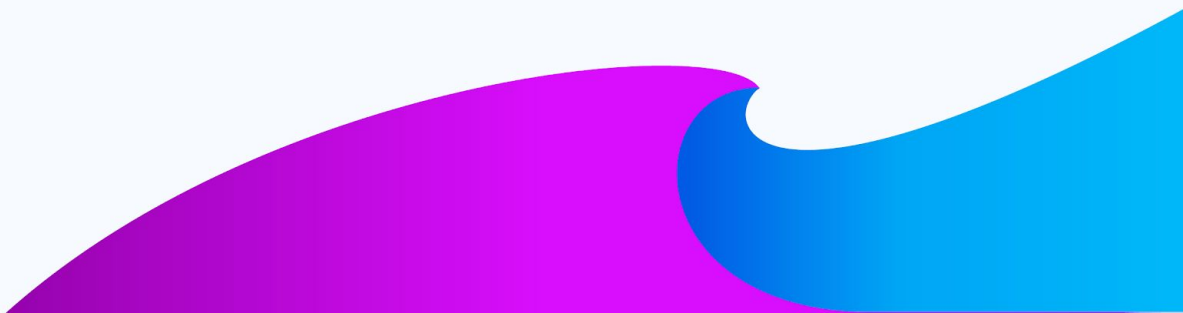
Bonaventura Aditya Perdana
baperdana-ctr@wikimedia.org

WIKIMEDIA ESEAP CONFERENCE
KOTA KINABALU 2024
COLLABORATION BEYOND THE HORIZON

Moderators

Users with extended rights, e.g.

- Stewards
- Functionaries
- Administrators
- Patrollers
- Rollbackers



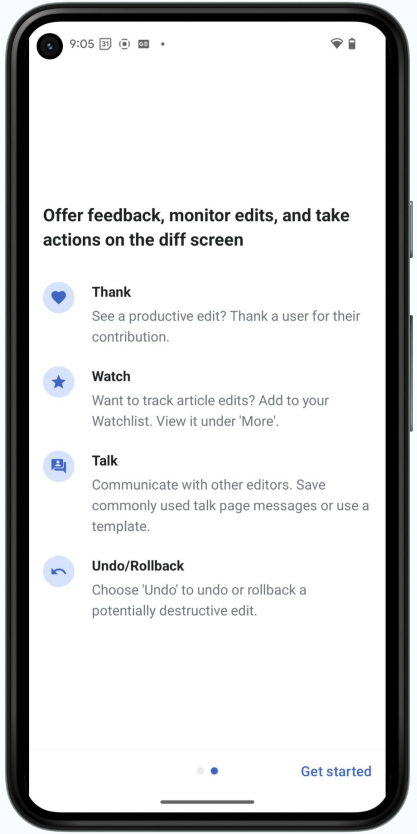
Edit Patrol

Edit Patrol

Patrol recent changes in the Android App

- This feature is part of the Wikimedia Foundation's Annual Plan to **Improve experience of editors with extended rights**
- The team noticed a theme of people wanting to take action on vandalism on the go.
- This feature hopes to increase moderator velocity across several language wikis, specifically when users are primarily using mobile devices.

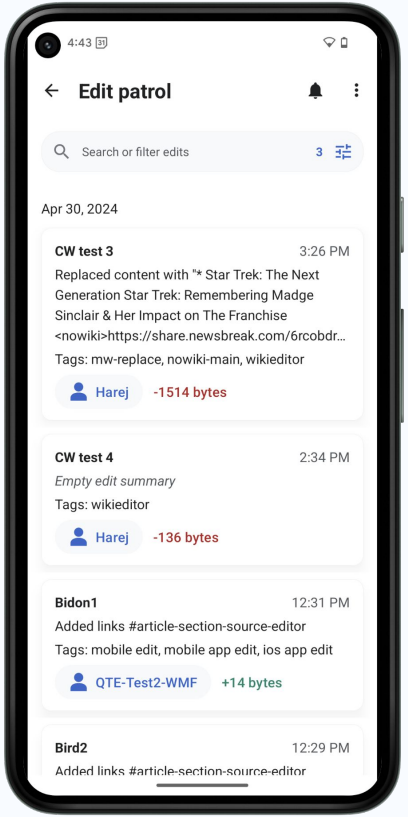
If you want to know more about how we approached this feature, please check out our presentation at Wikimania Singapore



Demonstration

Patrol recent changes in the Android App

https://commons.wikimedia.org/wiki/File:Edit_Patrol_Demo_-_Updated_April_29_2024.webm

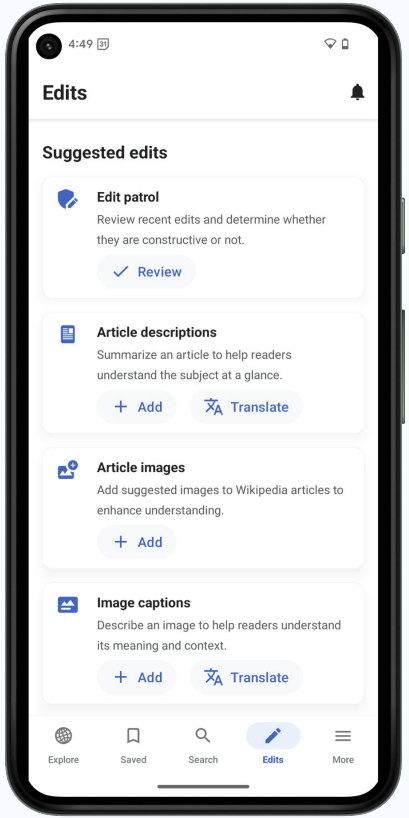


How to try it out

This feature is currently only available to users with rollback rights

1. Download the Wikipedia App from the Google Play Store
2. Set your primary language in the App
3. Sign in
4. Open the “Edits” tab
5. Open “Edit Patrol”
6. Begin reviewing edits!

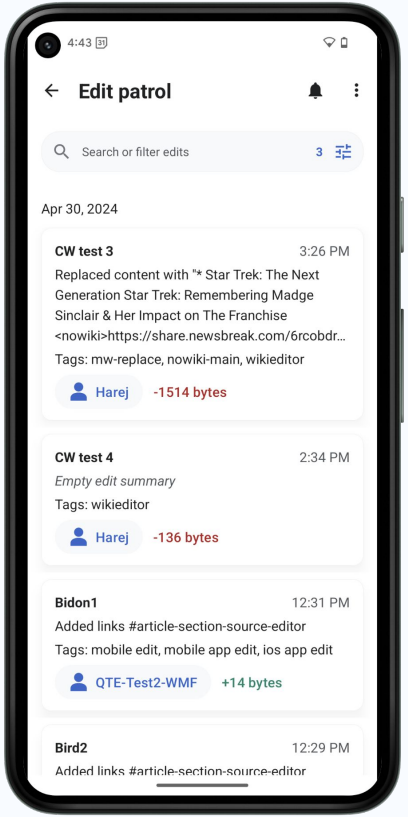
Wikipedia App



Incorporating Feedback

Subtitle

- Bullets



Feedback is welcome & valued

Let us know what you think after trying the tool

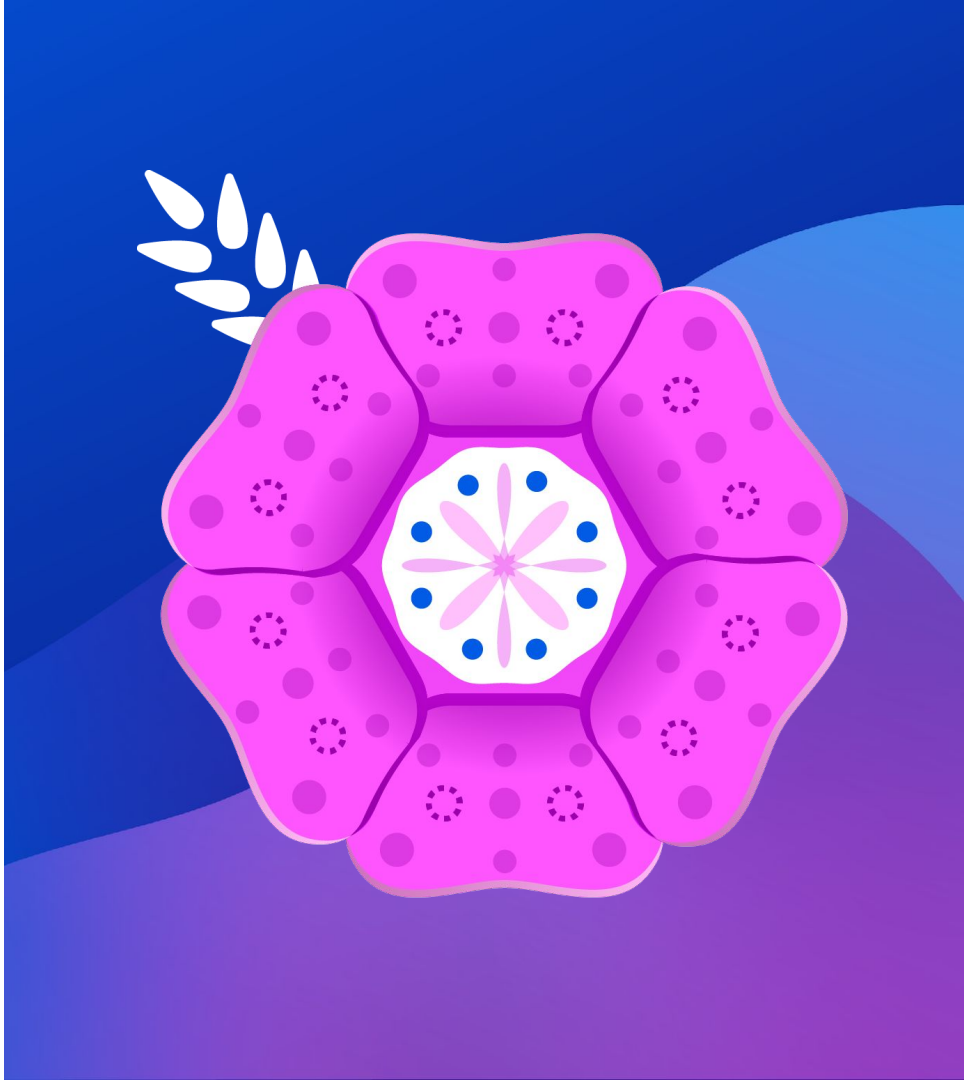
Fill out Google Form

or Share your answers on our Discussion Page

Automoderator

What if we could automatically find and revert vandalism, so that you didn't need to spend as much time reviewing new changes?

ESEAP
KK24



Some wikis have an anti-vandalism bot.

We're building something similar, but making it available to all wikis.

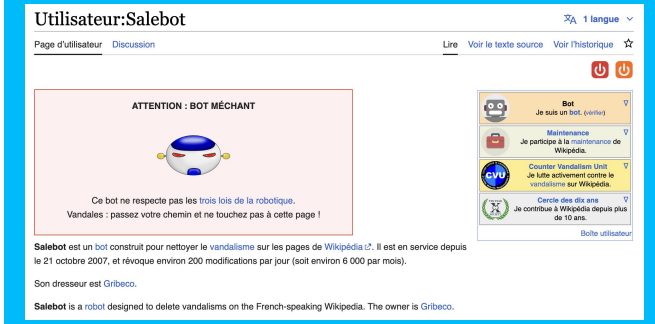
ClueBot NG

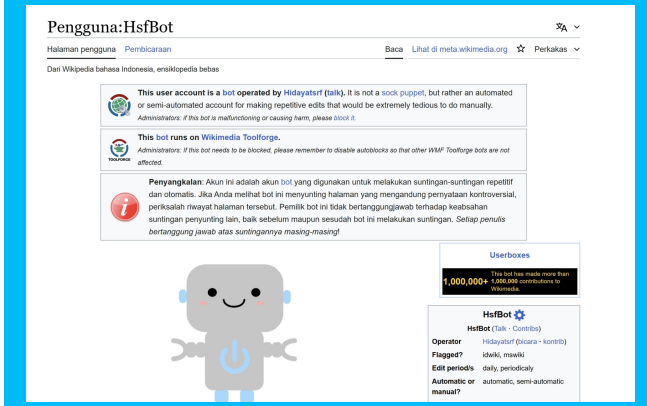
This user is a bot

(talk · contribs)

ClueBot NG aids in Operation Enduring Encyclopedia.

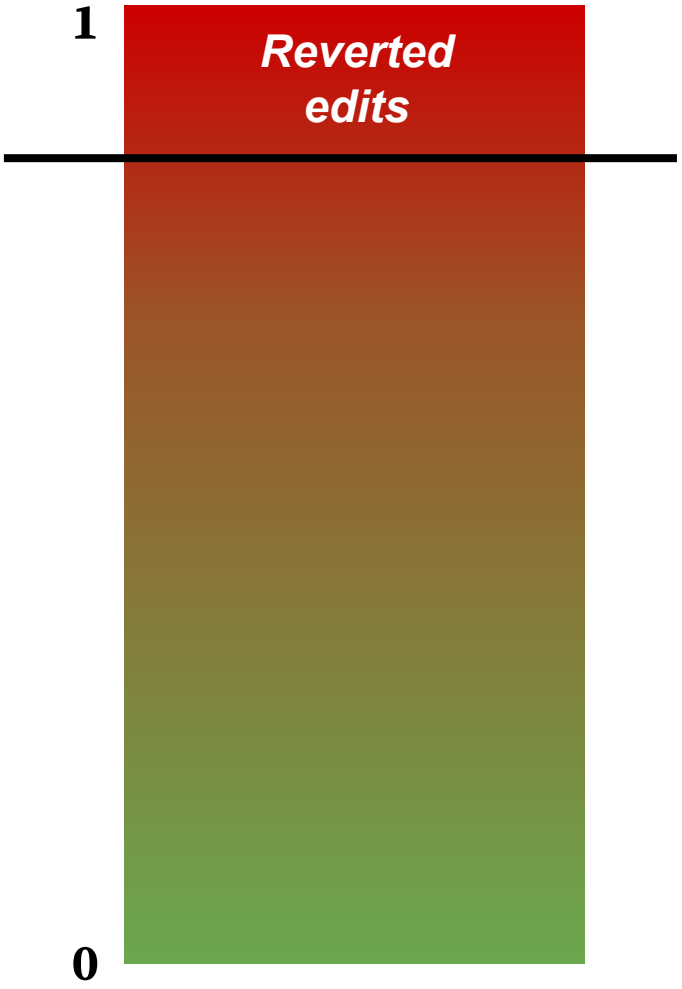
Operator	Rich Smith, DamianZaremba
Approved?	Yes, BRFA.
Flagged?	Yes.
Edit period(s)	Continually
Automatic or manual?	Automatic
Programming language(s)	C, C++, PHP, Python, Bash, and Java (more info)
Exclusion compliant?	Yes
Emergency shutoff-compliant?	Yes
Other information	ClueBot NG is run from the Wikimedia Toolforge infrastructure.





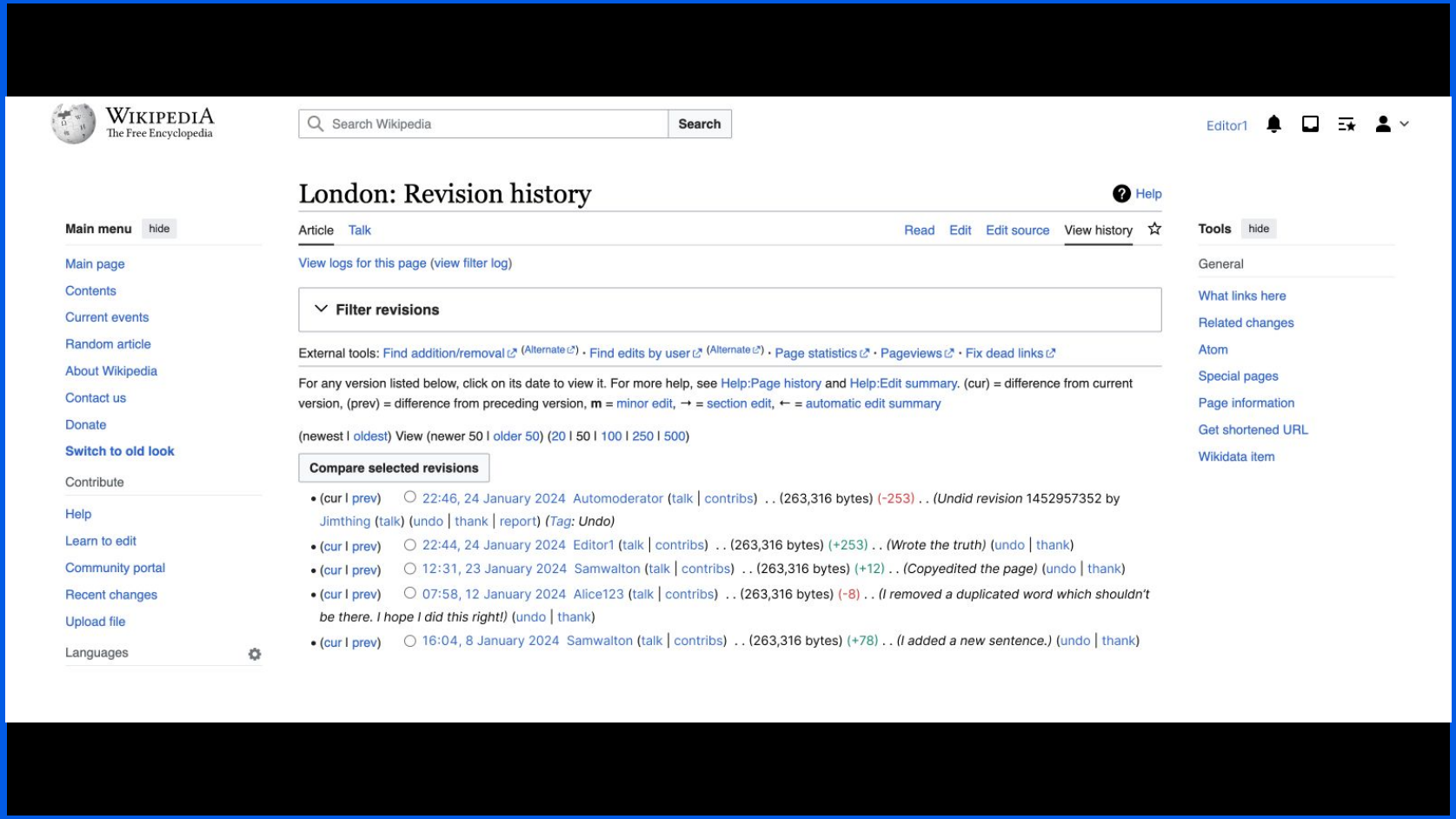
- (beda | riw) . . Hartarto Sastrosoenarto; 15.24 . . **(+678)** . . SASTROSOEMARTO (bicara | kontrib) *(Tag: menambah tag nowiki, gambar rusak, VisualEditor, Suntingan perangkat seluler, Suntingan peramban seluler)* (berterima kasih) → 0.65
- (Log pengguna baru); 15.23 . . SASTROSOEMARTO (bicara | kontrib) membuat akun pengguna *(Tag: Suntingan perangkat seluler, Suntingan peramban seluler)*
- (beda | riw) . . Universal Music Indonesia; 15.23 . . (+18) . . Eja Kurniawan (bicara | kontrib) *(Tag: Suntingan perangkat seluler, Suntingan aplikasi seluler, Suntingan aplikasi Android)* (berterima kasih) → 0.34
- (beda | riw) . . Nakusha (seri televisi); 15.22 . . (0) . . Abcdef242526 (bicara | kontrib) *(→Penayangan di Indonesia) (Tag: Suntingan perangkat seluler, Suntingan peramban seluler)* (berterima kasih) → 0.19
- (beda | riw) . . Nakusha (seri televisi); 15.22 . . (−98) . . Abcdef242526 (bicara | kontrib) *(Tag: Suntingan perangkat seluler, Suntingan peramban seluler)* (berterima kasih) → 0.20
- (beda | riw) . . Nakusha (seri televisi); 15.18 . . (−14) . . Abcdef242526 (bicara | kontrib) *(Tag: Suntingan perangkat seluler, Suntingan peramban seluler)* (berterima kasih) → 0.07
- (beda | riw) . . Di Antara Dua Cinta; 15.17 . . (+14) . . WillsonEP09 (bicara | kontrib) *(→Pemeran: updated) (Tag: Suntingan perangkat seluler, Suntingan peramban seluler, Suntingan seluler lanjutan)* (berterima kasih) → 0.53
- (beda | riw) . . PSIW Wonosobo; 15.17 . . (+13) . . 36.71.85.175 (bicara) *(tentang jukukan tim psiw wonosobo, dimana jukukan ini dari kesepakatan teman teman suporter psiw) (Tag: VisualEditor, Suntingan perangkat seluler, Suntingan peramban seluler)* → 0.93
- (beda | riw) . . Accor; 15.17 . . (−230) . . Rudyindarto (bicara | kontrib) *(→Accor di Indonesia)* (berterima kasih) → 0.97
- (beda | riw) . . Babak kualifikasi Kejuaraan Bulu Tangkis Beregu Putra dan Putri Eropa 2024; 15.17 . . (+376) . . Danfinklerr (bicara | kontrib) *(→Grup 1)* (berterima kasih) → 0.49
- (beda | riw) . . Daftar acara ANTV; 15.12 . . (+54) . . Abcdef242526 (bicara | kontrib) *(→Acara saat ini) (Tag: Suntingan perangkat seluler, Suntingan peramban seluler)* (berterima kasih) → 0.21

How does it work?



Edits scoring above a particular threshold will be reverted.

What will it look like?


Wikipedia
The Free Encyclopedia
Search Wikipedia
Search
Editor1
London: Revision history
Help
Main menu hide
Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate
Switch to old look
Contribute
Help
Learn to edit
Community portal
Recent changes
Upload file
Languages
Article Talk
Read Edit Edit source View history
Tools hide
General
What links here
Related changes
Atom
Special pages
Page information
Get shortened URL
Wikidata item
View logs for this page (view filter log)
Filter revisions
External tools: Find addition/removal (Alternate) · Find edits by user (Alternate) · Page statistics · Pageviews · Fix dead links
For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary. (cur) = difference from current version, (prev) = difference from preceding version, m = minor edit, → = section edit, ← = automatic edit summary
(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)
Compare selected revisions
• (cur | prev) 22:46, 24 January 2024 Automoderator (talk | contribs) . . (263,316 bytes) (-253) . . (Undid revision 1452957352 by Jimthing (talk) (undo | thank | report) (Tag: Undo)
• (cur | prev) 22:44, 24 January 2024 Editor1 (talk | contribs) . . (263,316 bytes) (+253) . . (Wrote the truth) (undo | thank)
• (cur | prev) 12:31, 23 January 2024 Samwalton (talk | contribs) . . (263,316 bytes) (+12) . . (Copyedited the page) (undo | thank)
• (cur | prev) 07:58, 12 January 2024 Alice123 (talk | contribs) . . (263,316 bytes) (-8) . . (I removed a duplicated word which shouldn't be there. I hope I did this right!) (undo | thank)
• (cur | prev) 16:04, 8 January 2024 Samwalton (talk | contribs) . . (263,316 bytes) (+78) . . (I added a new sentence.) (undo | thank)

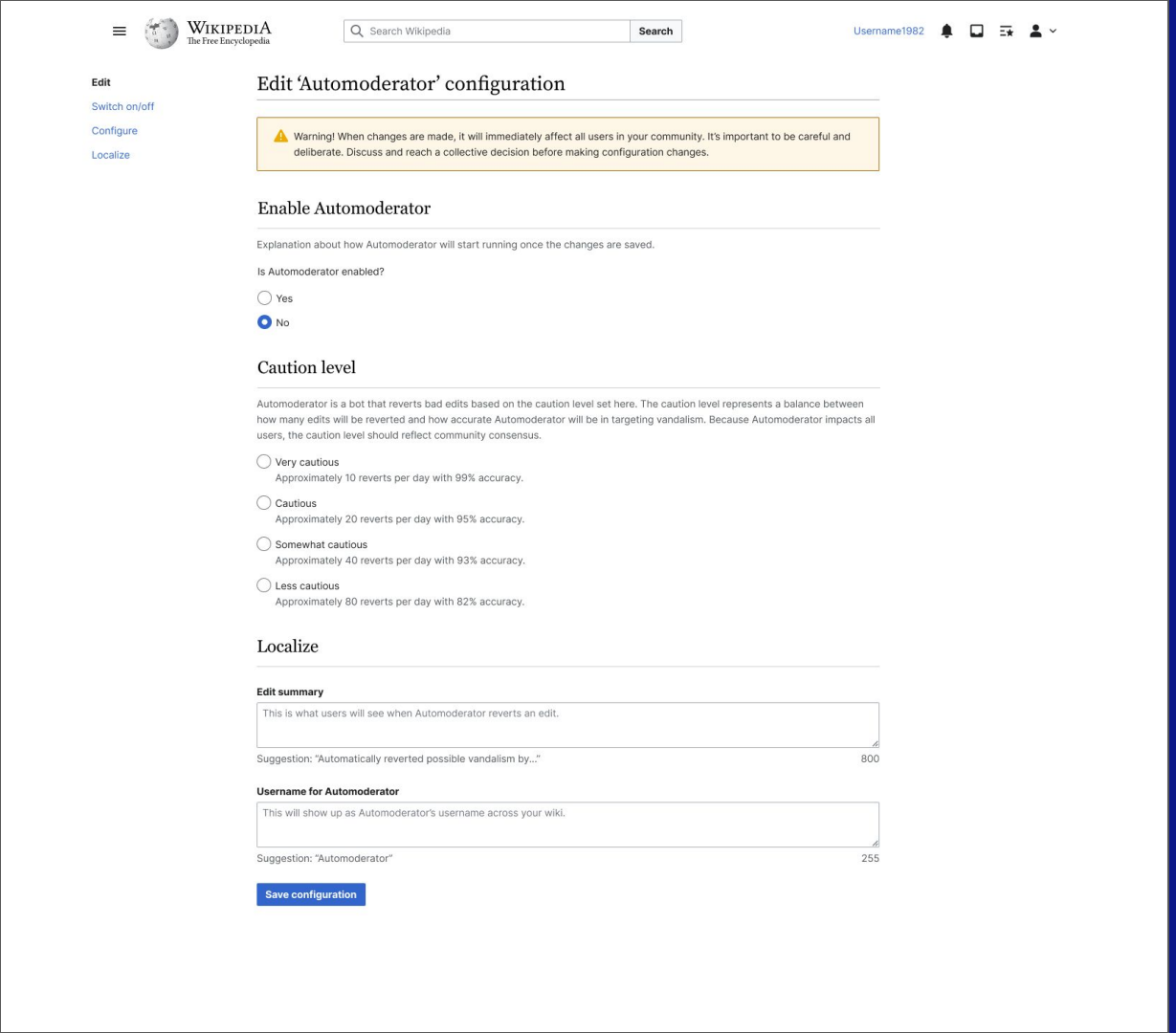
External tools: Find addition/removal (Alternate) · Find edits by user (Alternate) · Page statistics · Pageviews · Fix dead links

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary. (cur) = difference from current version, (prev) = difference from preceding version, **m** = minor edit, → = section edit, ← = automatic edit summary

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

- (cur | prev) ○ 22:46, 24 January 2024 Automoderator (talk | contribs) . . (263,316 bytes) (-253) . . *(Undid revision* 1452957352 by Jimthing (talk) (undo | thank | report) *(Tag: Undo)*
- (cur | prev) ○ 22:44, 24 January 2024 Editor1 (talk | contribs) . . (263,316 bytes) (+253) . . *(Wrote the truth)* (undo | thank)
- (cur | prev) ○ 12:31, 23 January 2024 Samwalton (talk | contribs) . . (263,316 bytes) (+12) . . *(Copyedited the page)* (undo | thank)
- (cur | prev) ○ 07:58, 12 January 2024 Alice123 (talk | contribs) . . (263,316 bytes) (-8) . . *(I removed a duplicated word which shouldn't be there. I hope I did this right!)* (undo | thank)
- (cur | prev) ○ 16:04, 8 January 2024 Samwalton (talk | contribs) . . (263,316 bytes) (+78) . . *(I added a new sentence.)* (undo | thank)



Administrators will have full control over Automoderator

Automoderator will not be 100% accurate.

Project page with updates

WIKIPEDIA
Ensiklopedia Bebas

Telusuri Wikipedia | Cari

Samwalton9 (WMF)

Wikipedia:Automoderator

Tambah bahasa

Halaman proyek | Pembicaraan | Baca | Sunting sumber | Lihat riwayat

Dari Wikipedia bahasa Indonesia, ensiklopedia bebas

Automoderator adalah perkakas yang dikembangkan oleh tim Moderator Tools dari Wikimedia Foundation. Perkakas ini dapat mengembalikan suntingan berpotensi vandalisme secara otomatis berdasarkan model pembelajaran mesin. Selain itu, perkakas ini dapat dikonfigurasi berdasarkan keputusan dan/atau kebutuhan komunitas (misalnya mengaktifkan atau menonaktifkan perkakas).

Melalui perkakas ini, diharapkan dapat mengurangi beban kontributor dalam melakukan patroli untuk mencegah dan/atau memberantas vandalisme. Perkakas ini masih dalam tahap awal pengembangan dan akan diujicoba pertama kali di Wikipedia bahasa Indonesia, maka kami harapkan masukan dan partisipasi Anda.

Informasi selengkapnya mengenai perkakas ini dapat dilihat di halaman MediaWiki Automoderator.

Ilustrasi dari peralatan Automoderator ketika mengembalikan suntingan.

Apa itu Automoderator? [sunting sumber]

Automoderator masih dalam tahap pengembangan. Perkakas ini menggunakan model pembelajaran mesin terbaru untuk mengidentifikasi suntingan yang perlu dikembalikan (dalam hal ini adalan vandalisme). Jika ada suntingan yang ditandai untuk dikembalikan, maka perkakas akan melakukan pengembalian suntingan secara otomatis berdasarkan preferensi yang ditentukan oleh komunitas.

Diperkirakan perkakas akan mengembalikan rata-rata sebanyak 7 hingga 29 suntingan per hari (lihat tabel di bawah).

Berhubung Automoderator menggunakan model pembelajaran mesin, maka sudah pasti tidak akan akurat sepenuhnya. Terkadang akan terjadi kesalahan pemeriksaan dan mengembalikan suntingan berniat baik secara tidak sengaja. Kami perkirakan, berdasarkan model ini, bahwa tingkat keberhasilan perkakas ini mencapai 99%.

Presentasi di Wikimania mengenai Automoderator (13:50).

Cara menggunakan [sunting sumber]

Menu utama sembunyikan

- Halaman Utama
- Daftar isi
- Perubahan terbaru
- Artikel pilihan
- Peristiwa terkini
- Halaman baru
- Halaman sembarang
- **Beralih ke tampilan lama**

Komunitas

- Warung Kopi
- Portal komunitas
- Bantuan

Wikipedia

- Tentang Wikipedia
- Pancapilar
- Kebijakan
- Menyumbang
- Hubungi kami
- Bak pasir

Bagikan

Perkakas sembunyikan

Tindakan

- Pindahkan
- Berlangganan

Umum

- Pranala balik
- Perubahan terkait
- Halaman istimewa
- Pranala permanen
- Informasi halaman
- Lihat URL pendek
- Unduh kode QR
- Pranala menurut ID
- Tambahkan pranala interwiki

Cetak/ekspor

- Buat buku
- Unduh versi PDF
- Versi cetak

During the pilot we will monitor lots of data

- Number of reverts
- Number of false positive reports
- Activity level of patrollers
- Impact on new editors
- ...

We will keep working on Automoderator

- More configuration options (e.g. skip certain user groups)
- Send a talk page message when reverted
- Integrating Community Configuration user interface
- Features for testing Automoderator's caution levels
- …
- *Features requested by you!*


Questions?