

Universitat  
Pompeu Fabra  
Barcelona



# Характеристика маніпуляцій знаннями в альтернативній версії російської Вікіпедії [a.k.a. RWFork]

**Authors:**  
Mykola Trokhymovych  
Oleksandr Kosovan  
Nathan Forrester  
Pablo Aragón  
Diego Saez-Trumper  
Ricardo Baeza-Yates

# Про мене



# Mykola Trokhymovych

trokhymovych.com | mykola.trokhymovych@upf.edu

- PhD-кандидат @ Universitat Pompeu Fabra
  - Працюю під науковим керівництвом Diego Saez-Trumper та Ricardo Baeza-Yates
- Запрошений дослідник, Indiana University (OSoMe)
  - Досліджую виявлення контенту, згенерованого ШІ в соціальних мережах, разом із Fil Menczer
- Досвід дослідницької роботи з екосистемою Wikimedia більше 5 років.

## Конференції:

- **ACL'25, ICWSM'25, ACL'24, KDD'23, CIKM'21**

## Раніше:

- 🎓 Науки про дані @ УКУ; Системний аналіз @ ІПСА, КПІ
- 👜 Data Scientist / ML Engineer: Ciklum, Jooble, Surprise

**Чому Вікіпедія?**


**Альтернативна версія Вікіпедії**

**Збір і обробка даних**


**Аналіз змін в контенті сторінок**

**Висновки і дискусія**

# Чому Вікіпедія?

-  **ТОП-10 найвідвідуваніших сайтів світу**
  - 300+ мов, понад 60 млн статей у Вікіпедії, ~16 редагувань щосекунди.
  - 6-те місце серед найважливіших джерел інформації про історію, випередивши відвідування музеїв, університетські курси та соціальні мережі (Burkholder і Schaffer, 2021).
-  **Прозорість і open-source**
  - Вікіпедія існує завдяки добровільним пожертвам мільйонів користувачів у всьому світі.
  - Створюється та підтримується спільнотою з метою стати «сумою всіх знань людства» (Sutcliffe, 2016).
  - У Вікіпедії діють чіткі правила, і це дає змогу запобігти впливу прихованих інтересів.
-  **Вплив поза Wikimedia ecosystem:**
  - З'являється в результатах Google, Siri, Alexa, та інші.
  - Використовується для навчання штучного інтелекту (включно з ChatGPT).

# Вікіпедія і цензурування

 Деякі гравці не погоджуються з контентом і політиками Вікіпедії.

## Блокування

Online encyclopedia Wikipedia blocked in China ahead of Tiananmen anniversary

By Josh Horwitz

May 15, 2019 12:02 PM GMT+2 · Updated May 15, 2019



Turkey blocks Wikipedia under law designed to protect national security

Users trying to access online encyclopaedia via Turkish internet providers receive 'connection timed out' error message



## Створення альтернативи

Conservapedia - the US religious right's answer to Wikipedia



How Baidu built an encyclopedia with 16 times more Chinese entries than Wikipedia

Transition from hard format encyclopaedias to open-source encyclopaedias has raised some quality-of-information issues though



**Hamichlol** (Hebrew: המכלול "The Entirety") is a **censored wiki-based Internet encyclopedia project** for the **Haredi public**. Most of the articles on the site are exact copies (as of the date they were copied) of articles from the **Hebrew Wikipedia** or re-edited versions of Hebrew Wikipedia articles, adapted for the Haredi reader, while



# Інтернет-врядування на РФ

Ранній розвиток Інтернету в Росії відбувався без державного контролю, що призвело до складної й децентралізованої мережі (Glasze et al. 2023)

Роскомнадзор змушений покладатися на добровільну співпрацю автономних систем та провайдерів, але нестача довіри знижує ефективність (Ermoshina & Musiani 2022)

Великі інтернет-компанії (Yandex, VK, Mail.ru) тісно пов'язані з державою, виступають посередниками цензури (Zuckerman 2010). Проте держава шукає нові форми онлайн-контролю (Wijermars 2021).



Yandex

@mail

# Створення альтернативної Вікіпедії (RWFork)

У липні 2023 року було запущено сайт Ruwiki (RWFork), відгалуження російської Вікіпедії, створене для відповідності вимогам російського законодавства.

Новий проект очолив багаторічний функціонер російської Вікіпедії.

Це сталося після того, як уряд РФ наклав штрафи на Вікімедіа на суму близько 255 мільйонів доларів.

## In Putin's Russia, 'fake news' now means real news

A term coined by misinformation researchers was coopted by President Trump. It has become a tool of authoritarians.

## Russia fines Wikimedia Foundation over Ukraine war entries

By Reuters

≡ Bloomberg

Subscribe



Businessweek | Technology

## Russian Wikipedia's Top Editor Leaves to Launch a Putin-Friendly Clone

The move is seen as a potential precursor to a ban on the original version.



TIME

SUBSCRIBE

WORLD • RUSSIA

## Russia to Create Its Own 'Alternative Wikipedia'

Яким чином інформація з Вікіпедії  
була зміненою в RWFork?

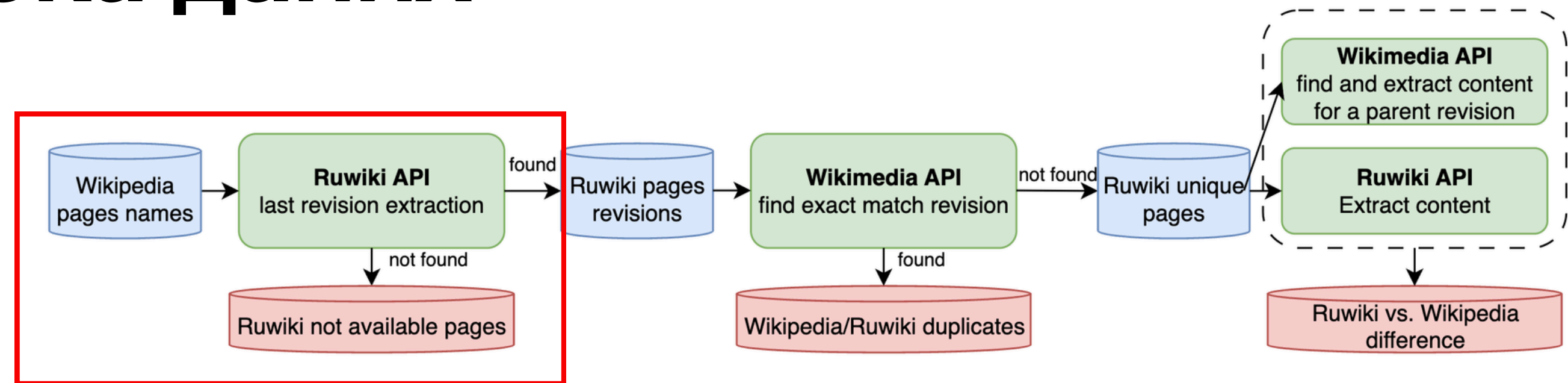
**33,664** (1.75%)

сторінок було змінено за перші  
три місяці існування проекту



**РУВИКИ** <sup>β</sup>

# Збір і обробка даних



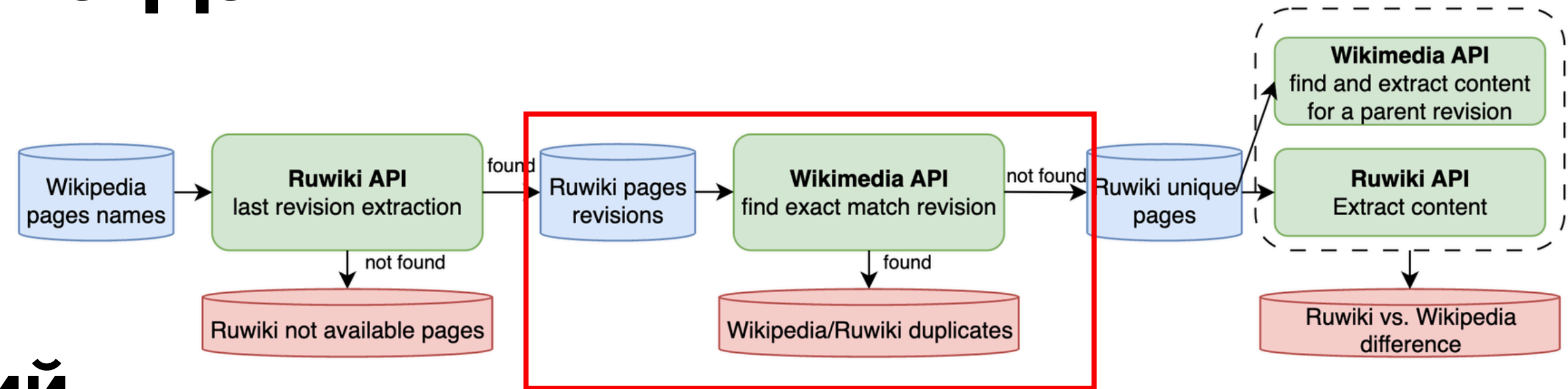
## Вибір статей

- ◆ Список із ~1,9 млн заголовків було зібрано на основі дампу Wikimedia.
- ◆ Здійснено пошук відповідних статей у RWFork.

## Wikimedia Downloads

If you are reading this on Wikimedia servers, please note that we have rate limited downloaders and we are capping the number of per-ip connections to 2. This will help to ensure that everyone can access the files with reasonable download times. Clients that try to evade these limits may be blocked. Our mirror sites do not have this cap.

# Збір і обробка даних



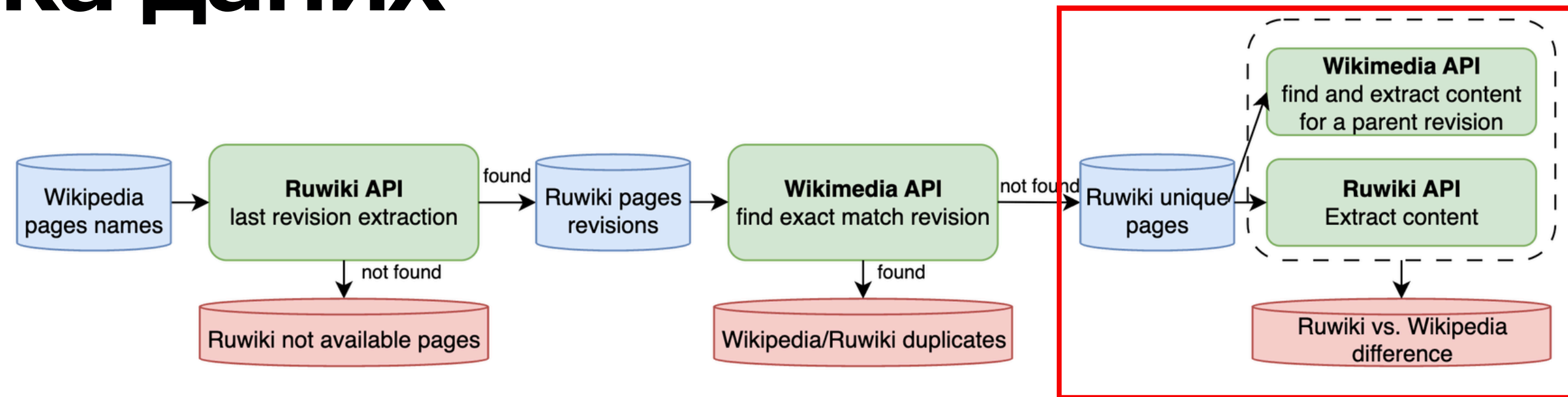
## Автоматизований обхід та збір даних з вебсайту

🔍 Визначення батьківської версії сторінки з Вікіпедії, на основі якої було створено сторінку в RWFork.

✍️ 33 664 сторінки (1,75%) містили редагування, зроблені вже в рамках RWFork.

👤 (текущ.   пред.)	🕒 16:40, 20 июня 2023	👤 Сапорус (обсуждение   вклад)	📄 (120 278 байт)	📊 (-559)	📄 (оформление)	🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 07:13, 19 июня 2023	👤 Sunshine (обсуждение   вклад)	📄 (120 837 байт)	📊 (+1296)		🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 21:35, 17 июня 2023	👤 Сапорус (обсуждение   вклад)	📄 (119 541 байт)	📊 (-17 258)	📄 (rv) (Метка: отмена)	🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 21:29, 17 июня 2023	👤 (имя автора стёрто)	📄 (136 799 байт)	📊 (-2)	📄 (описание правки стёрто)	🗑️ (Метка: отменено)	
👤 (текущ.   пред.)	🕒 21:26, 17 июня 2023	👤 (имя автора стёрто)	📄 (136 801 байт)	📊 (+17 262)	📄 (описание правки стёрто)	🗑️ (Метки: ручная отмена, отменено)	
👤 (текущ.   пред.)	🕒 21:23, 17 июня 2023	👤 (имя автора стёрто)	📄 (119 539 байт)	📊 (-2)	📄 (описание правки стёрто)	🗑️ (Метка: отменено)	
👤 (текущ.   пред.)	🕒 19:03, 7 июня 2023	👤 Сапорус (обсуждение   вклад)	📄 (119 541 байт)	📊 (-95)	📄 (оформление)	🗑️ (отменить)	🗑️ (Метка: отменено)
👤 (текущ.   пред.)	🕒 18:19, 7 июня 2023	👤 Сапорус (обсуждение   вклад)	📄 (119 636 байт)	📊 (-145)	📄 (оформление)	🗑️ (отменить)	🗑️ (Метка: отменено)
👤 (текущ.   пред.)	🕒 16:36, 6 июня 2023	👤 Сапорус (обсуждение   вклад)	📄 (119 781 байт)	📊 (-17 020)	📄 (707437-НВ)	🗑️ (отменить)	🗑️ (Метки: отменено, редактор вики-текста 2017)
👤 (текущ.   пред.)	🕒 18:16, 2 апреля 2023	👤 © Перенесено из ru.wikipedia.org/wiki/Херсон по лицензии CC BY-SA 3.0. Список авторов источника в «Истории». В страницу вносились изменения после переноса. (Unknown user)	📄 (136 801 байт)	📊 (0)	📄 (Изменил настройки защиты для «Херсон»: Автоматическая полузащита статьи из списка потенциально опасных для редакторов ([Редактирование=только автоподтверждённые] (бессрочно) [Переименование=только автоподтверждённые] (бессрочно)))	🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 09:48, 23 марта 2023	👤 © Перенесено из ru.wikipedia.org/wiki/Херсон по лицензии CC BY-SA 3.0. Список авторов источника в «Истории». В страницу вносились изменения после переноса. (Unknown user)	📄 (136 801 байт)	📊 (-21)	📄 (→Ссылки)	🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 13:32, 7 февраля 2023	👤 © Перенесено из ru.wikipedia.org/wiki/Херсон по лицензии CC BY-SA 3.0. Список авторов источника в «Истории». В страницу вносились изменения после переноса. (Unknown user)	📄 (136 822 байта)	📊 (-30)	📄 (→Фотогалерея Херсона)	🗑️ (отменить)	🔒 [отпатрулирована автоматически]
👤 (текущ.   пред.)	🕒 04:35, 4 февраля 2023	👤 © Перенесено из ru.wikipedia.org/wiki/Херсон по лицензии CC BY-SA 3.0. Список авторов источника в «Истории». В страницу вносились изменения после переноса. (Unknown user)	📄 (136 852 байта)	📊 (-2324)	📄 (→СМИ: по шаблону)	🗑️ (отменить)	🗑️ (Метка: правка через визуальный редактор) [отпатрулирована автоматически]

# Збір і обробка даних



- ◆ Окремо було витягнуто вміст останньої ревізії з RWForк та відповідної батьківської версії з Вікіпедії.

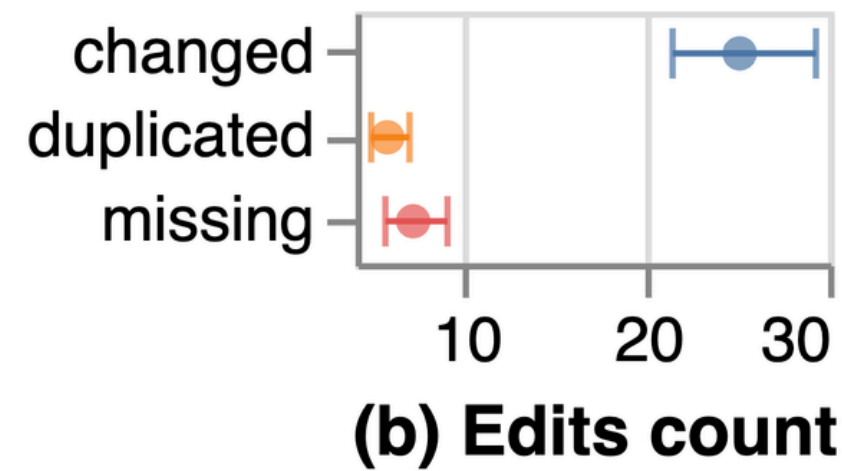
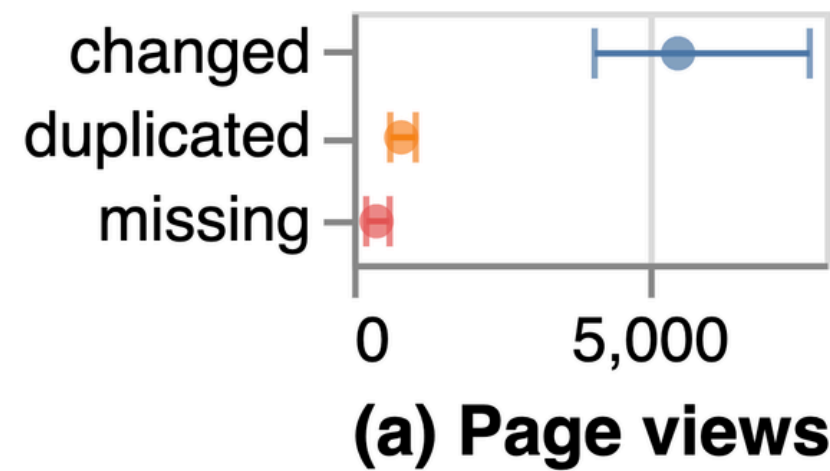
- ◆ Для виявлення та обробки текстових змін використовувався інструмент mwedittypes.

- ◆ Додатково було витягнуто країни, теми статей та інші статистичні дані сторінки з Вікіпедії.

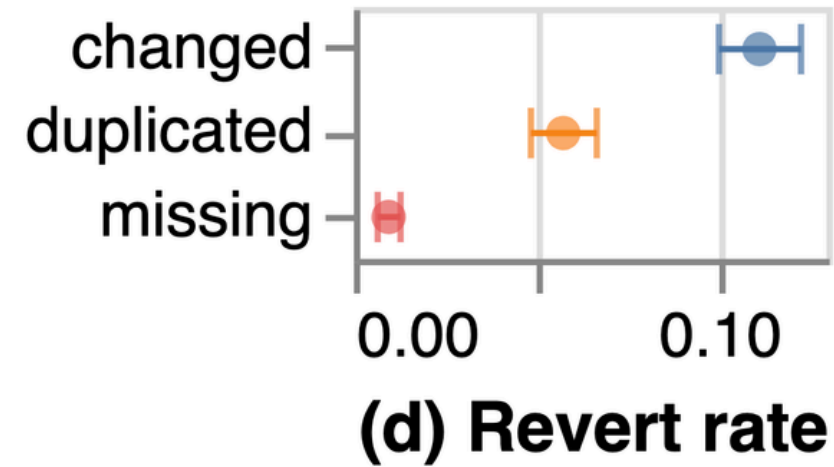
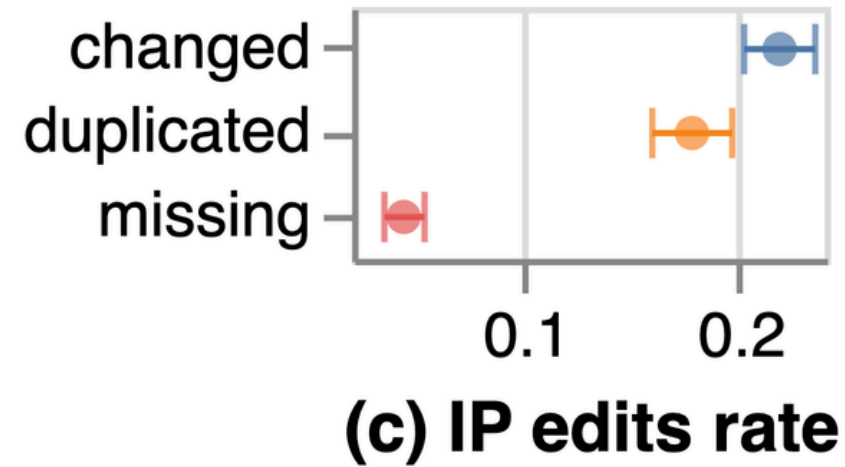
The screenshot shows a Wikipedia difference page for 'Laurit Krasniqi: Difference between revisions'. The page title is 'Laurit Krasniqi: Difference between revisions'. Below the title, it says 'From Wikipedia, the free encyclopedia'. There are navigation links: 'Browse history into', 'Revision as of 11:13, 29 July 2022 (edit)', 'ManiacOfSport (talk | contribs)', '← Previous edit', '(Tags: Reverted, Mobile edit, Mobile web edit)', and 'Next edit →'. A red box highlights the 'Parsed text changes from wikitext:' section, which lists: (1) text inserts: [], (2) texts removed: [], (3) text changes: [(['Born in Belgium, he has chosen to represent the Kosovo national team.', 'Born in Kosovo, he has chosen to represent the Kosovo national team.').]. Below this, two side-by-side text blocks show the difference between two versions of the text. The left block shows the original text with a minus sign (-) and the right block shows the updated text with a plus sign (+). The updated text includes the change from 'Born in Belgium' to 'Born in Kosovo'.

Приклад отримання різниці двох версій сторінки

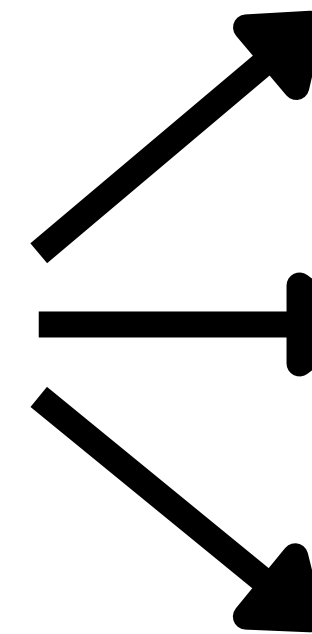
# Зміни здебільшого у впливових та суперечливих статтях



**Group:**  
● changed  
● duplicated  
● missing



**1.75%**  
змінених статей

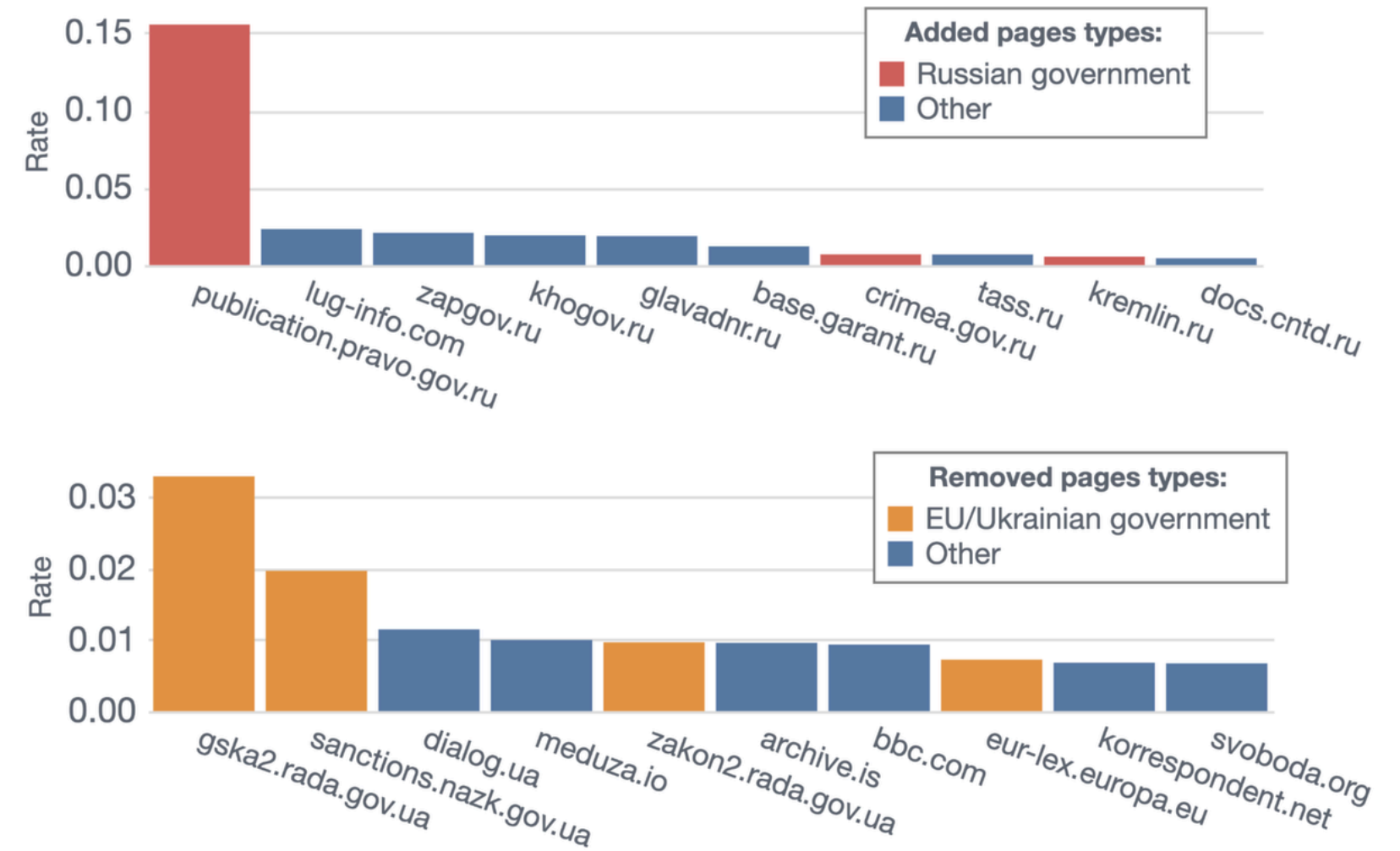
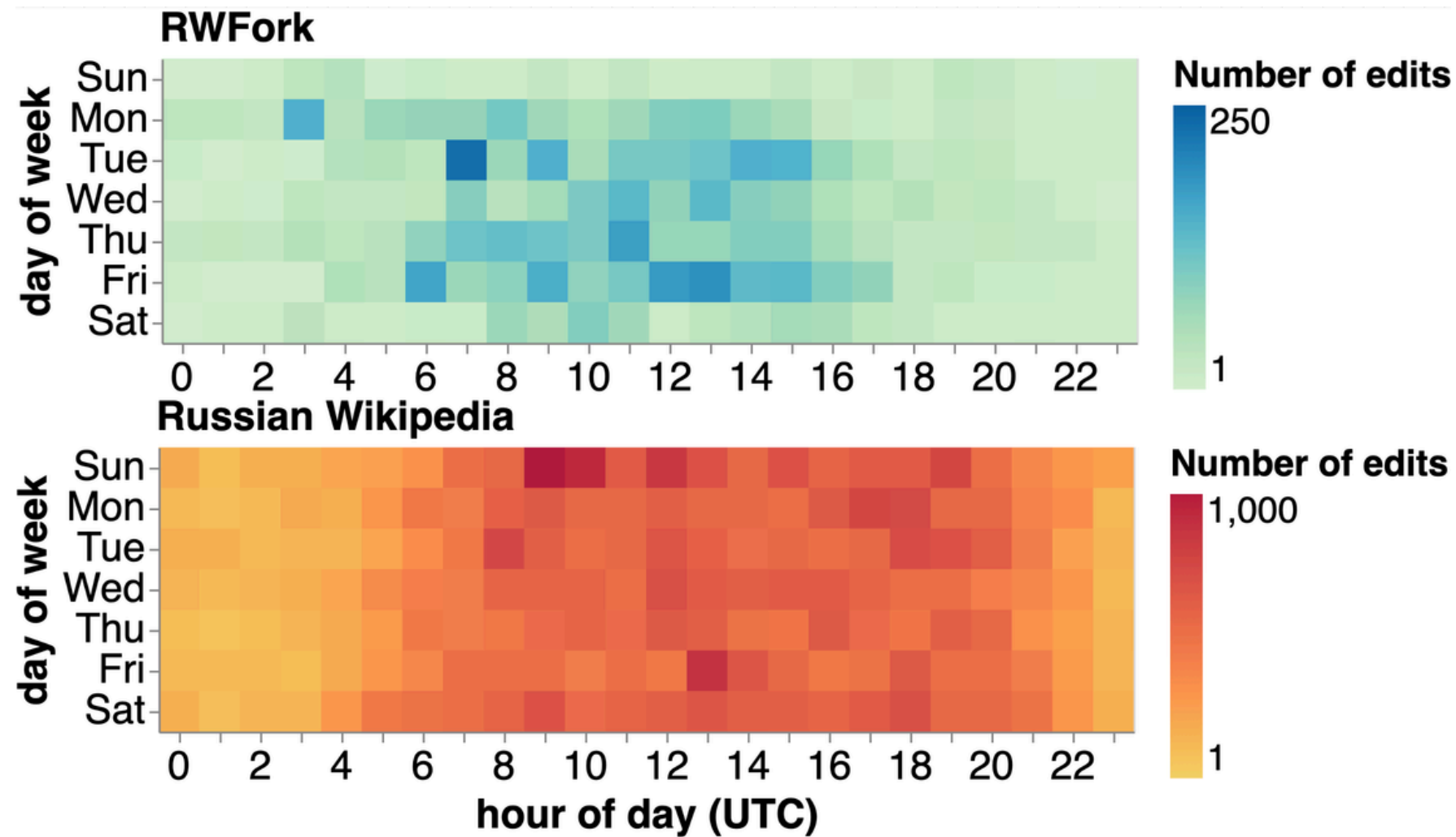


**14.2%**  
переглядів

**~x2**  
Частота відкатів

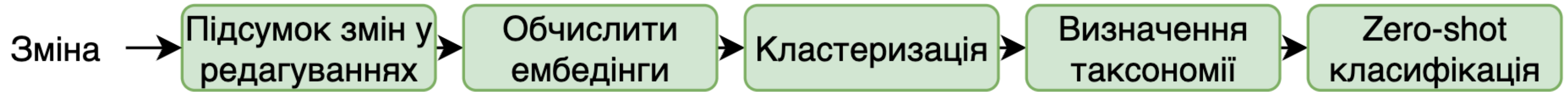
**~x5**  
частота змін

# Скоординована активність у редагуванні



- Піки активності спостерігаються у **робочі години в будні дні**
- Фокус на сторінках, пов'язаних з Україною, Росією та Білоруссю
- Додає джерела російського уряду, видаляє джерела ЄС та України

# Систематизація змін



*Text to summarize example:*

**page title:** Калиновское;

**lines changed:**

*Old value:* Калиновское — село в Белокуракинском районе Луганской области

*New value:* Калиновское — село в Белокуракинском районе Луганской Народной Республики

*Summary:*

The designation of the administrative region was changed from 'Луганской области' (Luhansk Oblast) to 'Луганской Народной Республики' (Luhansk People's Republic), reflecting a political and territorial update.

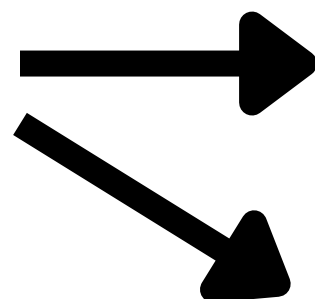
*Cluster:* Territorial Claims Dispute

*Cluster summary:* Edits reflect changes in territorial designations and governance, emphasizing claims by the Donetsk and Luhansk People's Republics while removing Ukrainian references and administrative details.

# Систематична маніпуляція знаннями



~90%  
ЗМІН



8  
основні категорії

~44%

пов'язано з російсько-українською війною

# Підсумки

Перше дослідження того, як оригінальний контент Вікіпедії було відгалужено та змінено для відповідності вимогам національного законодавства.

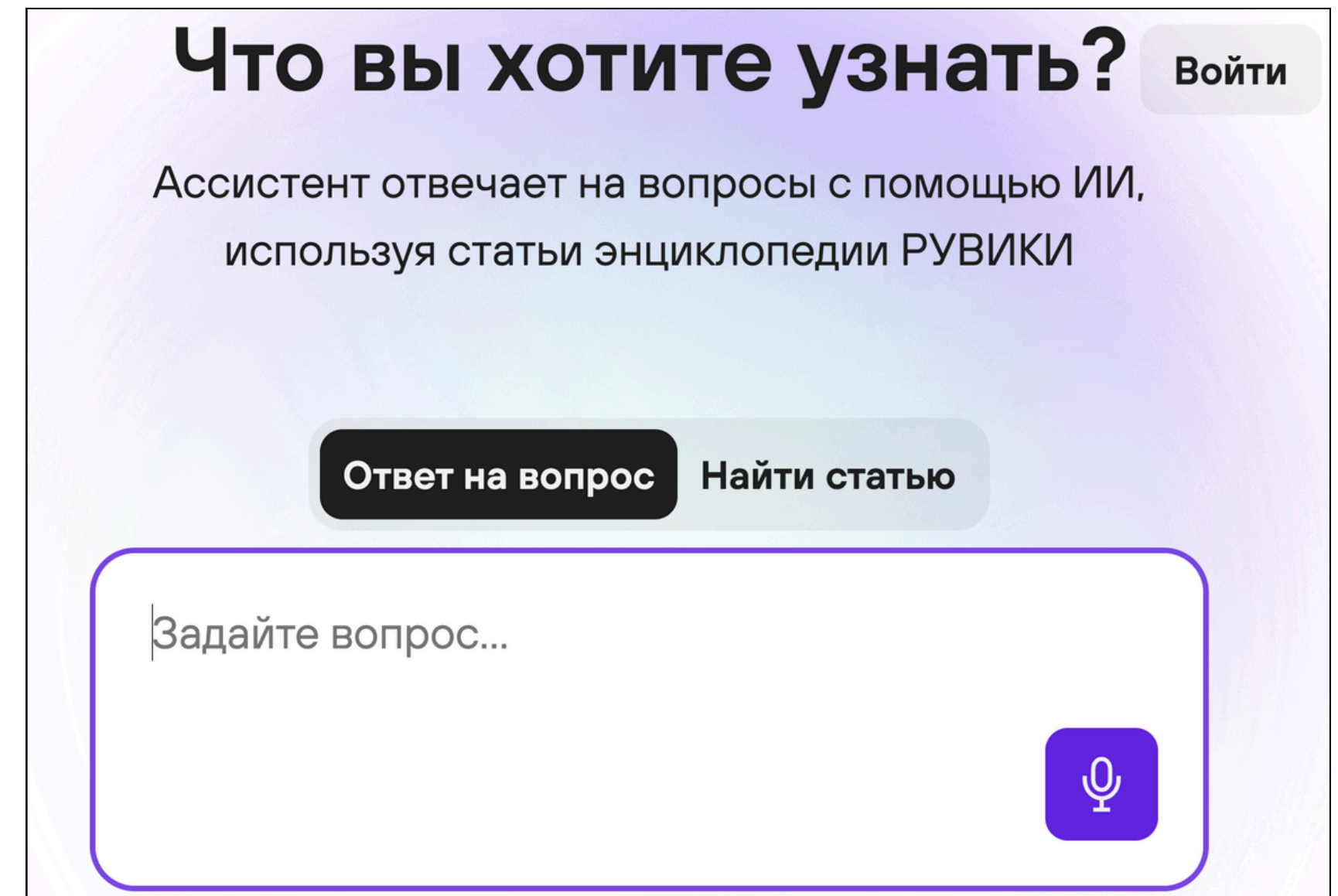
- Редагування у RWFork спрямовані на найвідвідуваніші та суперечливі сторінки.
- Редактори діють за моделлю активності в робочі години.
- Зміни систематичні і переважно стосуються російсько-української війни.

Набір даних доступний для подальших досліджень:  
<https://zenodo.org/records/15073728>



# Чому це важливо?

- Інформаційне підсилення та інституціалізація окремих сайтів
- Небезпека використання маніпулятивного контенту в навчанні та використанні ШІ. Верифікації фактів



# Епілог

**Відокремлюйте емоції від фактів.**

Світ цього тижня може отримати можливість припинити війну, — Єрмак

★ 1 😊 178 🙏 103 🤖 14 ❤️ 8 👁️ 14.1K

**Не забуваймо думати та перевіряти!**

**ТСН** ВТ, 5 СЕРПНЯ Курс валют: \$41.45 €47.86 ☀️ +29° UA ▾

13:35, 27.07.25 ■ Наука та IT ■ 👁️ 52k ■ ⌚ 3 хв

**У листопаді цього року інопланетяни можуть атакувати Землю – New York Post**

**Дякую!**