state of

Machine learning at Wikipedia

Santhosh Thottingal
Principal Software Engineer, Wikimedia Foundation

Projects using Machine Learning at Wikipedia

- Use cases
- Guiding principles
- Product design
- Challenges
- Impact

WIKIMEDIA FOUNDATION

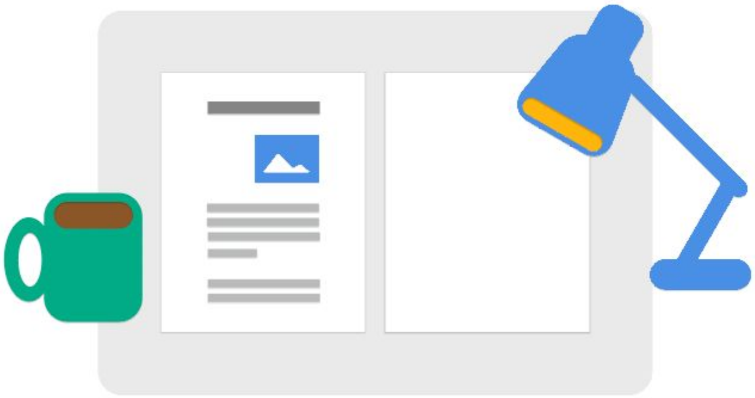
01 Content Translation

Machine Translation

Content Translation

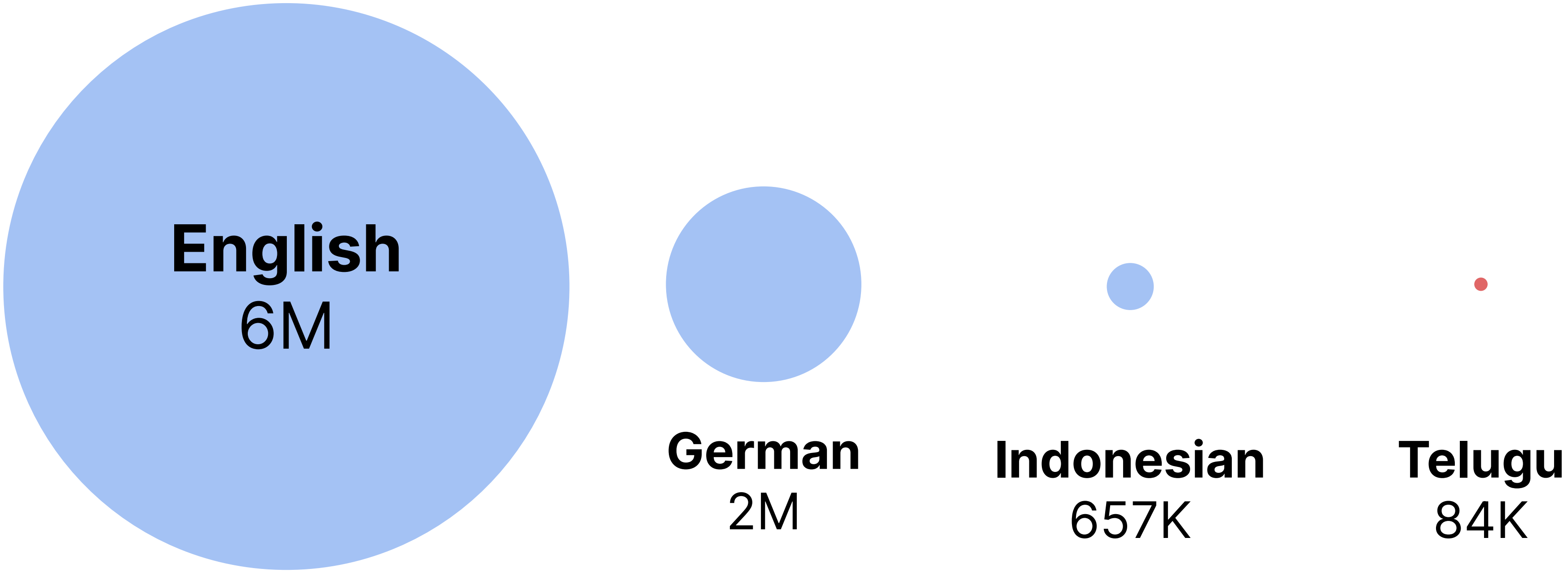
Easier translation of articles between languages.

Reusing work done by another community (notability, verifiability....) lowers the risk of deletion.



It also expands the number of people who can contribute, as it requires a different set of skills compared to writing completely new content.

Language gap



All translations

Saved 3 minutes ago

Publish

Plants in space

English

view page

4 categories

Use dmy dates

Zinnia flower in space

Plants in space are plants grown in outer space typically in a weightless but pressurized controlled environment in specific space gardens.[1] In the context of human spaceflight, they can be consumed

Plants in space

español

No categories

+ Add translation

1 Click paragraphs to translate

You don't need to add them all

2 Make the text read naturally

Machine translation is useful but you'll need to revise the text to make it accurate.

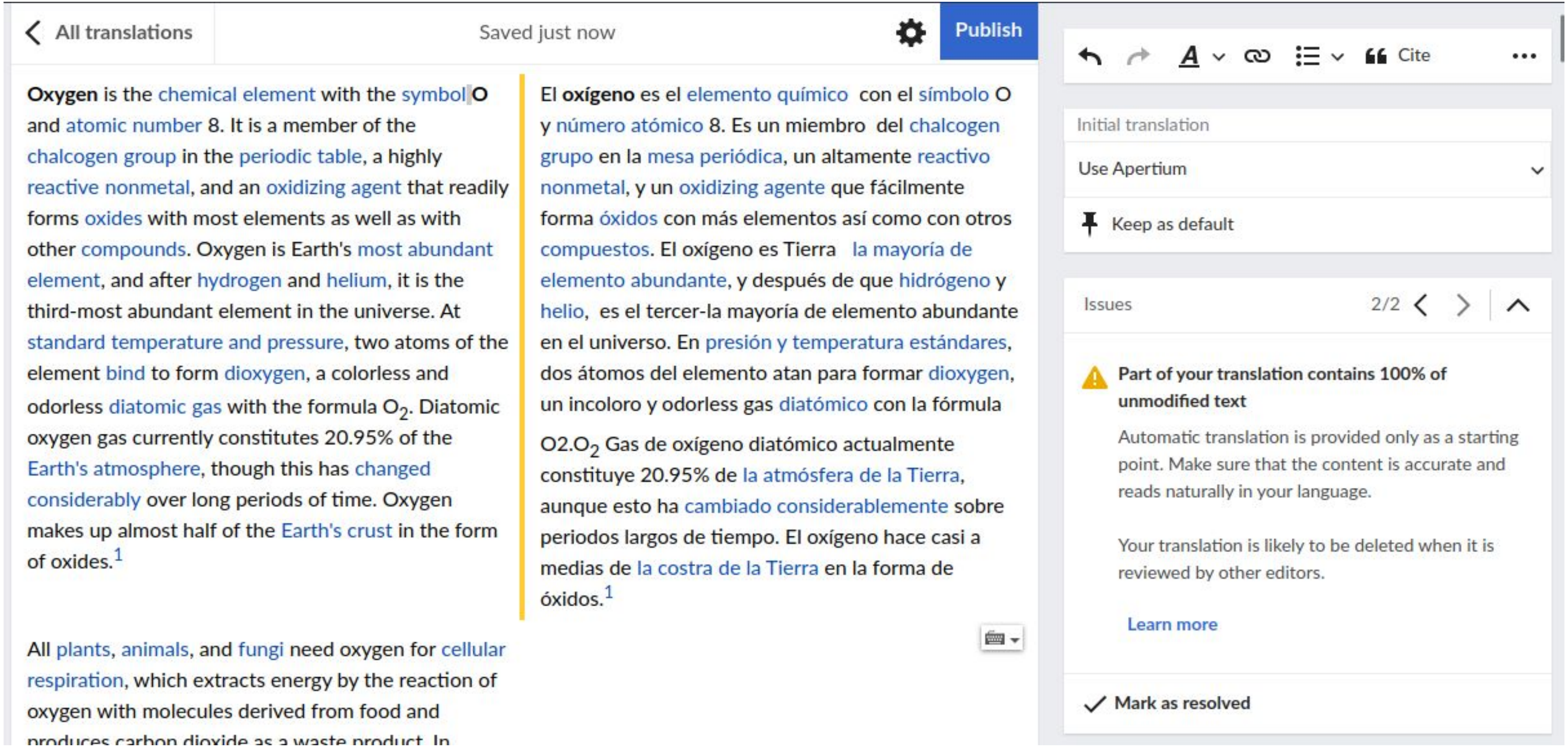
3 Publish the translation

When you are happy with the result, press "Publish".

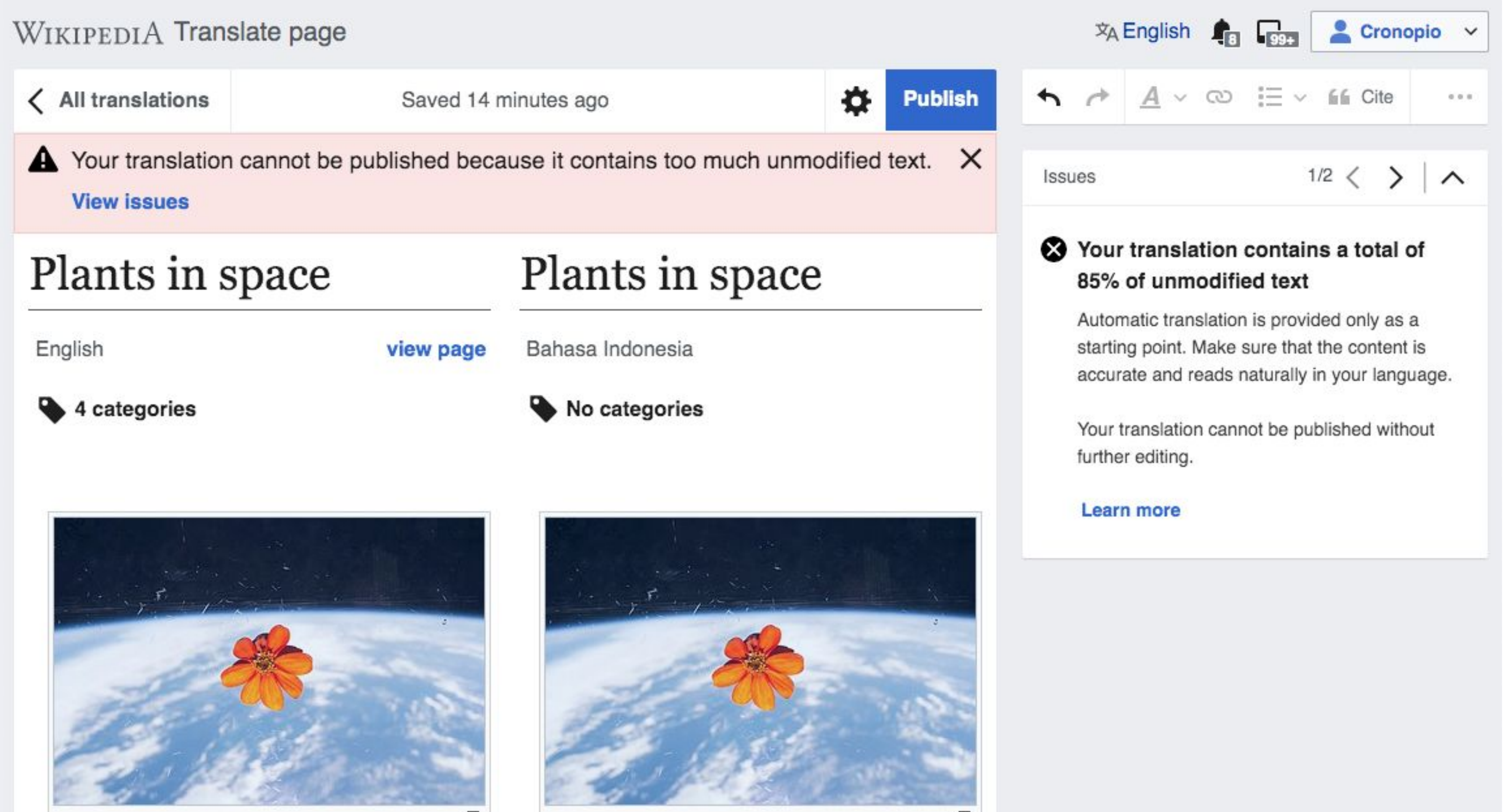
View translation guidelines

Provide feedback

Human curation of Machine translation



Machine Translation misuse prevention



Content Translation **Impact**

1.6 Million+

Articles published by translating

Combined, this would be a top 10 wikipedia

4%

Low deletion rate

Compared with **13%** deletion rate of articles created without translation

Machine translation services

Apertium

Google

Yandex

LingoCloud

Elia

MinT

MinT: Supporting underserved languages with open machine translation

13 June 2023 by Pau Giner

Translate This Post

Our vision is a world in which every single human being can freely share in the sum of all knowledge. Machine translation has the potential to help us achieve that vision by enabling more people to contribute content to Wikipedia in their native or preferred languages.

Content Translation, the tool used by Wikipedia editors to translate over one and a half million articles, uses machine translation as a starting point when it is available, making sure to keep humans in the loop by encouraging them to improve the initial translation and controlling how much it is edited. In this case, providing automation while keeping the humans in control helps Wikipedia editors to become more productive while producing quality content. However, not all languages have good quality machine translation available for editors to benefit from.

We are launching MinT in order to expand the current machine translation support. MinT ("Machine in Translation") is a new translation service by the Wikimedia Foundation

MinT

A self hosted Neural Machine Translation service by Wikipedia

Serves multiple MT models and provides a single API interface

- **NLLB**

  Generic model by Meta
- **NLLB-Wikipedia**

  Wikipedia Optimized models
- **OpusMT**

  For low resource languages
- **SoftCatala**

  For English-Catalan
- **IndicTrans2**

  for 22 indic languages and english

MinT

A self hosted Neural Machine Translation service by Wikipedia

Serves multiple MT models and provides a single API interface

198

Languages

35924

Language pairs

WIKIMEDIA FOUNDATION

02 Knowledge Integrity

AI article & edit quality assessment, vandalism patrol/prevention

Artificial Intelligence Aims to Make Wikipedia Friendlier and Better

The nonprofit behind Wikipedia is turning to machine learning to combat a long-standing decline in the number of editors.

By Tom Simonite

December 1, 2015

Software trained to know the difference between an honest mistake and intentional vandalism is being rolled out in an effort to make editing Wikipedia less psychologically bruising. It was developed by the Wikimedia Foundation, the nonprofit organization that supports Wikipedia.

ORES: Lowering Barriers with Participatory Machine Learning in Wikipedia

AARON HALFAKER*, Microsoft, USA
R. STUART GEIGER†, University of California, San Diego, USA

Algorithmic systems—from rule-based bots to machine learning classifiers—have a long history of supporting the essential work of content moderation and other curation work in peer production projects. From counter-vandalism to task routing, basic machine prediction has allowed open knowledge projects like Wikipedia to scale to the largest encyclopedia in the world, while maintaining quality and consistency. However, conversations about how quality control should work and what role algorithms should play have generally been led by the expert engineers who have the skills and resources to develop and modify these complex algorithmic systems. In this paper, we describe ORES: an algorithmic scoring service that supports real-time scoring of wiki edits using multiple independent classifiers trained on different datasets. ORES decouples several activities that have typically all been performed by engineers: choosing or curating training data, building models to serve predictions, auditing predictions, and developing interfaces or automated agents that act on those predictions. This meta-algorithmic system was designed to open up socio-technical conversations about algorithms in Wikipedia to a broader set of participants. In this paper, we discuss the theoretical mechanisms of social change ORES enables and detail case studies in participatory machine learning around ORES from the 5 years since its deployment.

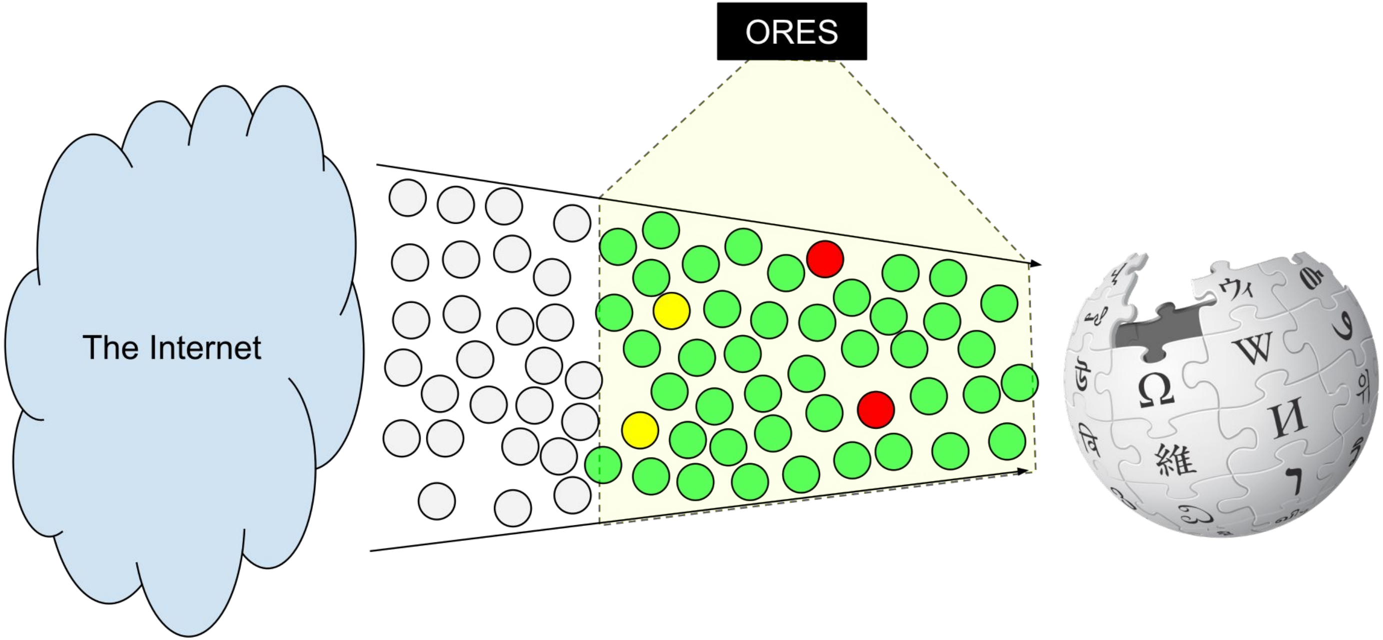
CCS Concepts: • **Networks** → **Online social networks**; • **Computing methodologies** → **Supervised learning by classification**; • **Applied computing** → **Sociology**; • **Software and its engineering** → **Software design techniques**; • **Computer systems organization** → Cloud computing;

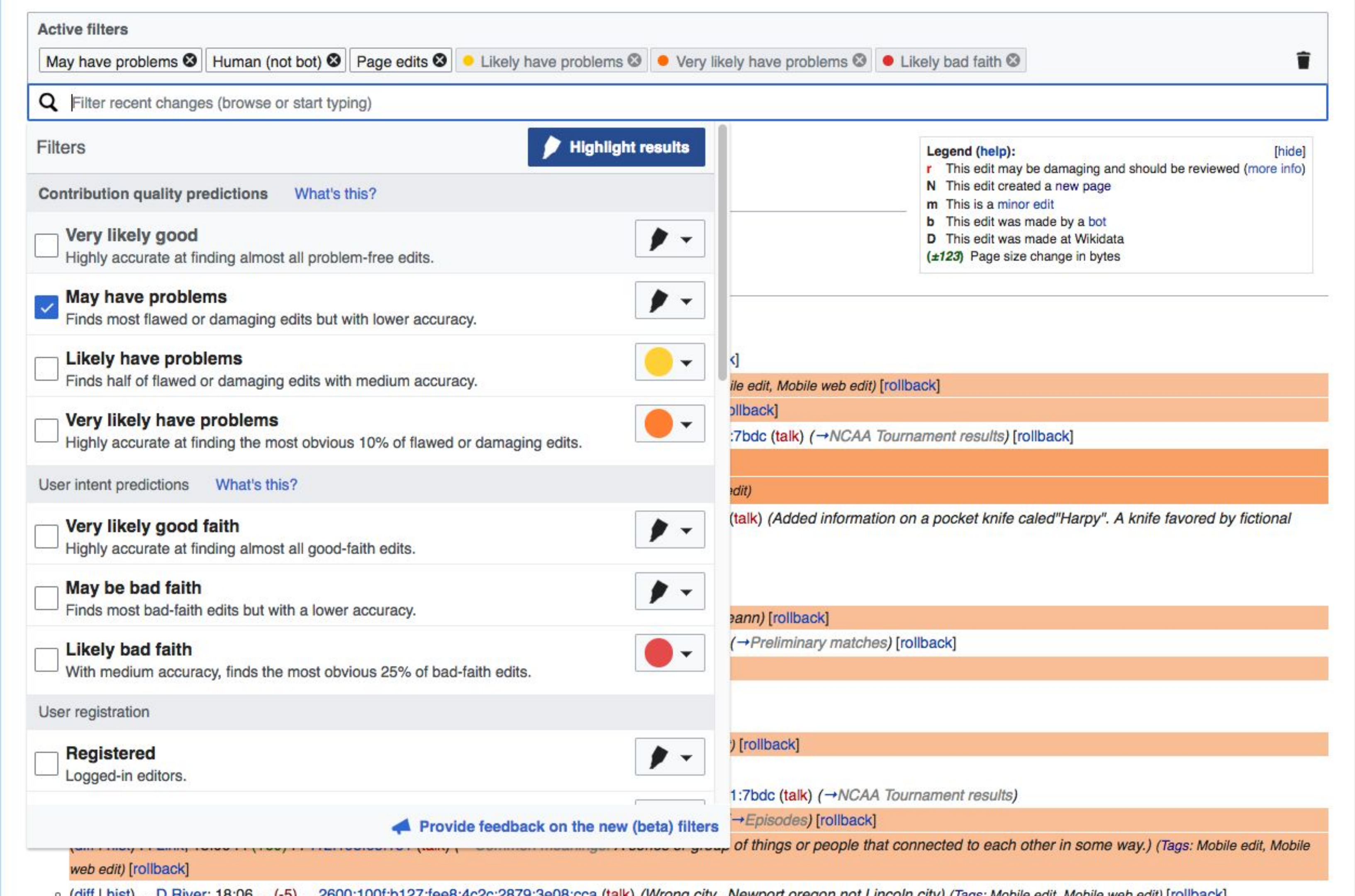
Additional Key Words and Phrases: Wikipedia; Reflection; Machine learning; Transparency; Fairness; Algorithms; Governance

ACM Reference Format:

Objective Revision Evaluation Service (ORES)

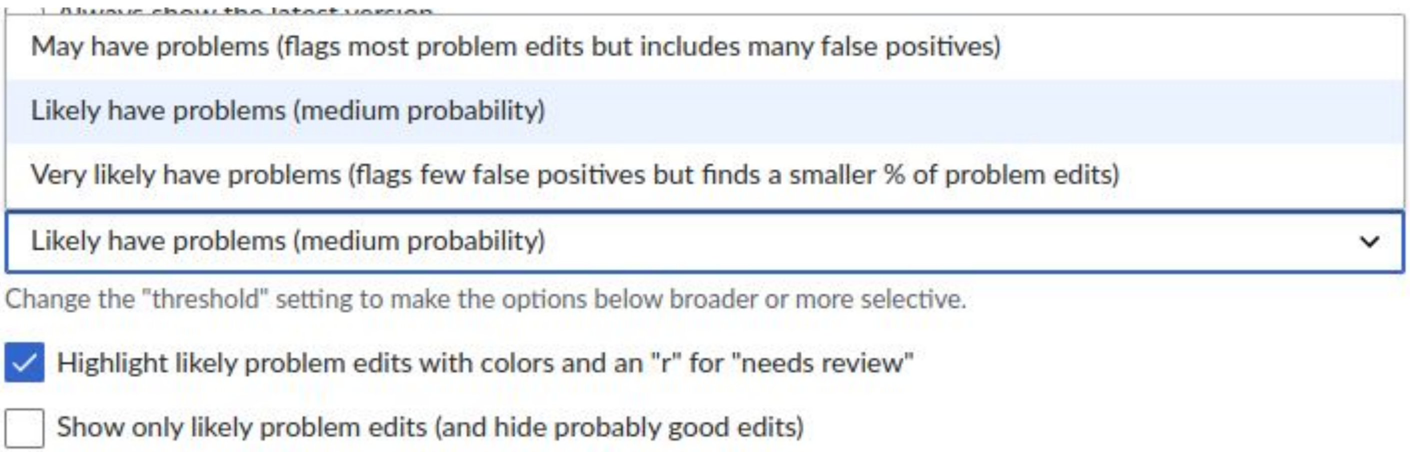
Edit quality: unknown good needs review damaging




Active filters
May have problems
Human (not bot)
Page edits
Likely have problems
Very likely have problems
Likely bad faith
Filter recent changes (browse or start typing)
Filters
Highlight results
Contribution quality predictions
What's this?
Very likely good
Highly accurate at finding almost all problem-free edits.
May have problems
Finds most flawed or damaging edits but with lower accuracy.
Likely have problems
Finds half of flawed or damaging edits with medium accuracy.
Very likely have problems
Highly accurate at finding the most obvious 10% of flawed or damaging edits.
User intent predictions
What's this?
Very likely good faith
Highly accurate at finding almost all good-faith edits.
May be bad faith
Finds most bad-faith edits but with a lower accuracy.
Likely bad faith
With medium accuracy, finds the most obvious 25% of bad-faith edits.
User registration
Registered
Logged-in editors.
Provide feedback on the new (beta) filters
Legend (help):
[hide]
r This edit may be damaging and should be reviewed (more info)
N This edit created a new page
m This is a minor edit
b This edit was made by a bot
D This edit was made at Wikidata
(±123) Page size change in bytes
ile edit, Mobile web edit) [rollback]
:7bdc (talk) (→NCAA Tournament results) [rollback]
(talk) (Added information on a pocket knife caled"Harpy". A knife favored by fictional
(→Preliminary matches) [rollback]
1:7bdc (talk) (→NCAA Tournament results)
→Episodes) [rollback]
of things or people that connected to each other in some way.) (Tags: Mobile edit, Mobile
web edit) [rollback]

Prediction Threshold preferences

Prediction Threshold



Revert Risk is now a service hosted in Lift Wing system

Rate limits
Reference
Revscoring Score object
Revert risk score object
Get revscoring goodfaith prediction
Get revscoring damaging prediction
Get revscoring reverted prediction
Get revscoring drafttopic prediction
Get revscoring draftquality prediction
Get revscoring articlequality prediction
Get revscoring articletopic prediction
Get reverted risk multilingual prediction

Get reverted risk language agnostic prediction

Discussion Updated 19 June 2023

POST /service/lw/inference/v1/models/revertrisk-language-agnostic:predict

The goal of this model is to detect revisions that might be reverted independently if they were made in good faith or with the intention of creating damage. Check the model card for more info.

Examples

curl Python

Anonymous access

```
# Get the revert risk probability for the edit on English Wikipedia identified by the revision id 123456.
$ curl https://api.wikimedia.org/service/lw/inference/v1/models/revertrisk-language-agnostic:predict -X
POST -d '{"rev_id": 123456, "lang": "en"}'
```

Logged in access

```
# Get the revert risk probability for the edit on English Wikipedia identified by the revision id 123456.
$ curl https://api.wikimedia.org/service/lw/inference/v1/models/revertrisk-language-agnostic:predict -X
POST -d '{"rev_id": 123456, "lang": "en"}' -H "Authorization: Bearer YOUR_ACCESS_TOKEN"
```

Technology

Revert Risk

	Revert Risk Language Agnostic	Revert Risk Multilingual
Characteristics	• Can run in all Wikipedia Language Editions • Mainly Based on Meta-Data	• Can run in the top-47 Wikipedia Language Editions • Uses an LLM (**mBert**)
Training Data	Implicit Annotations (past reverts)	
Pros	• Fast • Light on resources usage • Covers all languages	• Advanced NLP power • Fair on IP Edits
Cons	• Lower accuracy on IP Edits. • Basic NLP power.	• Covers just 47 languages • Heavy on computation resources.

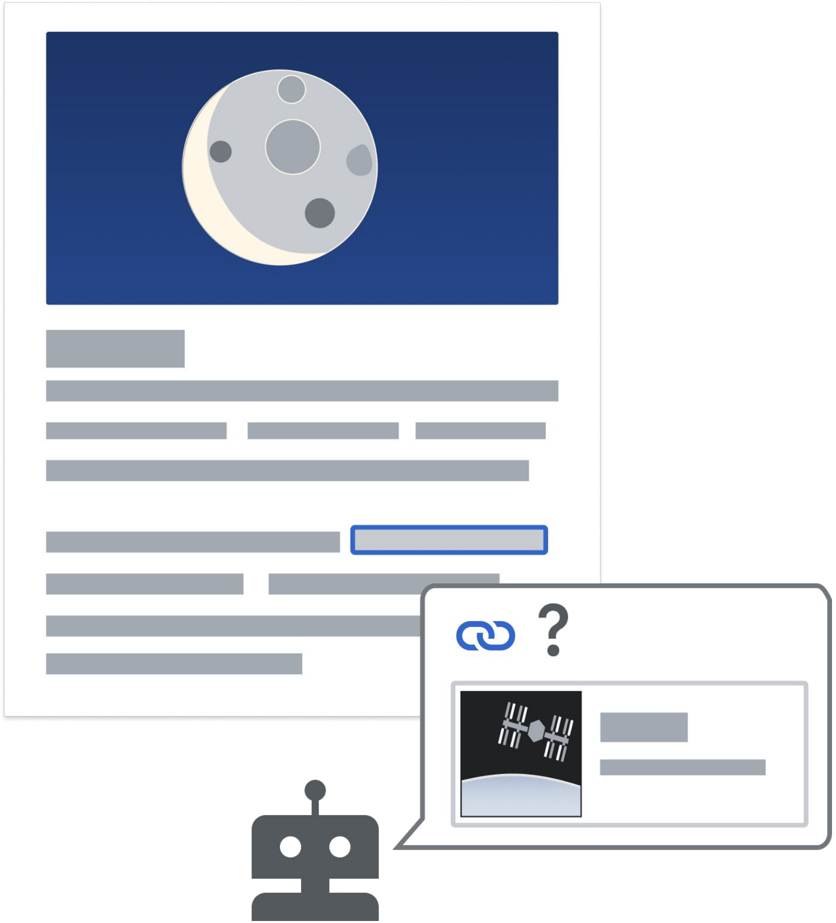
WIKIMEDIA FOUNDATION

03 Structured Tasks

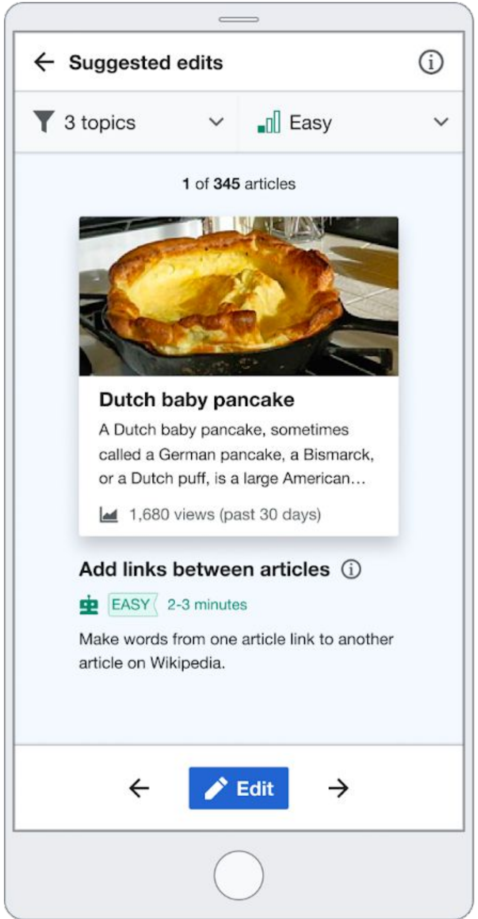
“add a link” and “add an image” to help new editors get started with easy tasks

Add a link
Newcomer task

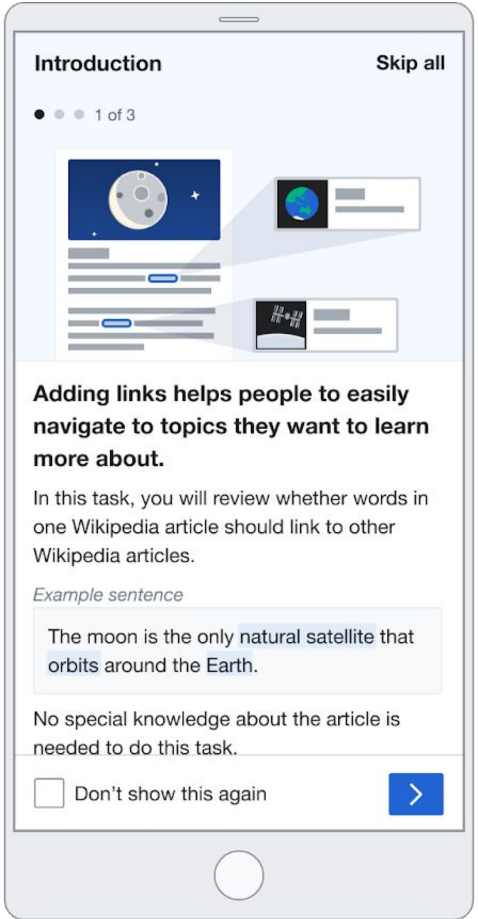
New editors review machine suggestions for making words in one Wikipedia article link to other Wikipedia articles.



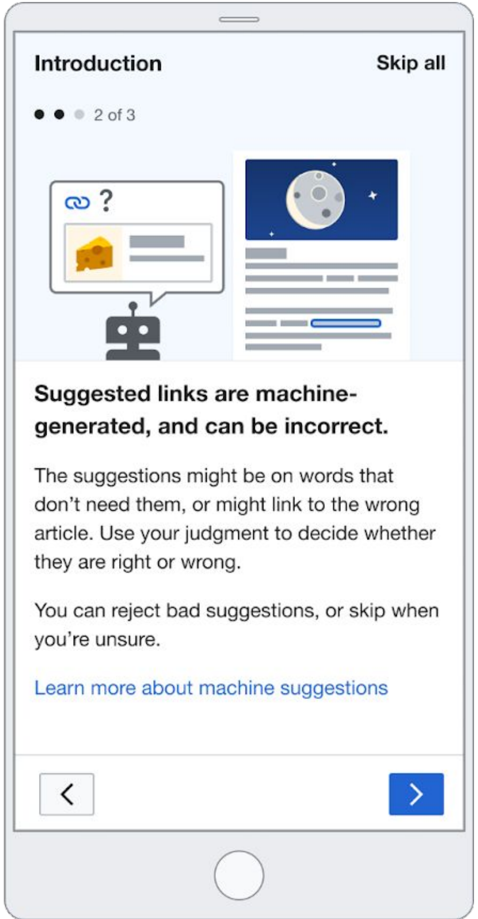
“Add a link” is available via the Suggested edits feed on Homepage



Onboarding 1: Explains value and impact of this small contribution

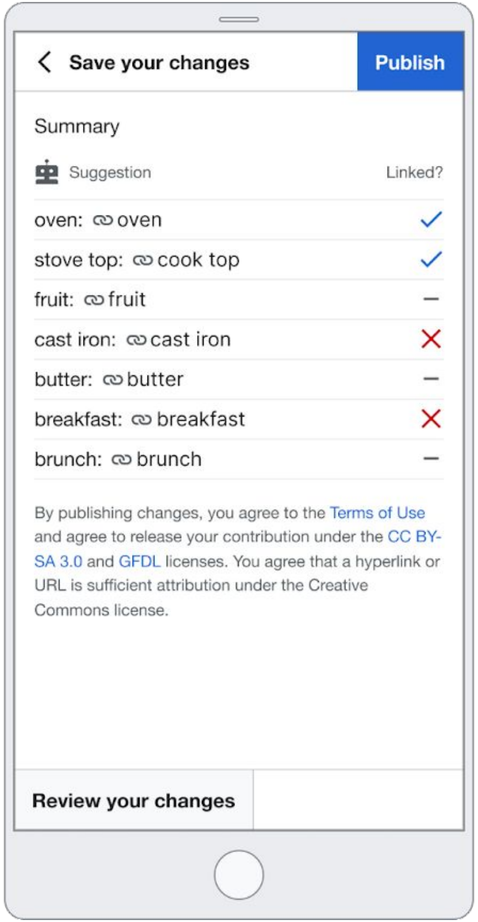


Onboarding 2: “Human in the loop” reviews machine suggestions

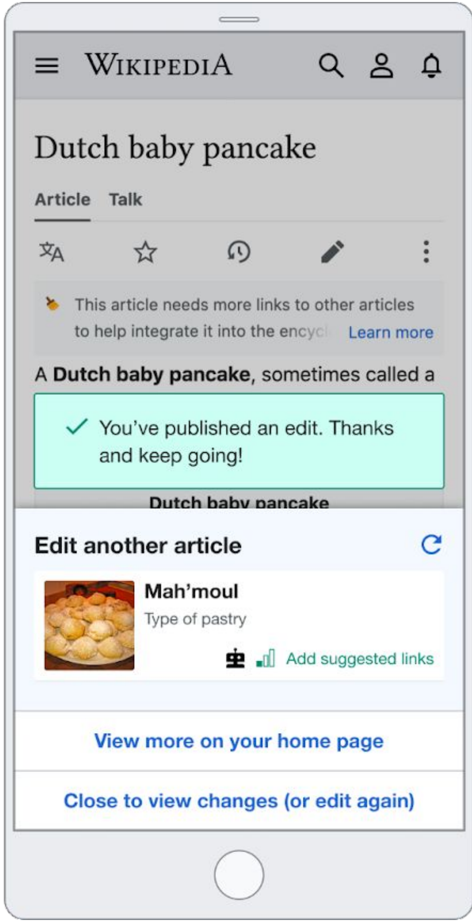


Evaluating machine suggestions of specific text to make into links…

…as an easy and fast way of contributing



Encouragement to do more post-edit



Add a link
Algorithm

Algorithm developed by the WMF Research team automatically generates link recommendations for Wikipedia articles.

The model's performance is evaluated based on precision and recall. Based on manual feedback from editors, *hard-coded rules* are implemented to avoid unwanted linking (e.g. links to dates).

Machine-learning model

The model predicts the probability of a link in the article (anchor-text + target-page).

- Identify *unlinked text* that could potentially contain a link
- Generate *candidate links* by looking up existing links with this text
- Score candidates and pick the most likely as the *target-page*

Training

The model is trained with existing sentences of millions of positive (what is linked) and negative examples (what is not linked).

Add a link

Impact

+17% **Activation**
increase in probability that a newcomer makes their first edit

+16% **Retention**
increase in probability that a new editor is retained

+18% **Productivity**
increase in the number of edits newcomers make during their first couple of weeks

-11% **Reverts**
decrease in revert rates compared to baseline newcomer edits (although this comparison is imperfect)

Add a link

Challenges

Moderation burden

Burden on patrollers: More edits = more work for patrollers.

Wider language support

Language characteristic and complexity affects parsing the sentences. ML models perform relatively poor on low resource languages

Data scarcity

Data scarcity for small wikis cause less performant ML models

WIKIMEDIA FOUNDATION

04 Optical Character Recognition

Document digitization

Optical Character Recognition

Tesseract	Self hosted Open source OCR engine
Transkribus	Externally hosted OCR system. Used for digitizing historical and handwritten documents relevant for Wikisource.
Google Cloud Vision OCR	External service

05 Lift Wing

Machine learning hosting platform

Lift Wing

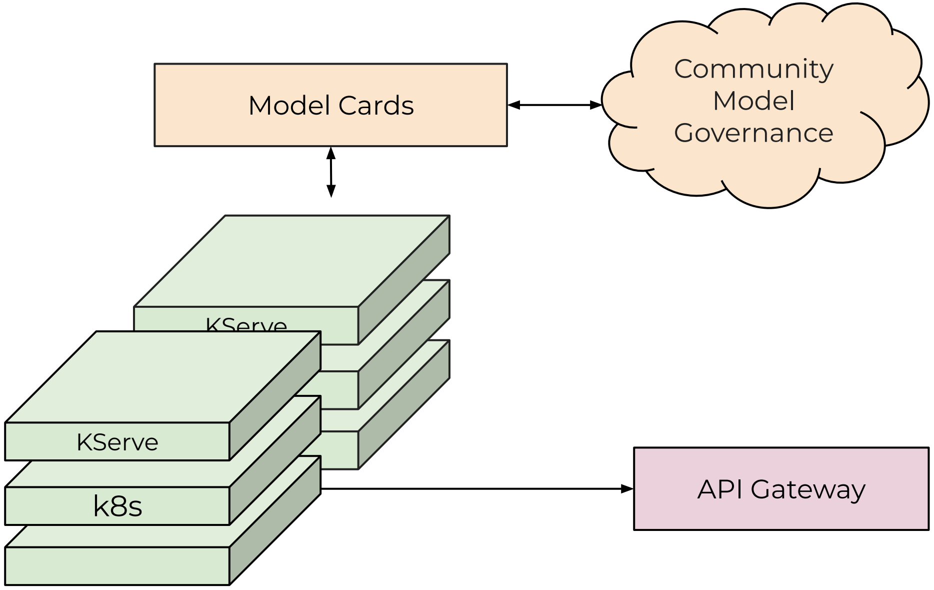
Scalability
Microservices can be independently scaled based on demand, allowing for more efficient resource utilization and improved performance.

Flexibility
Microservices architecture enables the use of different languages, and frameworks for each model service, providing greater flexibility in development.

Faster Deployment
Smaller codebases and independent deployment of microservices enable faster and more frequent releases, accelerating release to production

Fault Isolation:
Failure in one microservice is less likely to impact the entire system, improving overall system resilience and uptime.

Lift Wing



Lift Wing
Production environment

More machine learning use cases

Topic Classification system

Language agnostic link-based article topic classification - Label a given wikipedia article in any language to a topic

Language identification

Given a content snippet, this model can detect the language of the snippet. Supports ~200 languages.

Section alignment

Identify missing section between two existing article pairs in any languages. Used in Section translation feature of Content Translation

Third Party Machine Learning Services

Machine Translation	**Google**, **Yandex**, **Elia** machine translation services in Content Translation
Text to Speech	The Phonos extension to read IPA use **Google TTS API***
Machine Vision	Machine Vision use **Google's Cloud Vision API** to identify potential depicts statements for images in Commons.
Image to Text(OCR)	Wikisource use **Google's OCR API**, and **Transkribus**
Content moderation	Community Tech's CopyPatrol make use of **Turnitin**'s API for detecting plagiarism between passages added to Wikipedia and external documents
Named Entity Recognition	Architecture team used **Rosette** to identify Wikidata items from text

Model Cards

to make open source, transparent, human-centered machine learning

on-wiki model cards
for every model hosted
on WMF servers

- Use case, users
- Training data
- Ethical considerations
- Owners
- License
- Model architecture

How can we predict what general topic an article is in, and do so consistently across many languages? Answering this question would be useful for various analyses of Wikipedia dynamics. However, it is difficult to group a very diverse range of Wikipedia articles into coherent, consistent topics manually across all Wikipedia projects.

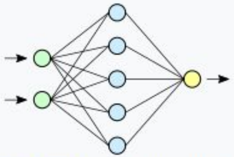
This model is a new, language-agnostic approach to predicting which topic an article might be relevant to. It uses the wikilinks in a given Wikipedia article to predict which (0 to many) of a set of 64 topics are relevant to a given article. For example, Mount Everest might reasonably be associated with South Asia, East Asia, Sports, Earth, and the Environment.

The training data for this model was over 30 million Wikipedia articles spanning all languages on Wikipedia. Each article was represented as the list of wikidata items associated with its outlinks. This data originated from the editing activities of Wikipedia and Wikidata editors, and was collected in an automated fashion.

This model is deployed on LiftWing. Right now, it can be publicly accessed through a beta testing site. This model may be useful for high-level analyses of Wikipedia dynamics (pageviews, article quality, edit trends), filtering articles, and cross-language analytics. It should not be used for projects outside of Wikipedia, namespaces outside of 0, disambiguations, or redirects.

Model card

This page is an on-wiki machine learning **model card**.



A model card is a document about a machine learning model that seeks to answer basic questions about the model.

Model Information Hub

Model creator(s)	Isaac Johnson, Martin Gerlach, and Diego Sáez-Trumper
Model owner(s)	WMF Research Team
Model interface	Beta Toolforge site
Past performance	Previous performance data
Publications	Language-agnostic Topic Classification for Wikipedia
Code	Github repository
Uses PII	No

This model uses **links in an article** to predict a **set of topics** that a **Wikipedia article** may be

Contents

- Motivation
- Users and uses
- Ethical considerations, caveats, and recommendations
- Model
  - Performance
    - Performance Notes
  - Implementation
- Data
- Licenses
- Citation

Machine learning at Wikipedia

Thank You

Santhosh Thottingal

WikipediaTutorials 0 0 Discusión Taller Preferencias Beta Lista de seguimiento Contribuciones Salir

Artículo Discusión Leer Editar código Editar Ver historial Buscar

WIKIPEDIA
La enciclopedia libre

- Portada
- Portal de la comunidad
- Actualidad
- Cambios recientes
- Páginas nuevas
- Página aleatoria
- Ayuda
- Donaciones
- Notificar un error

Imprimir/exportar
- Crear un libro
- Descargar como PDF
- Versión para imprimir

En otros proyectos
- Wikimedia Commons

Herramientas
- Lo que enlaza aquí
- Cambios en enlazadas
- Subir archivo
- Páginas especiales
- Enlace permanente
- Información de la página
- Elemento de Wikidata
- Citar esta página

Otros proyectos

Piedra de Rosetta

La **piedra de Rosetta** es un fragmento de una antigua estela egipcia de granodiorita inscrita con un decreto publicado en Menfis en el año 196 a. C. en nombre del faraón Ptolomeo V. El decreto aparece en tres escrituras distintas: el texto superior en jeroglíficos egipcios, la parte intermedia en escritura demótica y la inferior en griego antiguo. Gracias a que presenta esencialmente el mismo contenido en las tres inscripciones, con diferencias menores entre ellas, esta piedra facilitó la clave para el entendimiento moderno de los jeroglíficos egipcios.

Originalmente dispuesta dentro de un templo, la estela fue probablemente trasladada durante la época paleocristiana o la Edad Media y finalmente usada como material de construcción en un fuerte cerca de la localidad de Rashid (Rosetta), en el delta del Nilo. Allí fue hallada en 1799 por el soldado Pierre-François Bouchard durante la campaña francesa en Egipto. Las tropas británicas derrotaron a las francesas en Egipto en 1801 y la piedra original acabó en posesión inglesa bajo la Capitulación de Alejandría. Transportada a Londres, lleva expuesta al público desde 1802 en el Museo Británico, donde es la pieza más visitada.

Debido a que fue el primer texto plurilingüe antiguo descubierto en tiempos modernos, la Piedra de Rosetta despertó el interés público por su potencial para descifrar la hasta entonces ininteligible escritura jeroglífica egipcia, y en consecuencia sus copias litográficas y de yeso comenzaron a circular entre los museos y los eruditos europeos. La primera traducción completa del texto en griego antiguo apareció en 1803, pero no fue hasta 1822 cuando Jean-François Champollion anunció en París el descifrado de los textos jeroglíficos egipcios, mucho antes de que los lingüistas fueran capaces de leer con seguridad otras inscripciones y textos del antiguo Egipto. Los principales avances de [illegible] que el texto demótico usa caracteres fo[illegible]éticos para escribir nombres extranjeros (1802), que el texto jeroglífico también lo hace así y tiene similitudes generales con el demótico (Thomas Young en 1814) y que, además de ser usad[illegible] egipcias (Champollion entre 1822 y 1824).

Más tarde se descubrieron dos copias fragmentarias del mismo decreto, y en la actualidad se conocen varias inscripciones egipcias bilingües y trilingües, incluidos dos decretos Ptolemaicos, como el Decreto de Canopus del 238 a. C. y el Decreto de Menfis de Ptolomeo IV, c. 218 a. C. Por ello, aunque la Piedra de Rosetta ya no es única, fue un

Piedra de Rosetta

La piedra de Rosetta exhibida en el Museo Británico.

[illegible]	Granodiorita
Altura	112,3 cm
[illegible]	75,7 cm
[illegible]	28,4 cm
Peso	760 kg
Inscripción	Decreto de Ptolomeo V en tres escrituras diferentes.
Realización	196 a. C.

If a Wikipedia article exists in one language but doesn't exist in another,