Automatic approaches to bridge knowledge gaps in Wikimedia projects

WMF Research Team

WIKIMEDIA
FOUNDATION

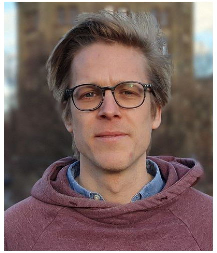
Wikimedia Research Showcase October 2021

The Research Team

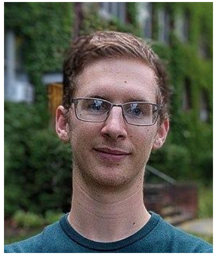
…and many formal collaborators

Leila Zia
Director, Head of Research

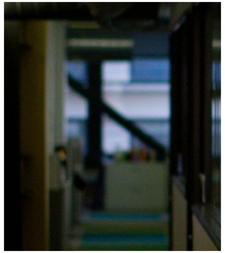
Pablo Aragón
Research Scientist



Martin Gerlach
Research Scientist



Isaac Johnson
Research Scientist



Fabian Kaelin
Senior Research Engineer

Emily Lescak
Senior Research Community Officer

Miriam Redi
Research Manager

Diego Sáez-Trumper
Senior Research Scientist

UNIVERSITY OF CAMBRIDGE

NATIONAL CHENG KUNG UNIVERSITY 1931

Università degli Studi di Torino

UNIVERSITY OF MICHIGAN

Telefonica

UNIVERSITY OF MINNESOTA

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

UNIVERSITY of WASHINGTON

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

جامعة نيويورك ابوظبي NYU ABU DHABI

Institute for Basic Science

Learn more: https://research.wikimedia.org

The strategic direction: Knowledge Equity
[from Movement Strategy]

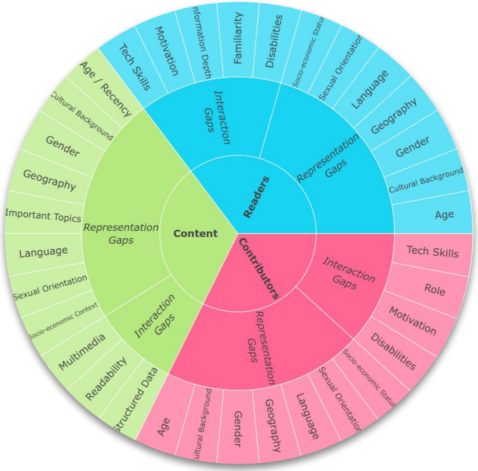
Knowledge equity: As a social movement, we will focus our efforts on the **knowledge and communities that have been left out by structures of power and privilege**. We will welcome people from every background to build strong and diverse communities. We will **break down the social, political, and technical barriers** preventing people from accessing and contributing to free knowledge.

The research program: Addressing Knowledge Gaps

Identify gaps

September 2020:

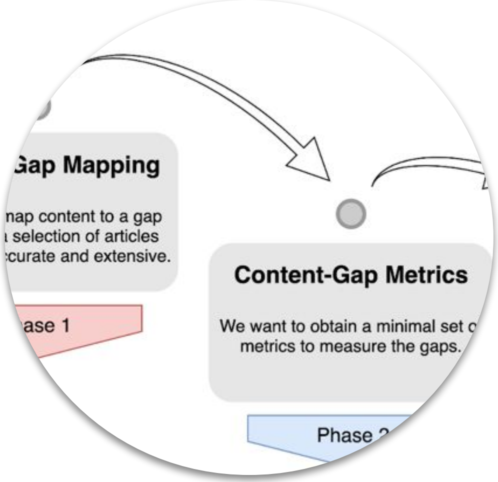
https://w.wiki/4DWU



Measure gaps

Ongoing:

https://w.wiki/4Exh

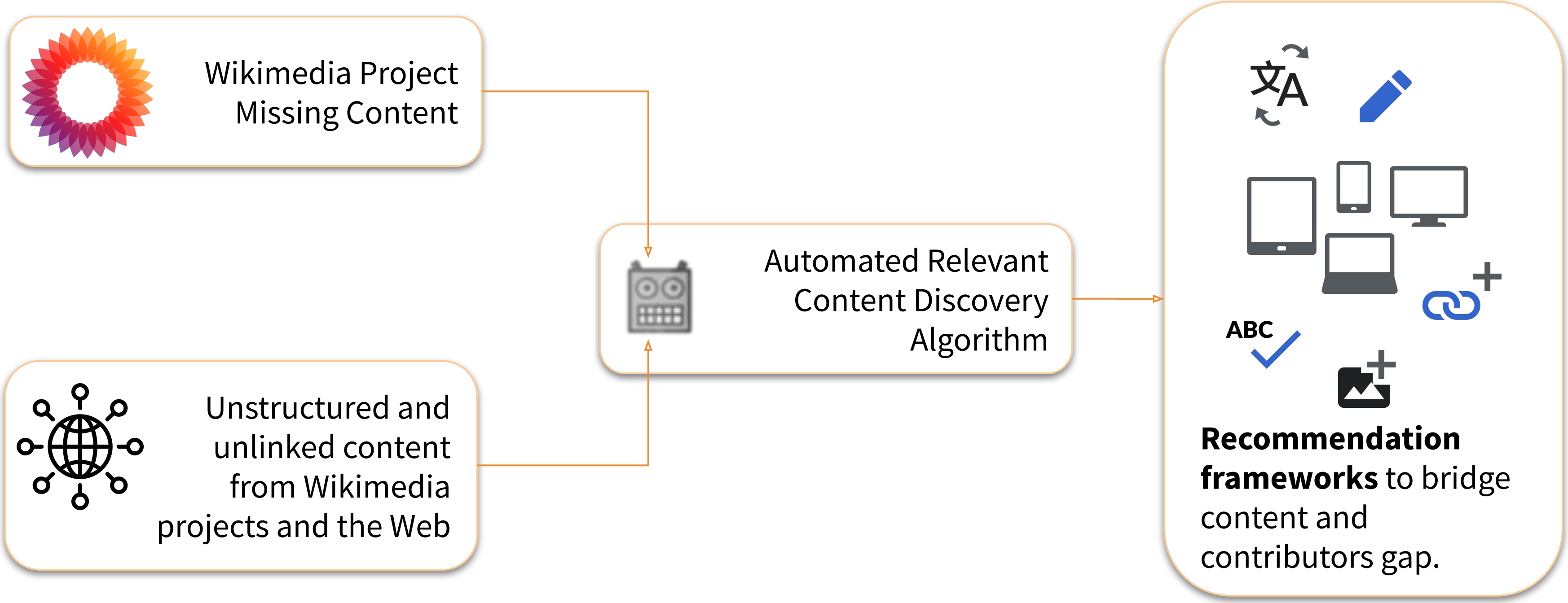


Bridging gaps

Today:

https://w.wiki/4DWe

Bridging knowledge gaps: Automated tools for content discovery and recommendation.



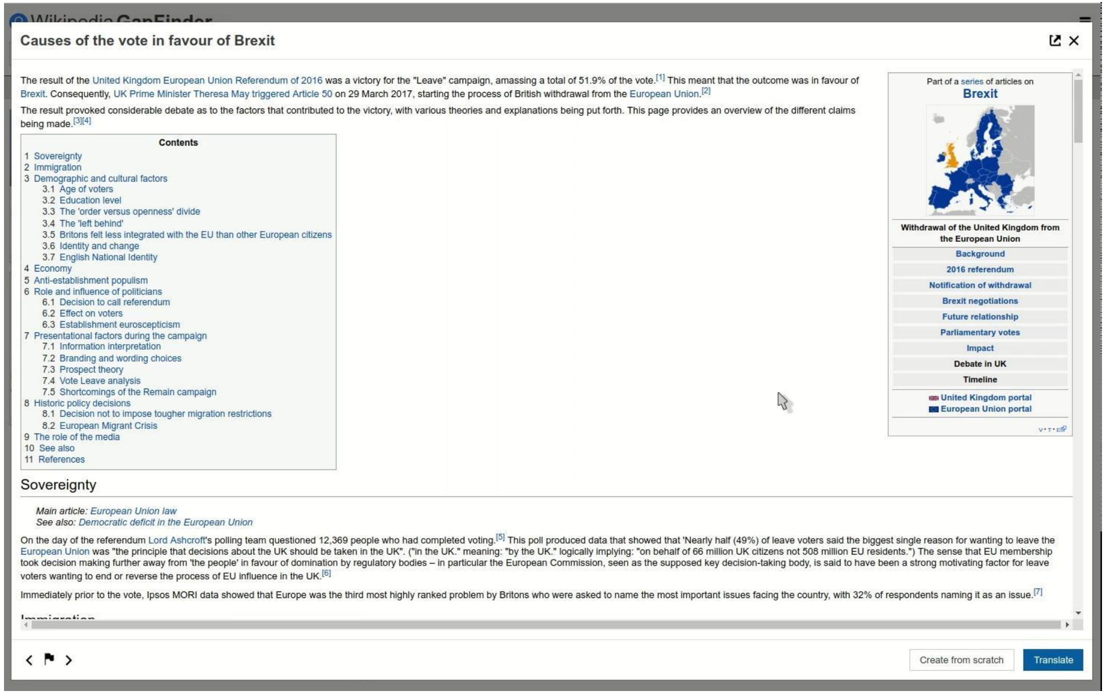
Wikimedia Project Missing Content

Unstructured and unlinked content from Wikimedia projects and the Web

Automated Relevant Content Discovery Algorithm

Recommendation frameworks to bridge content and contributors gap.

Bridging knowledge gaps:
Previous research work from our team



Causes of the vote in favour of Brexit

The result of the United Kingdom European Union Referendum of 2016 was a victory for the "Leave" campaign, amassing a total of 51.9% of the vote.[1] This meant that the outcome was in favour of Brexit. Consequently, UK Prime Minister Theresa May triggered Article 50 on 29 March 2017, starting the process of British withdrawal from the European Union.[2]

The result provoked considerable debate as to the factors that contributed to the victory, with various theories and explanations being put forth. This page provides an overview of the different claims being made.[3][4]

Contents

Part of a series of articles on
Brexit

Withdrawal of the United Kingdom from the European Union

Background
2016 referendum
Notification of withdrawal
Brexit negotiations
Future relationship
Parliamentary votes
Impact
Debate in UK
Timeline

United Kingdom portal
European Union portal

Sovereignty

Main article: European Union law
See also: Democratic deficit in the European Union

On the day of the referendum Lord Ashcroft's polling team questioned 12,369 people who had completed voting.[5] This poll produced data that showed that 'Nearly half (49%) of leave voters said the biggest single reason for wanting to leave the European Union was "the principle that decisions about the UK should be taken in the UK". ("in the UK." meaning: "by the UK." logically implying: "on behalf of 66 million UK citizens not 508 million EU residents.") The sense that EU membership took decision making further away from 'the people' in favour of domination by regulatory bodies – in particular the European Commission, seen as the supposed key decision-taking body, is said to have been a strong motivating factor for leave voters wanting to end or reverse the process of EU influence in the UK.[6]

Immediately prior to the vote, Ipsos MORI data showed that Europe was the third most highly ranked problem by Britons who were asked to name the most important issues facing the country, with 32% of respondents naming it as an issue.[7]

Bridging gaps:
Developing automatic tools

- Section alignment (Diego Sáez-Trumper) https://w.wiki/Bb6
    - In Collaboration with the Language team
- Link recommendation (Martin Gerlach) https://w.wiki/4BCP
    - In Collaboration with the Growth team
- Image recommendation (Miriam Redi) https://w.wiki/4DW8
    - In Collaboration with the Growth and Structured Data teams
- Equity in recommendations (Isaac Johnson) https://w.wiki/$4P

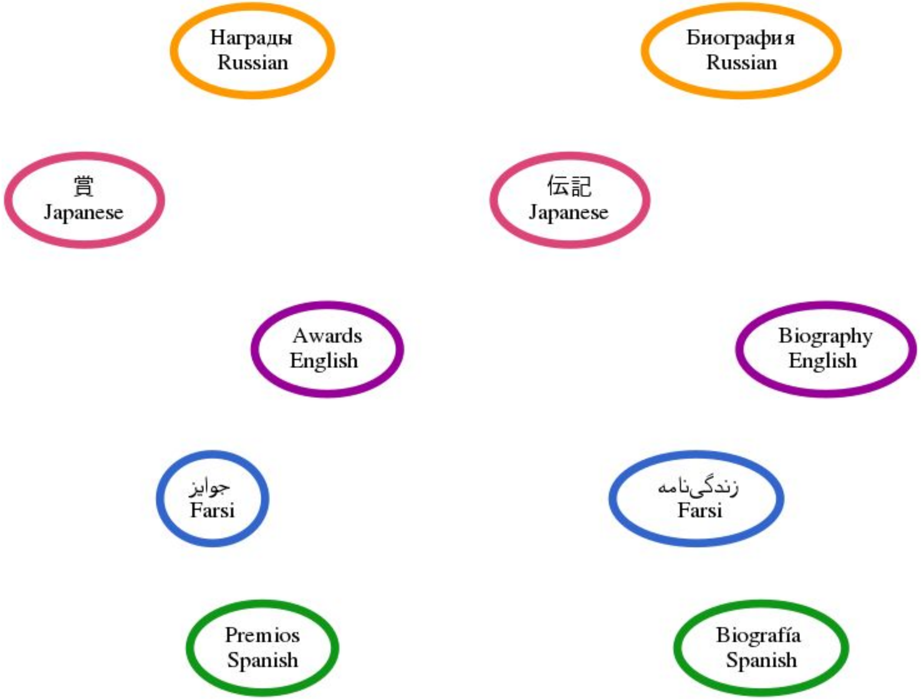
Section alignment

Diego

WIKIMEDIA
FOUNDATION

Problem: Translation engines can not be directly used on Wikipedia

- We want to compare sections and share content across languages.
- We can’t compare sections across languages.



Награды
Russian

Биография
Russian

賞
Japanese

伝記
Japanese

Awards
English

Biography
English

جوایز
Farsi

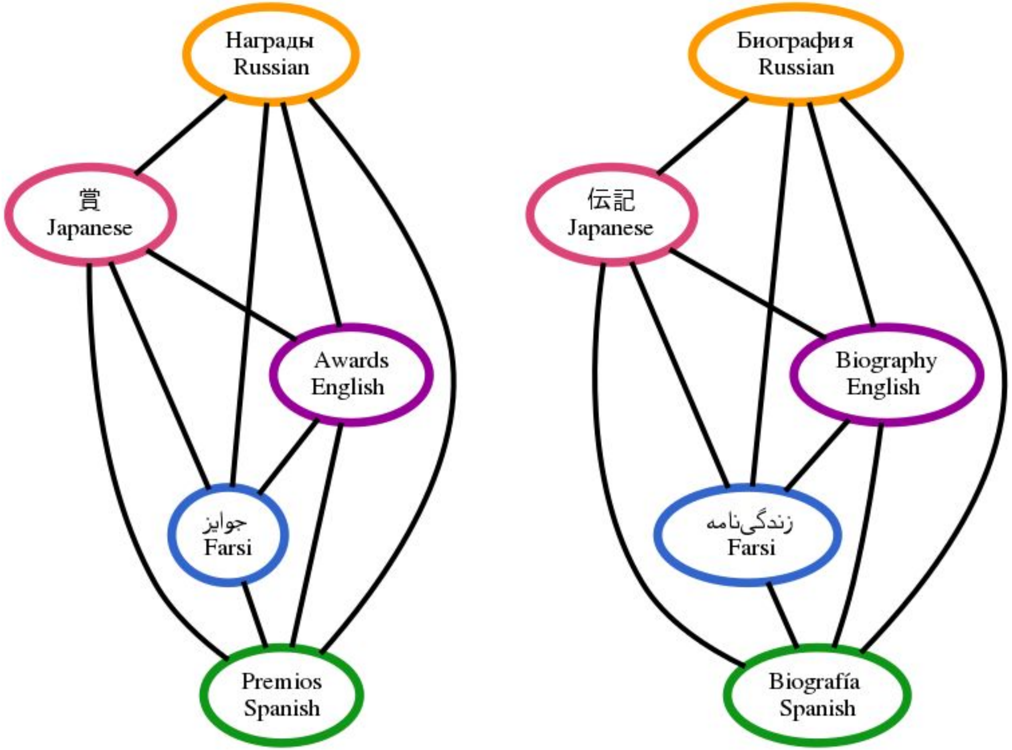
زندگی‌نامه
Farsi

Premios
Spanish

Biografía
Spanish

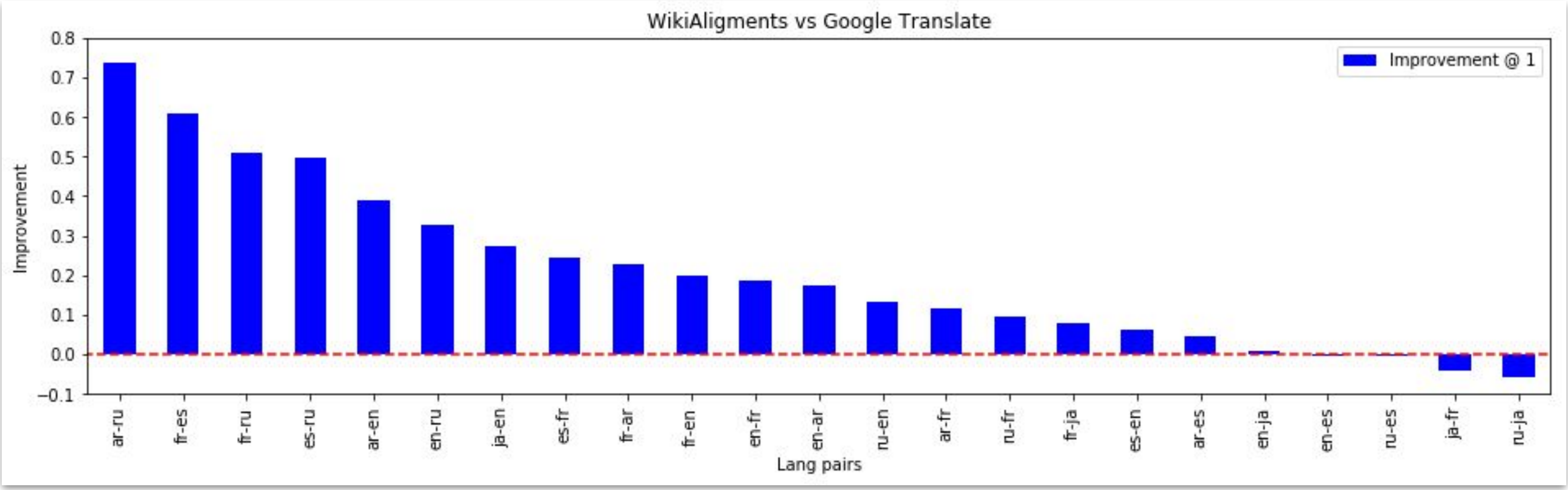
Solution: Cross-lingual section alignment

- Create an ML system to align sections across languages.



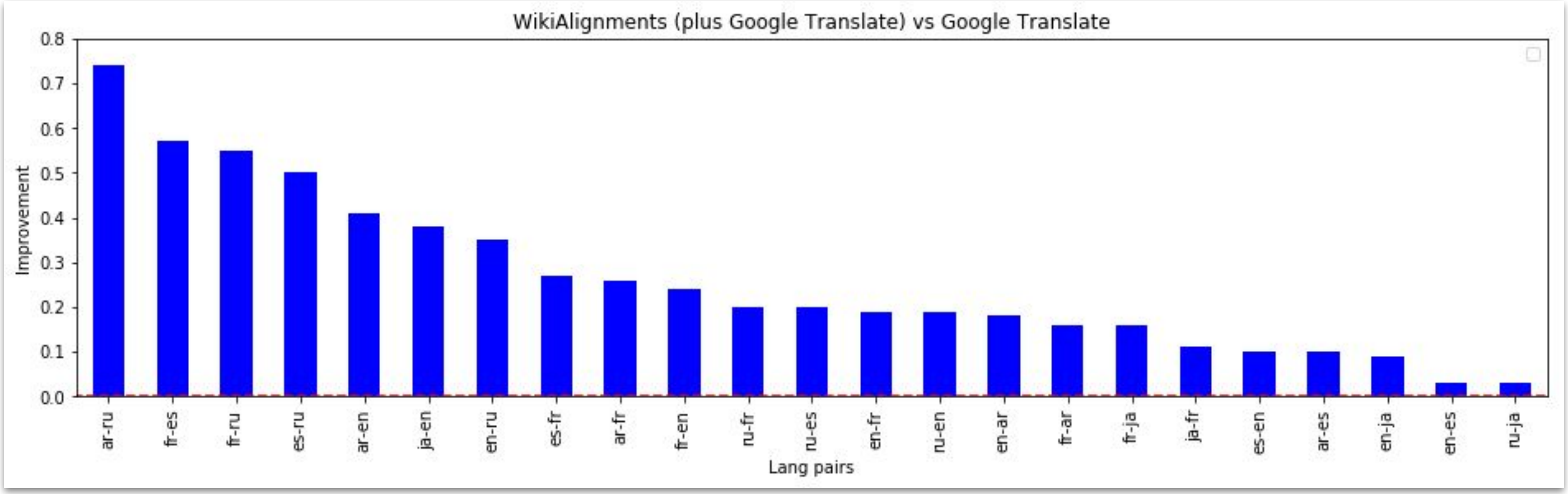
Step 1: Let’s try to do better than Automatic Machine Translation

- Ask the community to help annotate data:
  - ar,es,fr,ja,ru,en (all pairs)
- Cross-lingual word-embeddings + Wikipedia specific features.



Step 1: Let’s try to do better than Automatic Machine Translation

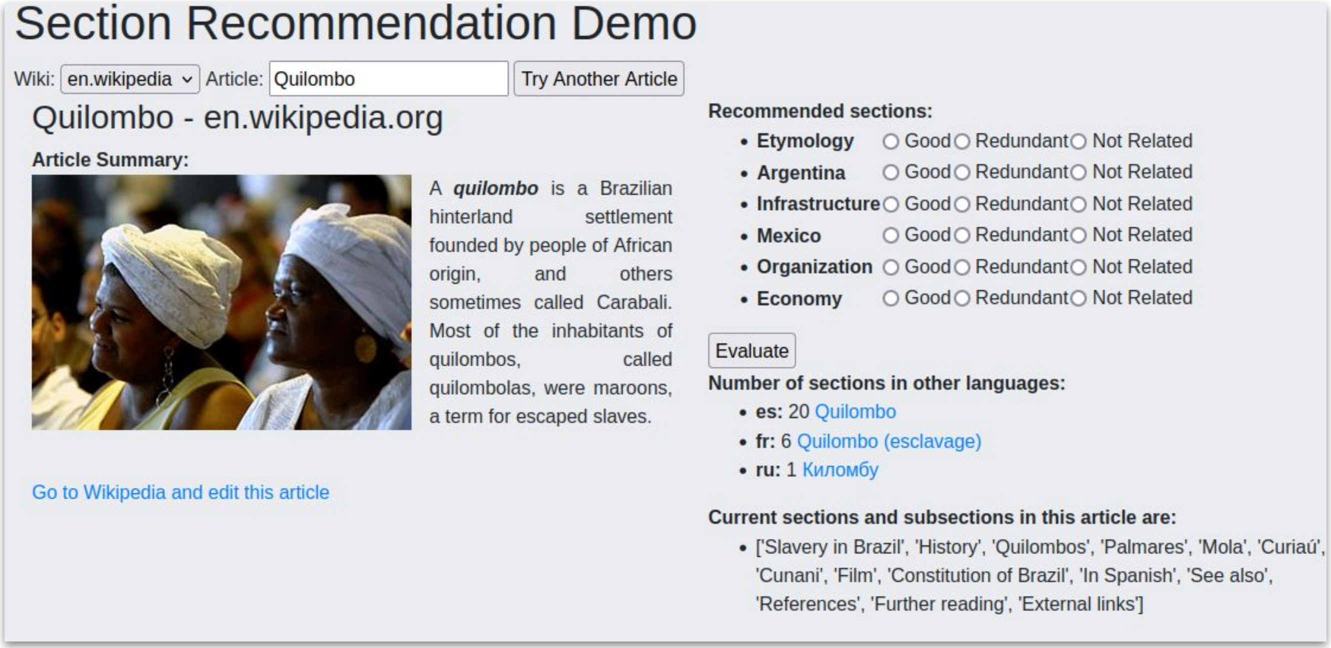
- Ask the community to help annotate data:
  - ar,es,fr,ja,ru,en (all pairs)
- Cross-lingual word-embeddings + Wikipedia specific features.



Learn more and test the system!

[[m:Research:Expanding_Wikipedia_articles_across_languages/Inter_language_approach]]

https://secrec.wmflabs.org



Section Recommendation Demo

Wiki: en.wikipedia Article: Quilombo Try Another Article

Quilombo - en.wikipedia.org

Article Summary:

A ***quilombo*** is a Brazilian hinterland settlement founded by people of African origin, and others sometimes called Carabali. Most of the inhabitants of quilombos, called quilombolas, were maroons, a term for escaped slaves.

Go to Wikipedia and edit this article

Recommended sections:

- **Etymology** ○ Good ○ Redundant ○ Not Related
- **Argentina** ○ Good ○ Redundant ○ Not Related
- **Infrastructure** ○ Good ○ Redundant ○ Not Related
- **Mexico** ○ Good ○ Redundant ○ Not Related
- **Organization** ○ Good ○ Redundant ○ Not Related
- **Economy** ○ Good ○ Redundant ○ Not Related

Evaluate

Number of sections in other languages:

- **es:** 20 Quilombo
- **fr:** 6 Quilombo (esclavage)
- **ru:** 1 Киломбу

Current sections and subsections in this article are:

- ['Slavery in Brazil', 'History', 'Quilombos', 'Palmares', 'Mola', 'Curiaú', 'Cunani', 'Film', 'Constitution of Brazil', 'In Spanish', 'See also', 'References', 'Further reading', 'External links']

Step 2: Scale up to 100+ languages

- Add new languages and deploy.
- Test new embeddings and language models.
- Improve the recommender system.

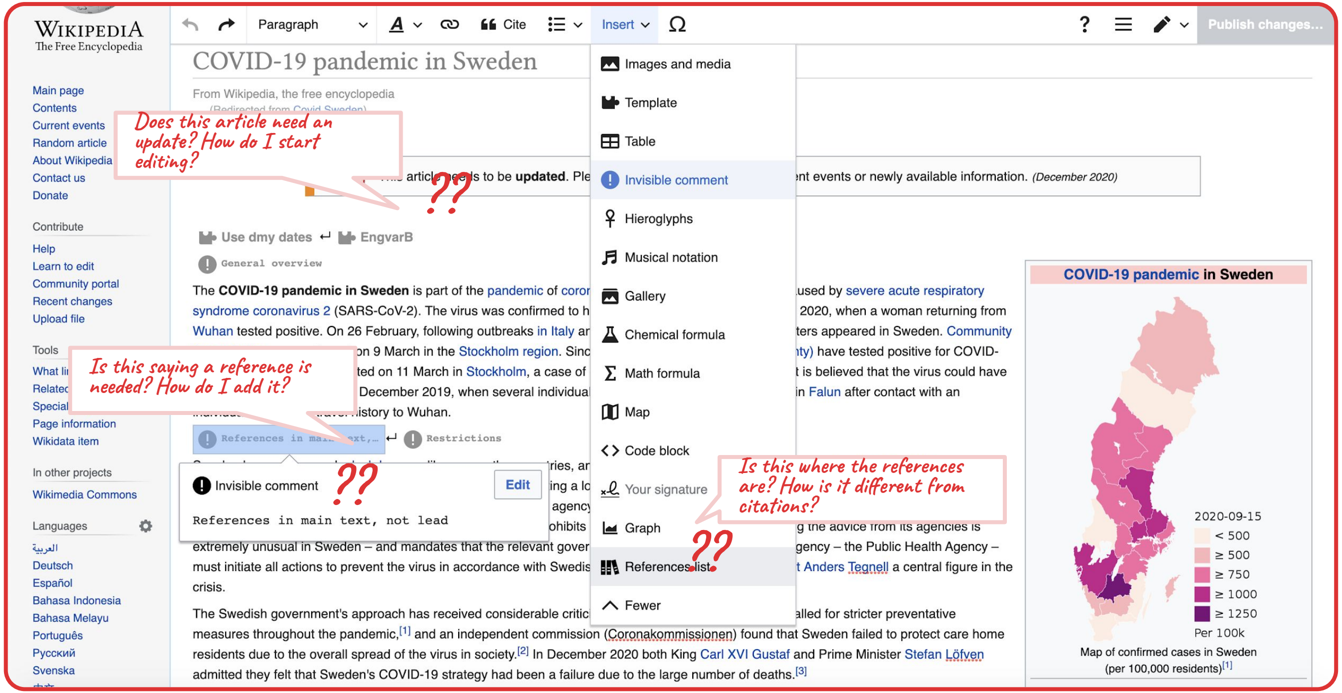
Link recommendation

Martin Gerlach (Research Team)
Growth Team (Marshall Miller, Rita Ho, Kosta Harlan, many more)
Djellel Difallah (NYU Abu Dhabi)

WIKIMEDIA
FOUNDATION

Problem

Editing is hard



COVID-19 pandemic in Sweden

From Wikipedia, the free encyclopedia

(Redirected from Covid Sweden)

This article needs to be updated. Ple... ...nt events or newly available information. *(December 2020)*

The **COVID-19 pandemic in Sweden** is part of the pandemic of coron... ...used by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus was confirmed to h... ...2020, when a woman returning from Wuhan tested positive. On 26 February, following outbreaks in Italy ar... ...ters appeared in Sweden. Community ...on 9 March in the Stockholm region. Sinc... ...ity) have tested positive for COVID- ...ted on 11 March in Stockholm, a case of... ...t is believed that the virus could have ...December 2019, when several individual... ...in Falun after contact with an ...travel history to Wuhan.

Invisible comment

References in main text, not lead

...ries, ar... ...ing a lo... ...agency ...ohibits... ...g the advice from its agencies is extremely unusual in Sweden – and mandates that the relevant gover... ...gency – the Public Health Agency – must initiate all actions to prevent the virus in accordance with Swedis... ...t Anders Tegnell a central figure in the crisis.

The Swedish government's approach has received considerable critic... ...alled for stricter preventative measures throughout the pandemic,[1] and an independent commission (Coronakommissionen) found that Sweden failed to protect care home residents due to the overall spread of the virus in society.[2] In December 2020 both King Carl XVI Gustaf and Prime Minister Stefan Löfven admitted they felt that Sweden's COVID-19 strategy had been a failure due to the large number of deaths.[3]

Map of confirmed cases in Sweden (per 100,000 residents)[1]

Does this article need an update? How do I start editing?

??

Is this saying a reference is needed? How do I add it?

??

Is this where the references are? How is it different from citations?

??

Technical

What is an infobox?

Conceptual

What is notability?

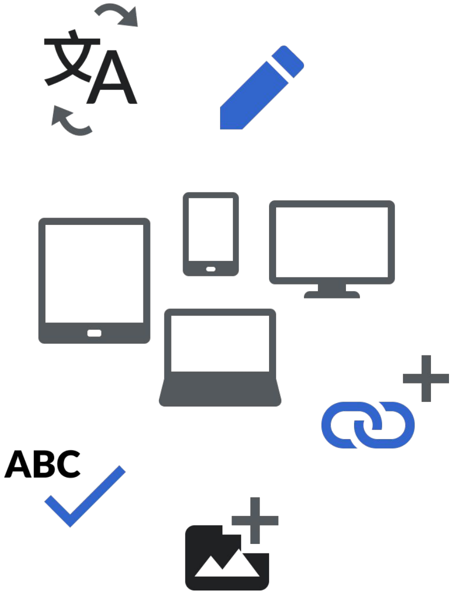
Cultural

Why are people so mean?

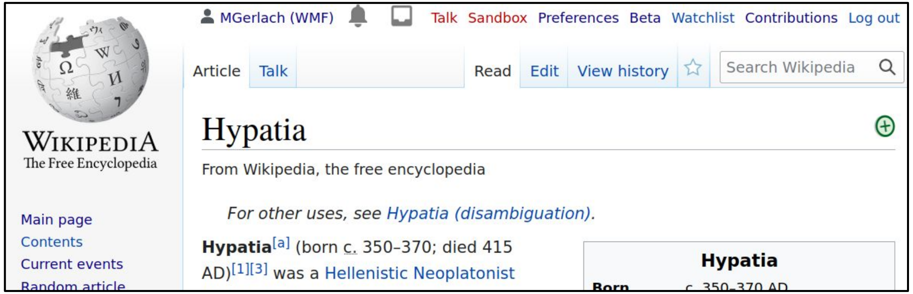
Solution

Structured task editing

- Break down editing into simpler tasks
  - Easier to: understand, do on mobile, get positive experience
- We picked the task of: "adding a link"
  - Well-defined, frequent, and attractive task type
- Machine-in-the-loop:
  - Generate recommendations using ML approaches.
  - Editors verify the output and validate the insertion.



Link recommendation



Hypatia

From Wikipedia, the free encyclopedia

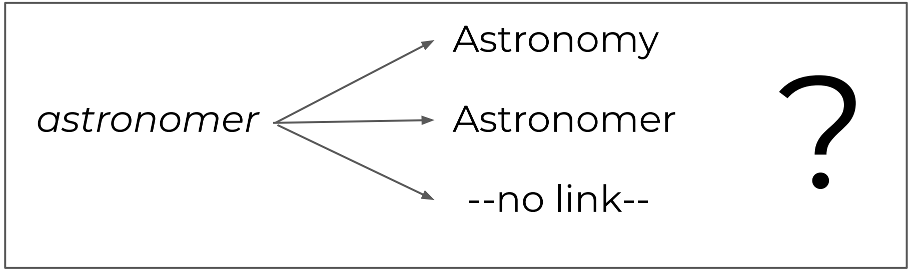
For other uses, see Hypatia (disambiguation).

Hypatia[a] (born c. 350–370; died 415 AD)[1][3] was a Hellenistic Neoplatonist

Hypatia

Born c. 350–370 AD

Hypatia (born c. 350–370; died 415 AD) was a Hellenistic Neoplatonist philosopher, astronomer, and mathematician, who lived in Alexandria, Egypt, then part of the Eastern Roman Empire.



Entity-linking task

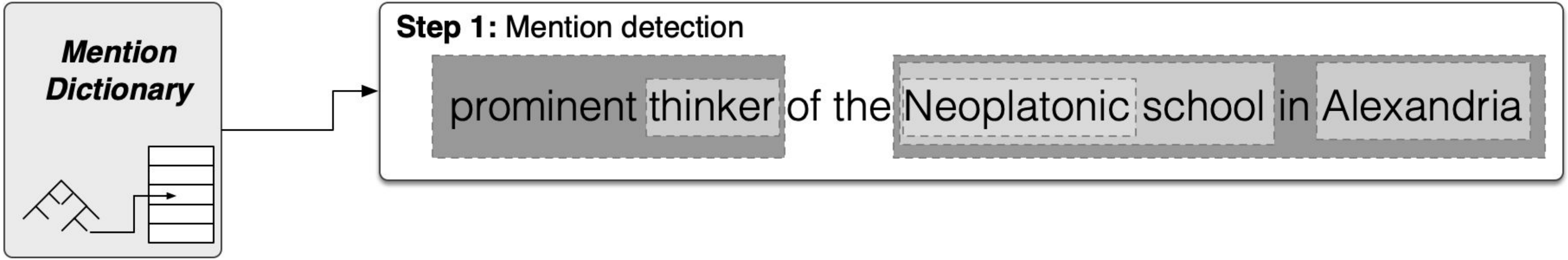
- Mention detection
- Link generation
- Link Disambiguation

The Add-a-link Task in Wikipedia

- Language support
  - 300+ language version of Wikipedia; highest impact for smaller communities.
- Other considerations
  - Manual of style constraints
  - Prefer simpler over complex models (scalability, transparency, etc)
  - Utility: find a balance between precision over recall

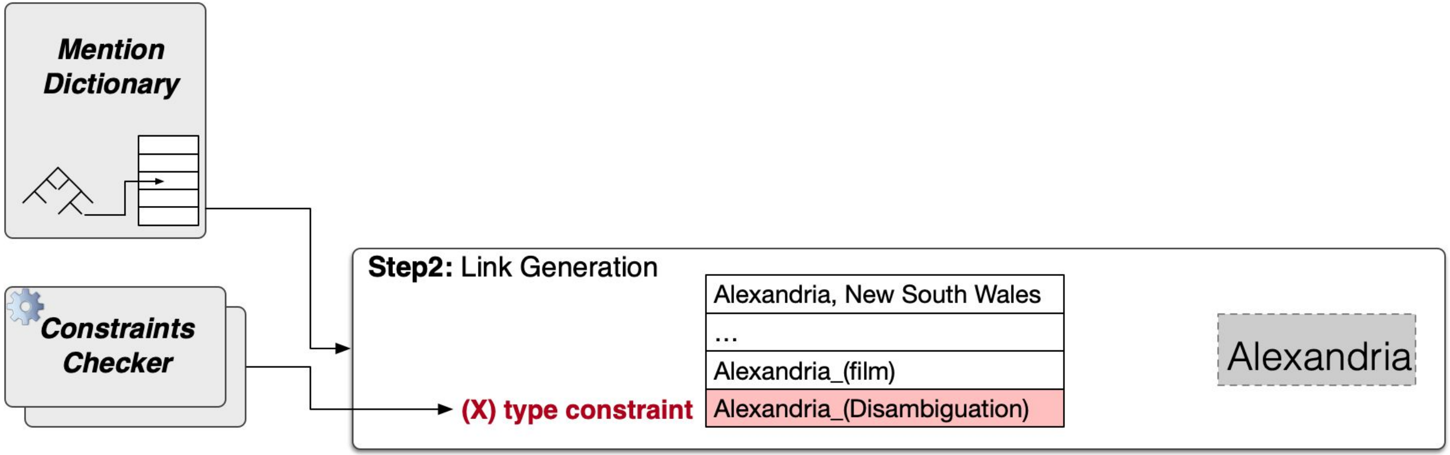
Step 1: Mention detection

- Build mention dictionary from all existing links
  - e.g. English Wikipedia: 7M anchors, 170M links
- String-matching sweeping window for all possible n-grams (N=1...10)
- Give preference to larger N


Mention Dictionary
Step 1: Mention detection
prominent thinker of the Neoplatonic school in Alexandria

Step 2: Link generation

- From the anchor-dictionary extract all used links
- Drops links based on constraints (type-based)
- Drop links based on link probability heuristic:
  - Text to Link ratio < 6.5% (picked empirically, and supported by previous work)

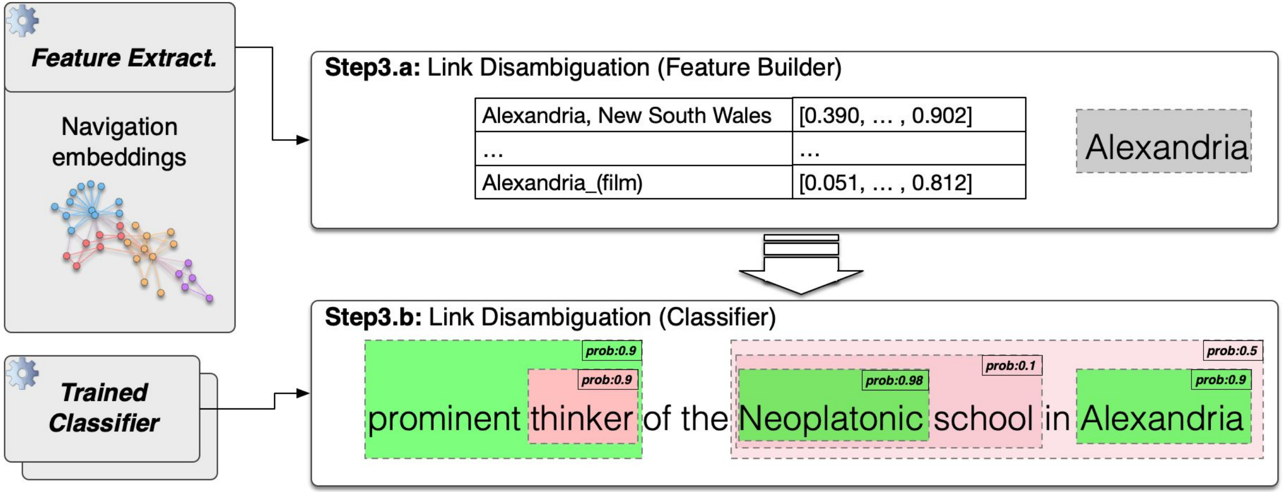


Step 3-a: Link disambiguation- features

- **N-gram size:** the number of tokens in the anchor (based on simple tokenization).
- **Frequency:** count of the anchor-link pair in the anchor-dictionary.
- **Ambiguity:** how many different candidate links exist for an anchor in the anchor-dictionary.
- **Kurtosis:** the kurtosis of the shape of the distribution of candidate-links for a given anchor in the anchor-dictionary
- **Levenshtein-distance:** a string similarity measure between the anchor and the link, e.g., the Levensthein-distance between “kitten” and “sitting” is 3.
- **Wiki2Vec Distance (entity embedding):** similarity between the article (source-page) and the link (target-page) based on the content of the pages.

Step 3-b: Link disambiguation- classifier

- Extract fully linked sentences from the lead sections
  - Positive example: linked mention with correct link
  - Negative example: linked mention with incorrect link, unlinked mention
- Train a binary classifier (XGBoost)



Step3.a: Link Disambiguation (Feature Builder)

Alexandria, New South Wales	[0.390, … , 0.902]
…	…
Alexandria_(film)	[0.051, … , 0.812]

Step3.b: Link Disambiguation (Classifier)

Evaluation

Held-out testset + **Manual evaluation**
(thanks: Bennoit Evellin, Habib Mhenni, Martin Urbanec, Bluetpp, -revi)

KaiO.Reid, CC BY-SA 4.0

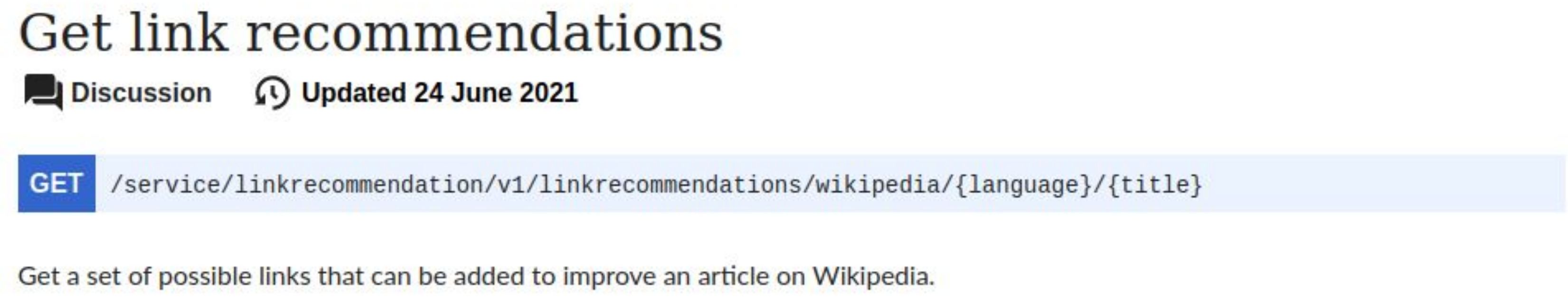
Tested Wikis: Arabic, Bengali, Czech, English, French, Vietnamese

Precision: 70% - 92%
How many suggestions are correct?

Recall: 30% - 66%
How many of the possible links captured?

Link recommendation model

- Training pipeline for each language.
- Models/datasets published publicly
- Link recommendation API on kubernetes



Get link recommendations

Discussion   Updated 24 June 2021

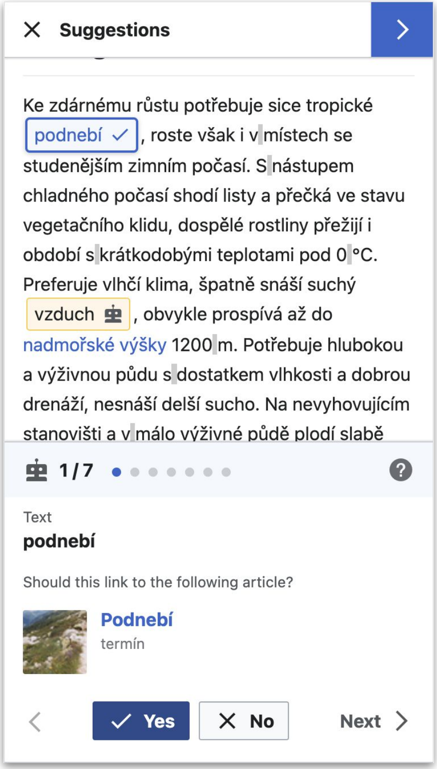
GET `/service/linkrecommendation/v1/linkrecommendations/wikipedia/{language}/{title}`

Get a set of possible links that can be added to improve an article on Wikipedia.

https://api.wikimedia.org/wiki/API_reference/Service/Link_recommendation

User interface

Evaluate the suggestion



✕ **Suggestions** >

Ke zdárnému růstu potřebuje sice tropické podnebí ✓, roste však i v místech se studenějším zimním počasí. S nástupem chladného počasí shodí listy a přečká ve stavu vegetačního klidu, dospělé rostliny přežijí i období s krátkodobými teplotami pod 0 °C. Preferuje vlhčí klima, špatně snáší suchý vzduch, obvykle prospívá až do nadmořské výšky 1200 m. Potřebuje hlubokou a výživnou půdu s dostatkem vlhkosti a dobrou drenáží, nesnáší delší sucho. Na nevyhovujícím stanovišti a v málo výživné půdě plodí slabě

1/7 ● ○ ○ ○ ○ ○ ○ ?

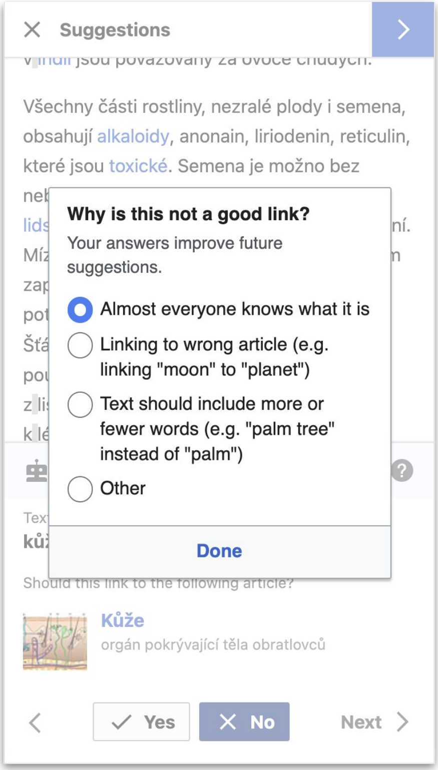
Text
podnebí

Should this link to the following article?

Podnebí
termín

< ✓ Yes ✕ No Next >

Feedback on algorithm



✕ **Suggestions** >

v Indii jsou považovány za ovoce chudých.

Všechny části rostliny, nezralé plody i semena, obsahují alkaloidy, anonain, liriodenin, reticulin, které jsou toxické. Semena je možno bez

Why is this not a good link?
Your answers improve future suggestions.

● Almost everyone knows what it is
○ Linking to wrong article (e.g. linking "moon" to "planet")
○ Text should include more or fewer words (e.g. "palm tree" instead of "palm")
○ Other

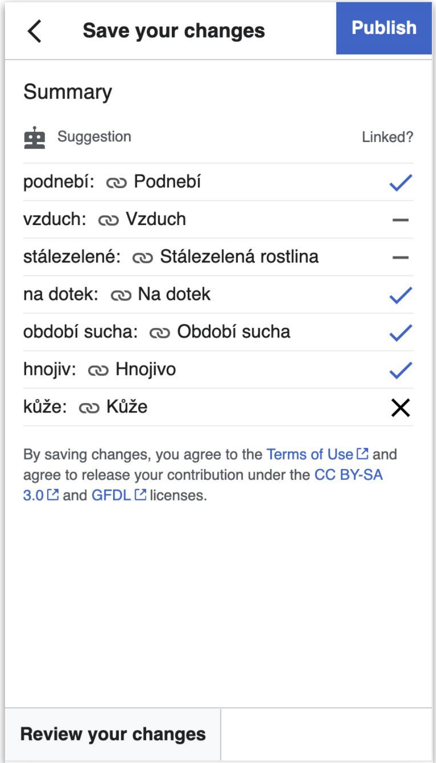
Done

Should this link to the following article?

Kůže
orgán pokrývající těla obratlovců

< ✓ Yes ✕ No Next >

Edit summary



< **Save your changes** **Publish**

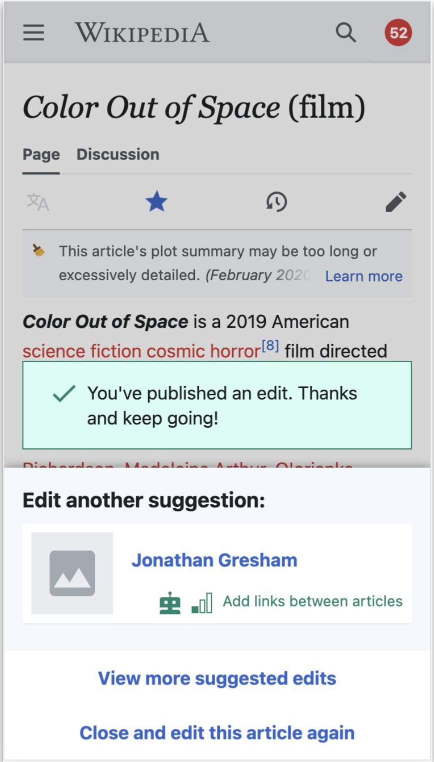
Summary

Suggestion	Linked?
podnebí: Podnebí	✓
vzduch: Vzduch	–
stálezelené: Stálezelená rostlina	–
na dotek: Na dotek	✓
období sucha: Období sucha	✓
hnojiv: Hnojivo	✓
kůže: Kůže	✕

By saving changes, you agree to the Terms of Use and agree to release your contribution under the CC BY-SA 3.0 and GFDL licenses.

Review your changes

Next suggestion



☰ WIKIPEDIA 52

Color Out of Space (film)

Page **Discussion**

This article's plot summary may be too long or excessively detailed. *(February 202* Learn more

Color Out of Space is a 2019 American science fiction cosmic horror[8] film directed

✓ You've published an edit. Thanks and keep going!

Edit another suggestion:

Jonathan Gresham
Add links between articles

View more suggested edits

Close and edit this article again

In practice

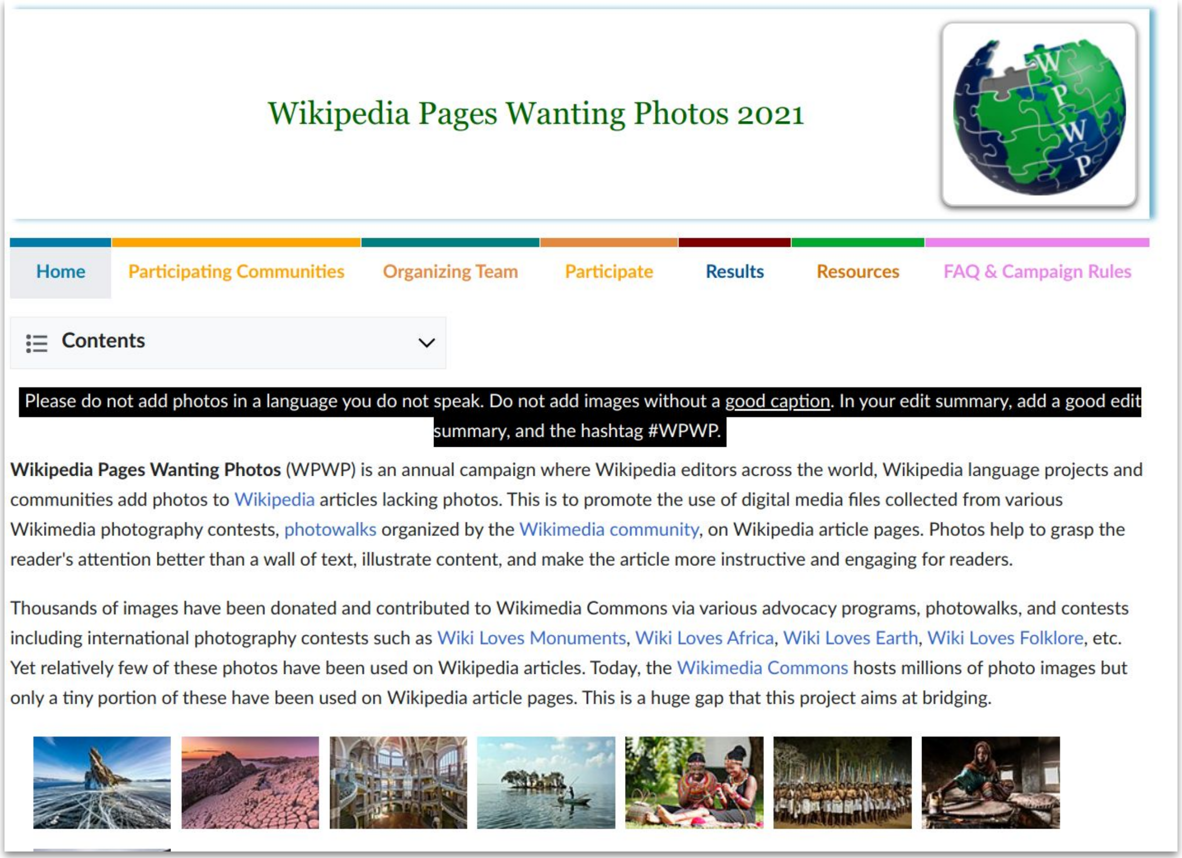
- Results from pilot-wikis (Arabic, Bengali, Czech, Vietnamese)
  - Newcomers prefer structured editing
  - Revert rate of structured edits is much lower (7.9% vs 25.5% for unstructured tasks)
  - Careless editing is rare
  - Reactions from community mostly positive
- Deployment
  - Currently: Arabic, Bengali, Czech, French, Hungarian, Persian, Polish, Romanian, Russian, Vietnamese
  - Planned: Catalan, Hebrew, Hindi, Korean, Norwegian, Portuguese, Simple English, Swedish, Ukrainian

Image recommendation

Miriam

WIKIMEDIA
FOUNDATION

Problem: **Wikipedia is missing images**



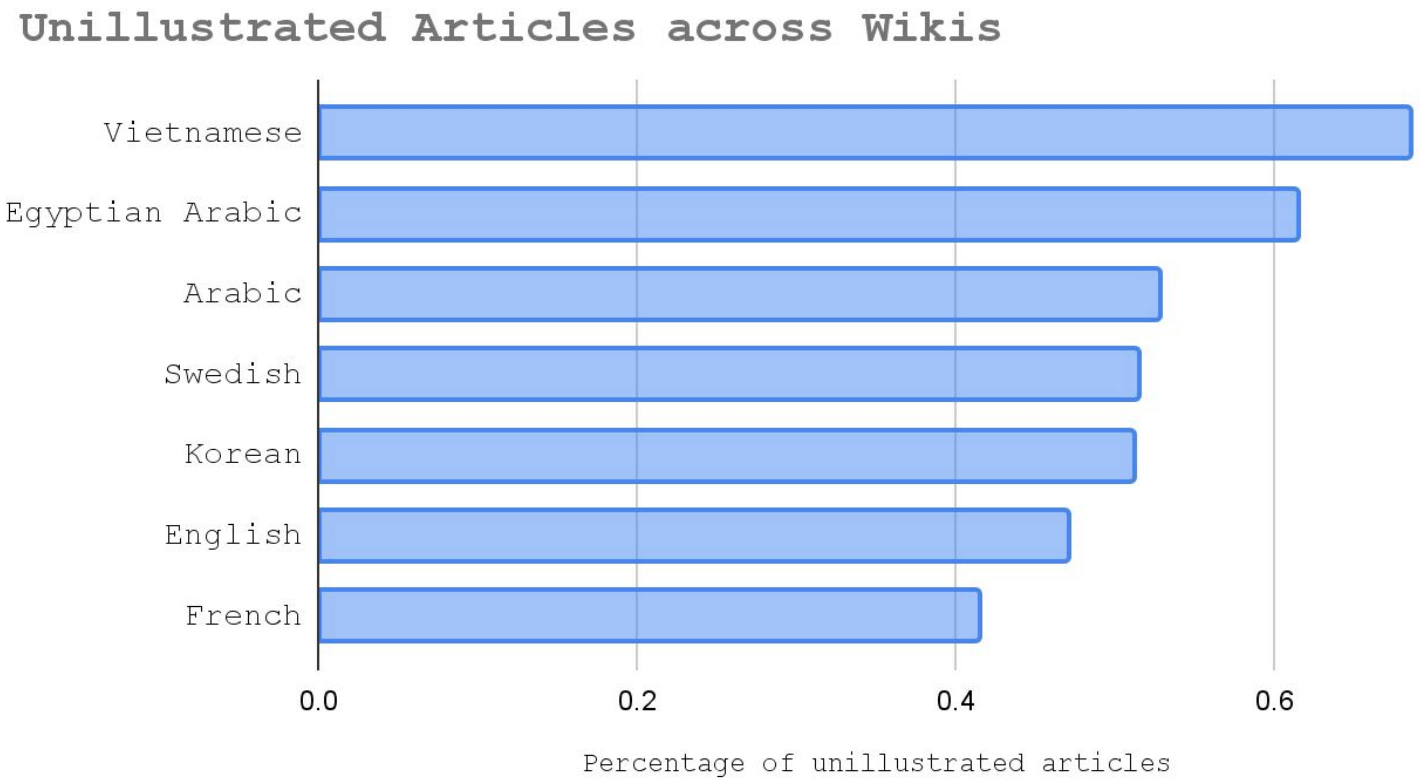
Wikipedia Pages Wanting Photos 2021

Please do not add photos in a language you do not speak. Do not add images without a good caption. In your edit summary, add a good edit summary, and the hashtag #WPWP.

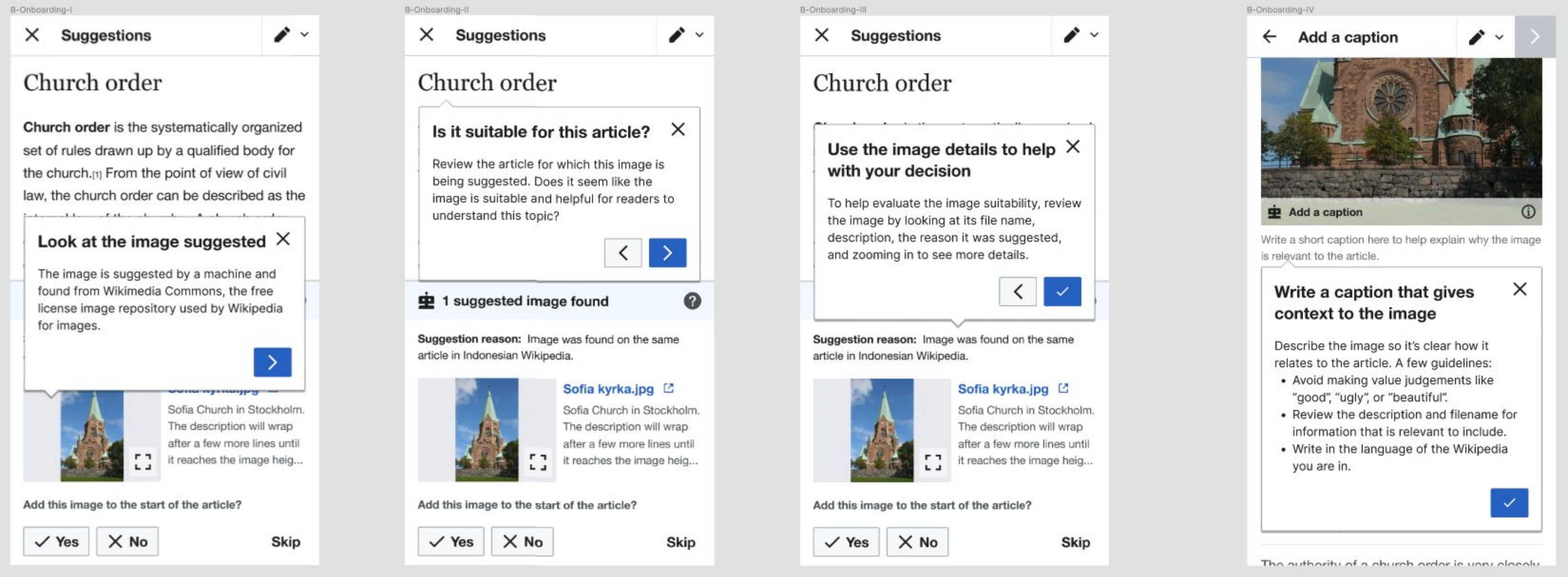
Wikipedia Pages Wanting Photos (WPWP) is an annual campaign where Wikipedia editors across the world, Wikipedia language projects and communities add photos to Wikipedia articles lacking photos. This is to promote the use of digital media files collected from various Wikimedia photography contests, photowalks organized by the Wikimedia community, on Wikipedia article pages. Photos help to grasp the reader's attention better than a wall of text, illustrate content, and make the article more instructive and engaging for readers.

Thousands of images have been donated and contributed to Wikimedia Commons via various advocacy programs, photowalks, and contests including international photography contests such as Wiki Loves Monuments, Wiki Loves Africa, Wiki Loves Earth, Wiki Loves Folklore, etc. Yet relatively few of these photos have been used on Wikipedia articles. Today, the Wikimedia Commons hosts millions of photo images but only a tiny portion of these have been used on Wikipedia article pages. This is a huge gap that this project aims at bridging.

Many Images!


Unillustrated Articles across Wikis
Vietnamese
Egyptian Arabic
Arabic
Swedish
Korean
English
French
0.0
0.2
0.4
0.6
Percentage of unillustrated articles

“Add an Image” Structured Task



B-Onboarding-I

Church order

Church order is the systematically organized set of rules drawn up by a qualified body for the church.[1] From the point of view of civil law, the church order can be described as the

Look at the image suggested ✕

The image is suggested by a machine and found from Wikimedia Commons, the free license image repository used by Wikipedia for images.

Sofia Church in Stockholm. The description will wrap after a few more lines until it reaches the image heig...

Add this image to the start of the article?

B-Onboarding-II

Church order

Is it suitable for this article? ✕

Review the article for which this image is being suggested. Does it seem like the image is suitable and helpful for readers to understand this topic?

1 suggested image found

Suggestion reason: Image was found on the same article in Indonesian Wikipedia.

Sofia kyrka.jpg

Sofia Church in Stockholm. The description will wrap after a few more lines until it reaches the image heig...

Add this image to the start of the article?

B-Onboarding-III

Church order

Use the image details to help with your decision ✕

To help evaluate the image suitability, review the image by looking at its file name, description, the reason it was suggested, and zooming in to see more details.

Suggestion reason: Image was found on the same article in Indonesian Wikipedia.

Sofia kyrka.jpg

Sofia Church in Stockholm. The description will wrap after a few more lines until it reaches the image heig...

Add this image to the start of the article?

B-Onboarding-IV

Add a caption

Write a short caption here to help explain why the image is relevant to the article.

Write a caption that gives context to the image ✕

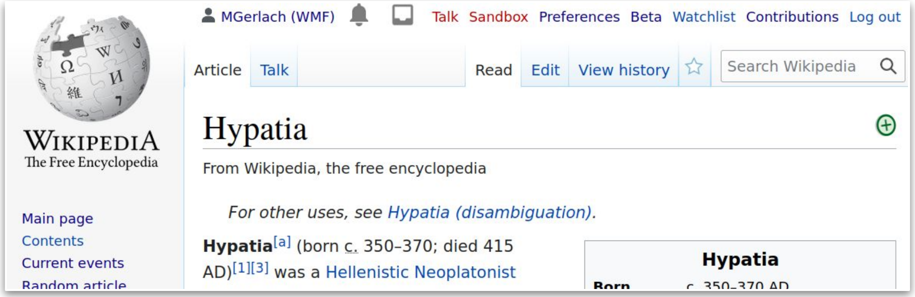
Describe the image so it’s clear how it relates to the article. A few guidelines:

- Avoid making value judgements like “good”, “ugly”, or “beautiful”.
- Review the description and filename for information that is relevant to include.
- Write in the language of the Wikipedia you are in.

The authority of a church order is very closely

By RHo (WMF) - Mockups from WMF designer, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=108208100

“Add an Image” Algorithm Logic



Hypatia

From Wikipedia, the free encyclopedia

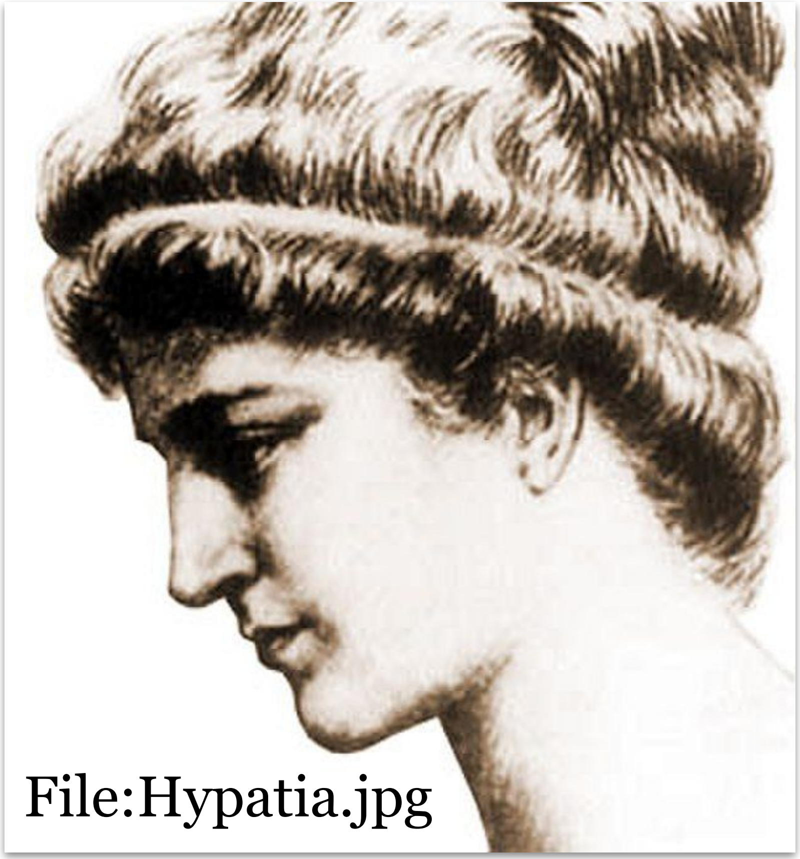
For other uses, see Hypatia (disambiguation).

Hypatia[a] (born c. 350–370; died 415 AD)[1][3] was a Hellenistic Neoplatonist

Hypatia	
Born	c. 350–370 AD

Unillustrated Article

Wikimedia Commons



File:Hypatia.jpg

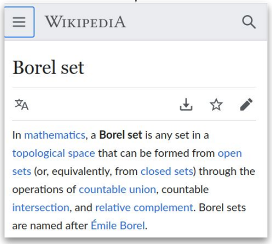
The “Add an Image” Task in Wikipedia

- Manually finding good images for articles is hard:
  - The space of search for good free-licensed image matches can be huge, especially for newcomers
  - Sometimes finding matches is difficult: image search is not very effective in absence of metadata
- Automatically finding good images for articles is also hard:
  - We need to have “explainable” image-article matches
  - Due to the impact and visibility of images, accuracy/quality of matches is crucial
  - Computer vision technologies are not able to identify fine-grained objects, and simpler models are preferred
- Language support
  - Similar to links 300+ language version of Wikipedia

A Simple Solution: **Leveraging Existing Matching Signals**

Discovering unillustrated articles as articles without images or with *icons* only.

Generate list of icons for each Wikipedia



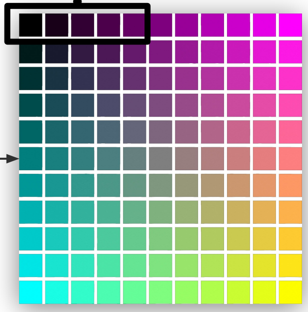
Borel set

In mathematics, a **Borel set** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.

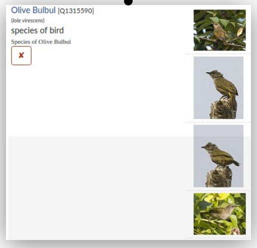
Unillustrated Article

Discovering related images from different sources

Filtering



Ranking images according to Relevance and Quality



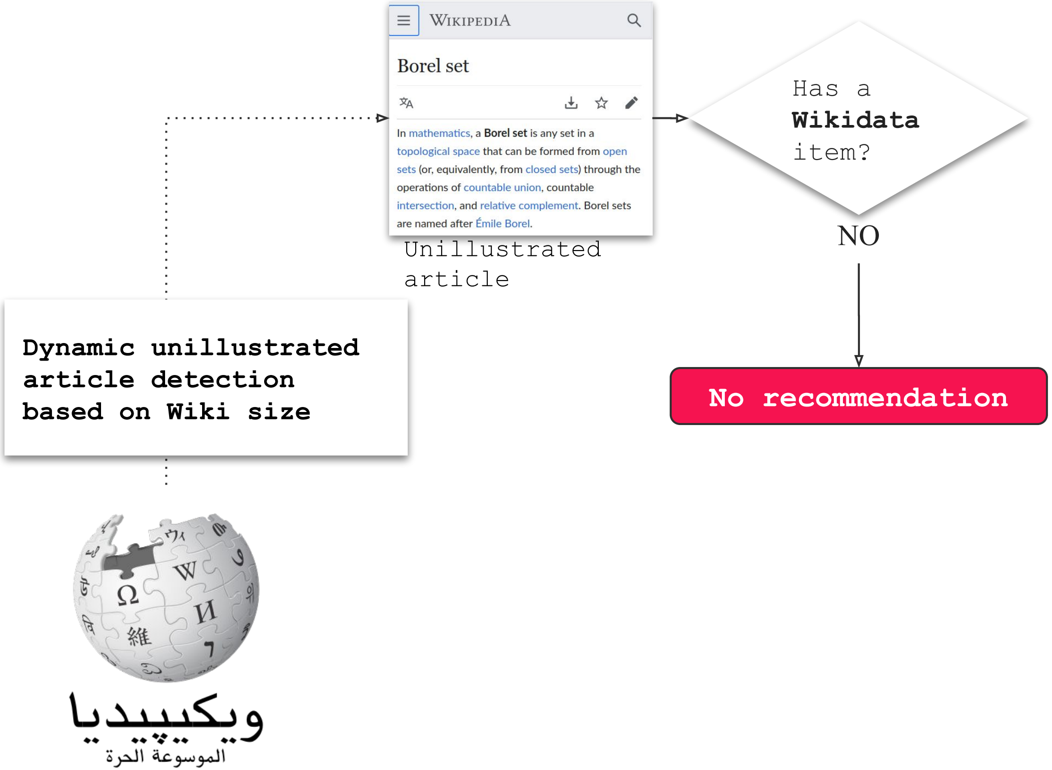
Olive Bulbul [Q1315590]

(*Iole virescens*)

species of bird

Species of Olive Bulbul

Task recommendation



WIKIPEDIA

Borel set

In mathematics, a **Borel set** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.

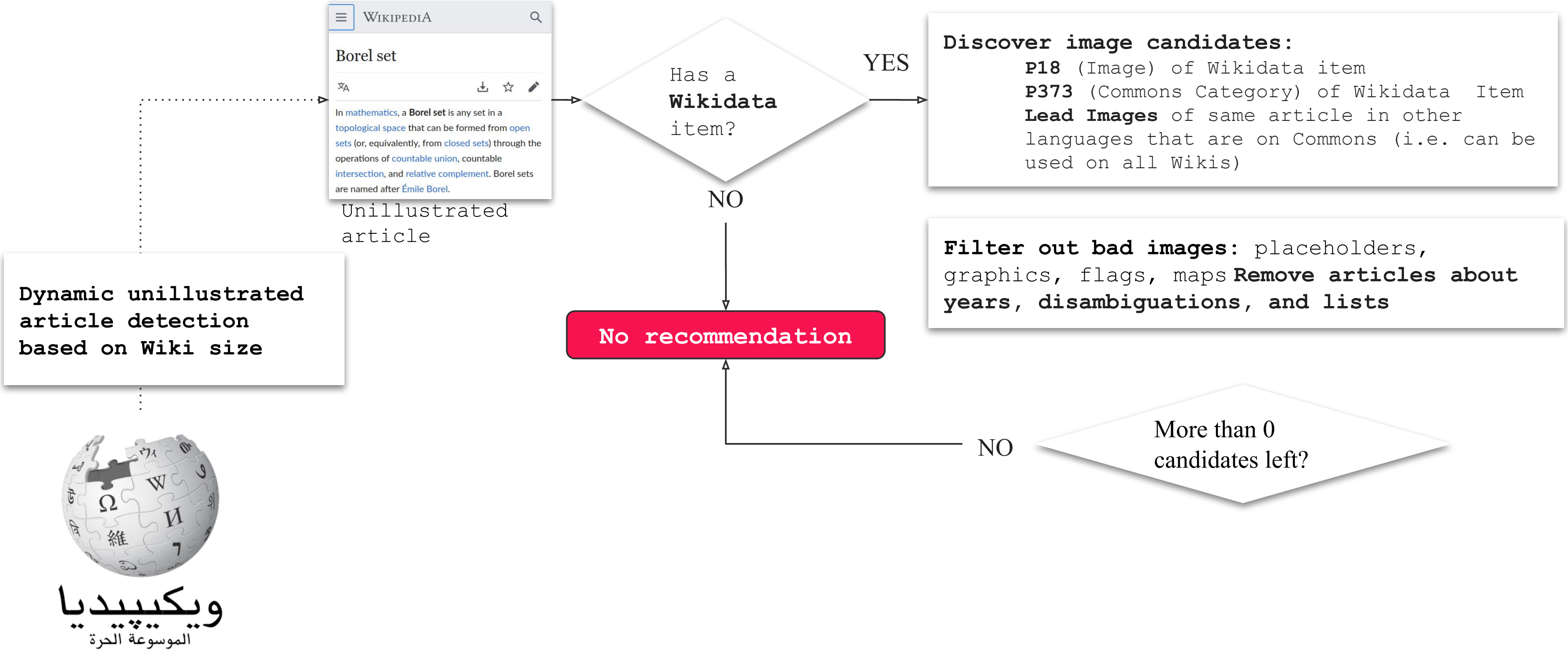
Unillustrated article

Has a **Wikidata** item?

NO

No recommendation

Dynamic unillustrated article detection based on Wiki size



Dynamic unillustrated article detection based on Wiki size

ويكيبيديا
الموسوعة الحرة

WIKIPEDIA

Borel set

In mathematics, a **Borel set** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.

Unillustrated article

Has a **Wikidata** item?

YES

NO

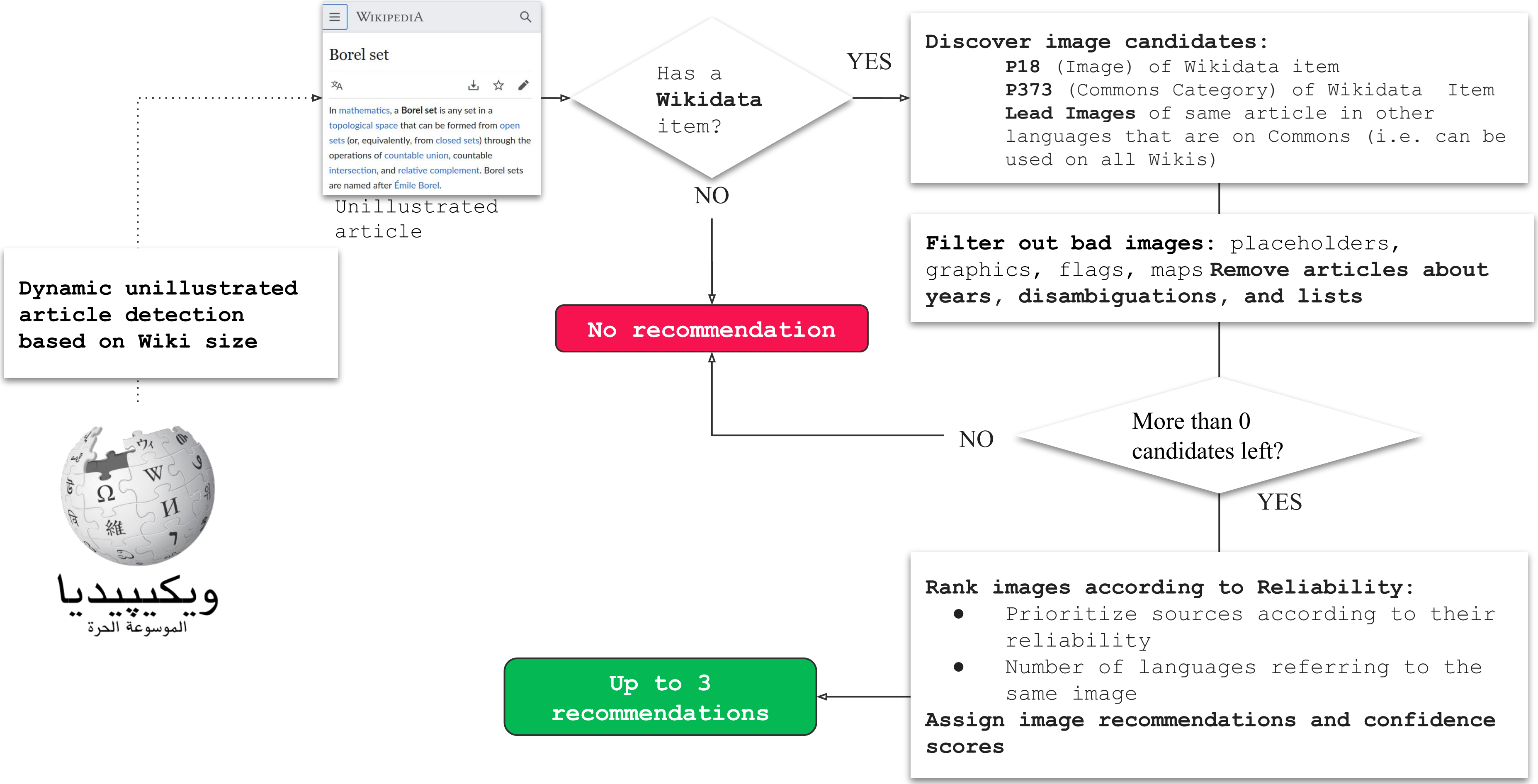
Discover image candidates:
P18 (Image) of Wikidata item
P373 (Commons Category) of Wikidata Item
Lead Images of same article in other languages that are on Commons (i.e. can be used on all Wikis)

Filter out bad images: placeholders, graphics, flags, maps **Remove articles about years, disambiguations, and lists**

No recommendation

NO

More than 0 candidates left?



Wikipedia

Borel set

In mathematics, a **Borel set** is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. Borel sets are named after Émile Borel.

Unillustrated article

Dynamic unillustrated article detection based on Wiki size

Has a **Wikidata** item?

Discover image candidates:

P18 (Image) of Wikidata item

P373 (Commons Category) of Wikidata Item

Lead Images of same article in other languages that are on Commons (i.e. can be used on all Wikis)

Filter out bad images: placeholders, graphics, flags, maps **Remove articles about years, disambiguations, and lists**

More than 0 candidates left?

No recommendation

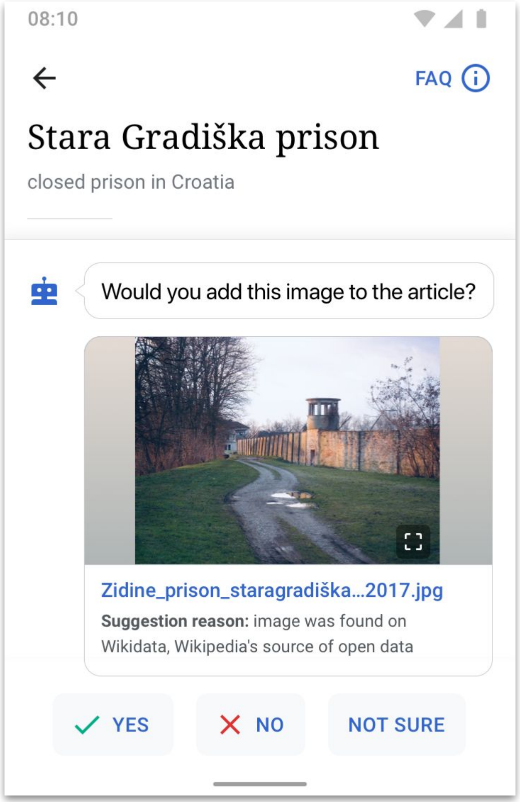
Rank images according to Reliability:

- Prioritize sources according to their reliability
- Number of languages referring to the same image

Assign image recommendations and confidence scores

Up to 3 recommendations

Evaluation



08:10

FAQ

Stara Gradiška prison

closed prison in Croatia

Would you add this image to the article?

Zidine_prison_staragradiška...2017.jpg

Suggestion reason: image was found on Wikidata, Wikipedia's source of open data

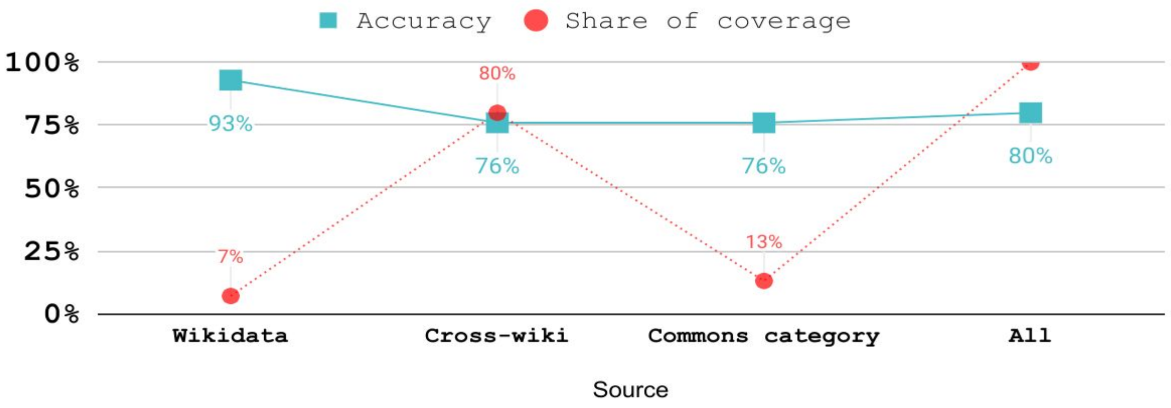
✓ YES ✗ NO NOT SURE

Android MVP “Train Image Algorithm”

Image Rec Task deployed on plwiki, eswiki, ruwiki, hewiki, ptwiki, fawiki, viwiki, frwiki, trwiki, itwiki, arwiki, enwiki, dewiki (many thanks to the Growth, Android and Platform Engineering Teams!!)

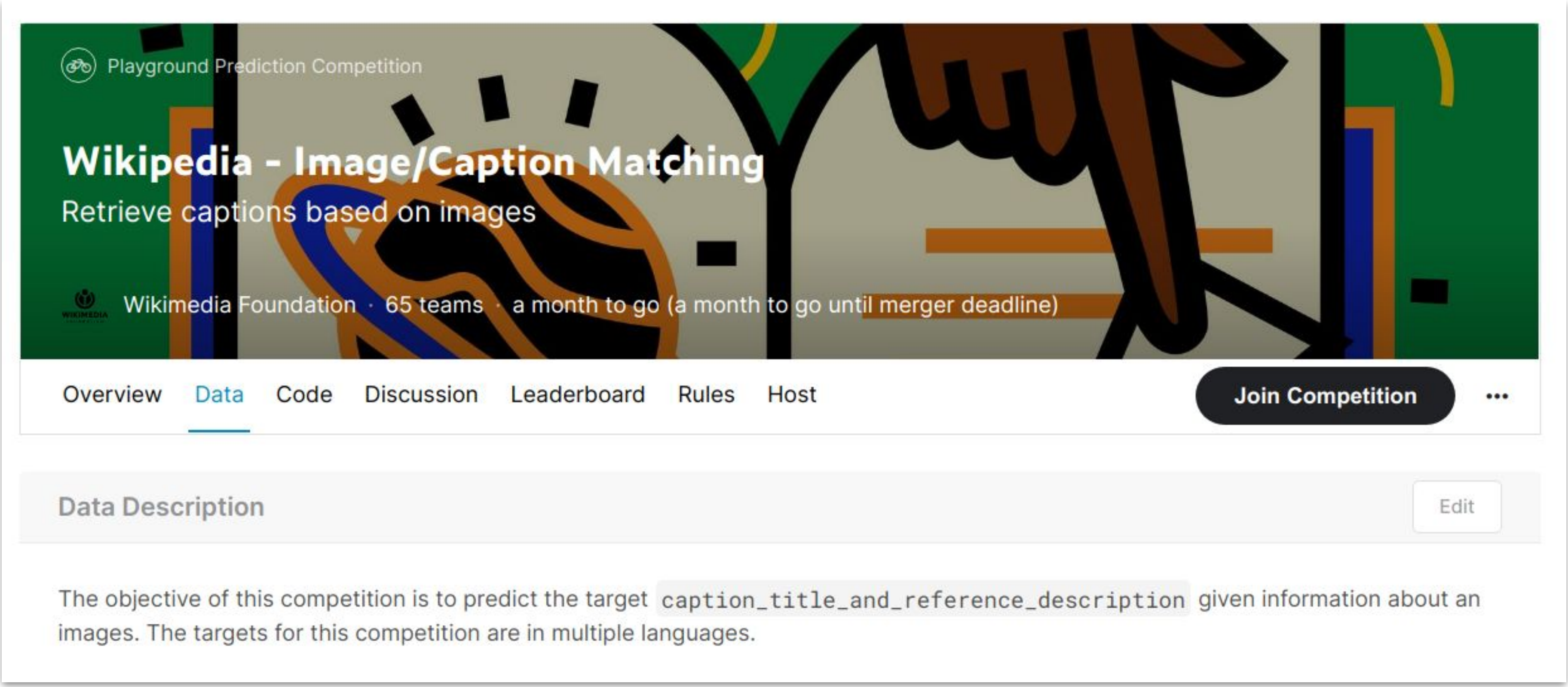
Engagement is high: average 9 edits/day

Precision is high but depends on source:



Coverage: 4% - 26% of unillustrated articles

Gathering our Scientific Communities around the problem of image/text matching



Playground Prediction Competition

Wikipedia - Image/Caption Matching

Retrieve captions based on images

Wikimedia Foundation · 65 teams · a month to go (a month to go until merger deadline)

Data Description

The objective of this competition is to predict the target `caption_title_and_reference_description` given information about an images. The targets for this competition are in multiple languages.

Equity in recommendations

Isaac

WIKIMEDIA
FOUNDATION

Equity in Recommendations

- Recommender systems facilitate a growing number of Wikimedia edits:
    - Suggested Edits (300K edits Commons; 825K edits Wikidata)
    - Newcomer Tasks (190K edits across 70 Wikipedias)
    - Content Translation (1M articles created across all Wikipedias)
- How do they align with knowledge equity goals?
    - What languages are these tools available in?
        - Read more: https://w.wiki/4Hhg
    - What types of content receives the greatest benefits? should be prioritized?
        - Read more: https://w.wiki/$4P

Newcomer Tasks -- Russian Wikipedia -- Topic Filters

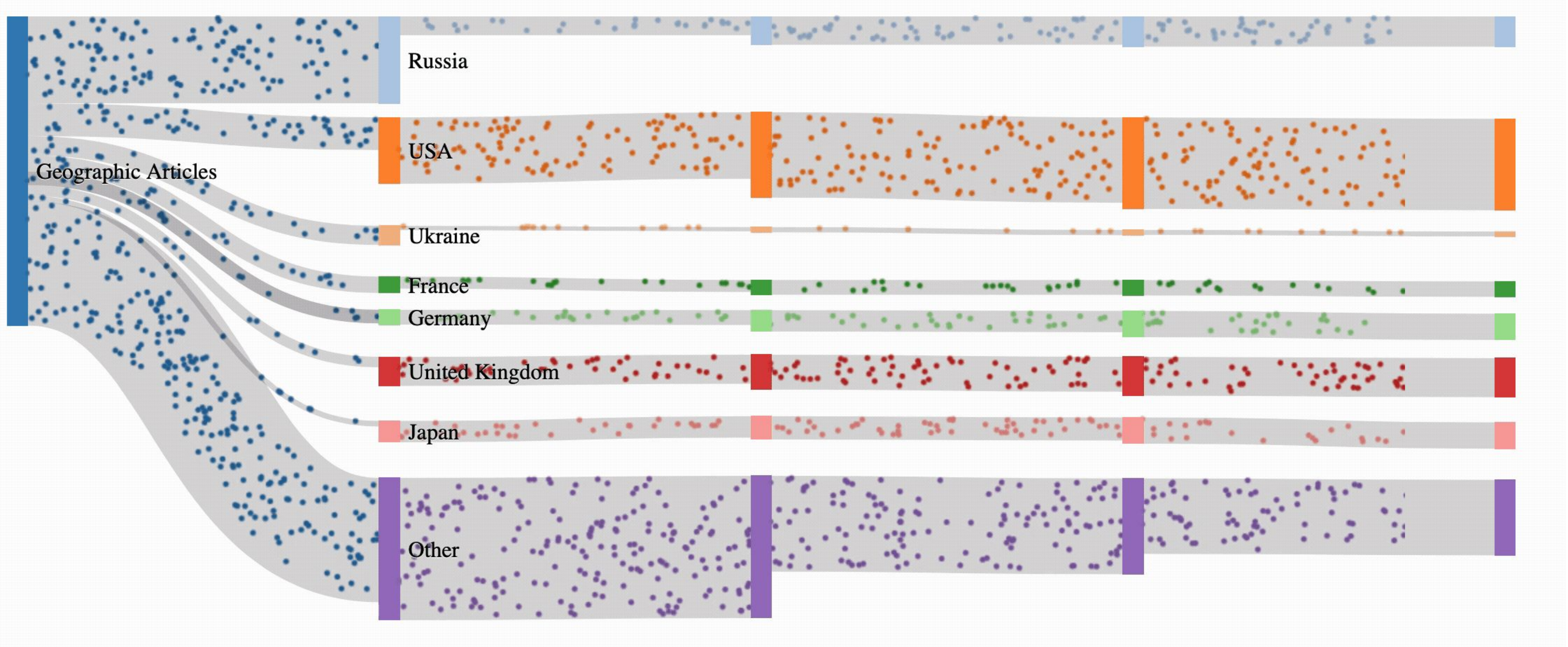
Flow of geographic "bias" induced by algorithmic filters (Baseline -> Impressions) and editor behavior (Impressions -> Clicks -> Edits)

Baseline
54% of Russian Wikipedia articles are geographic

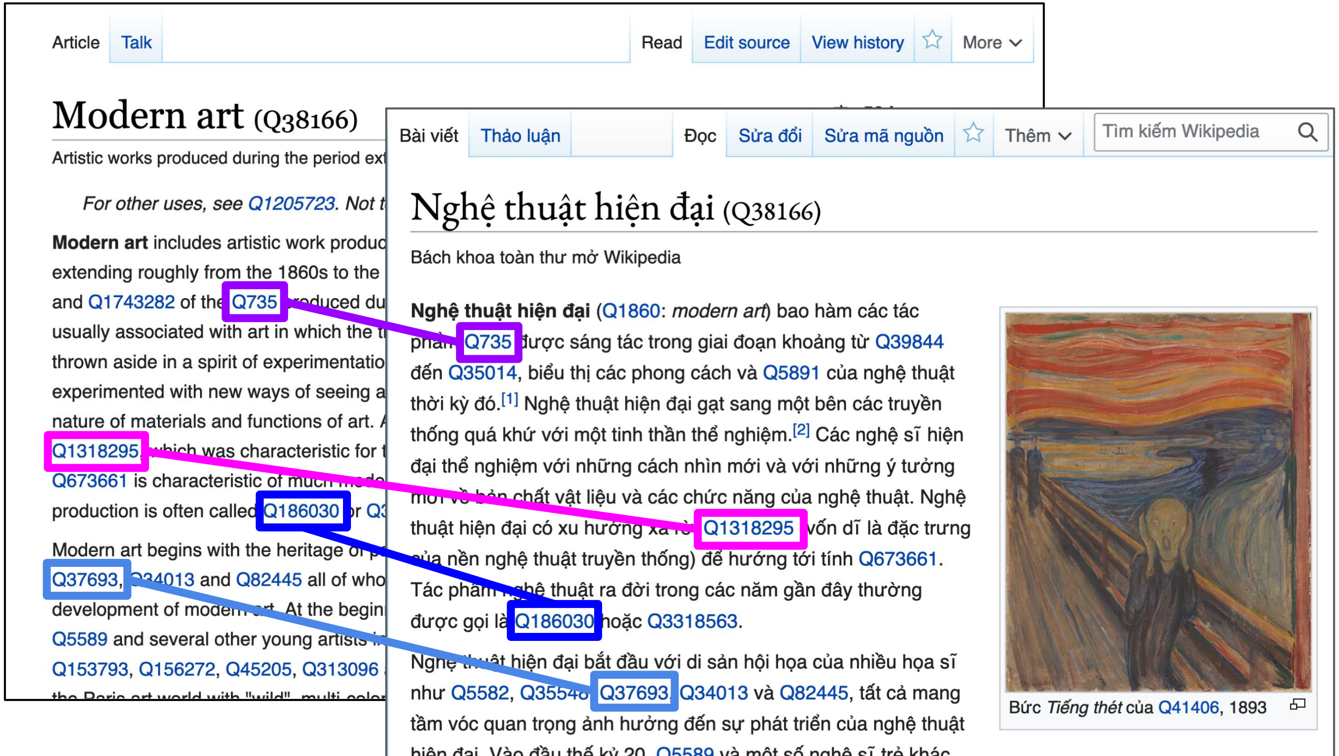
Impressions
Proportion of geographic impressions by country

Clicks
Proportion of geographic clicks by country
(5.7% of all impressions)

Edits
Proportion of geographic impressions by country
(56.5% of all clicks)



What challenges do you face in expanding section alignment to more Wikipedia languages?



Modern art (Q38166)

Artistic works produced during the period ext

For other uses, see Q1205723. Not t

Modern art includes artistic work produc
extending roughly from the 1860s to the
and Q1743282 of the Q735 produced du
usually associated with art in which the t
thrown aside in a spirit of experimentatio
experimented with new ways of seeing a
nature of materials and functions of art. A
Q1318295, which was characteristic for t
Q673661 is characteristic of much mode
production is often called Q186030 or Q3

Modern art begins with the heritage of p
Q37693, Q34013 and Q82445 all of who
development of modern art. At the begin
Q5589 and several other young artists i
Q153793, Q156272, Q45205, Q313096
the Paris art world with "wild", multi-color

Nghệ thuật hiện đại (Q38166)

Bách khoa toàn thư mở Wikipedia

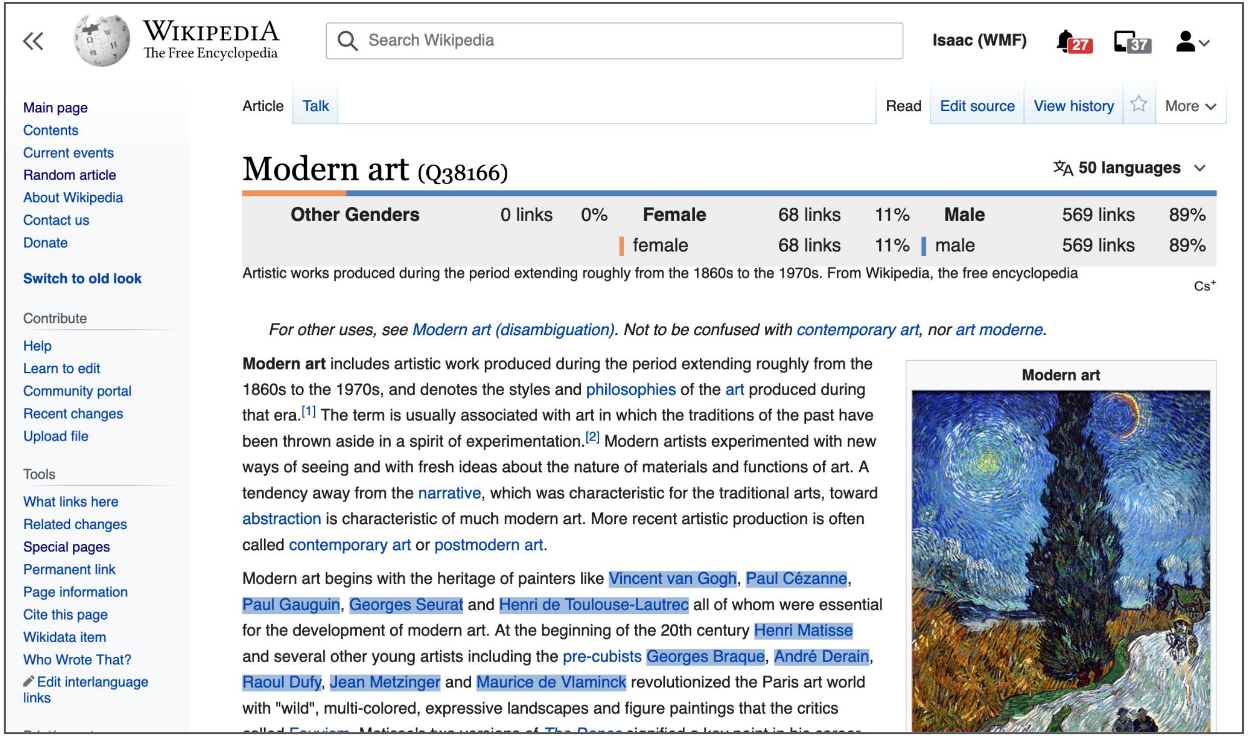
Nghệ thuật hiện đại (Q1860: *modern art*) bao hàm các tác phẩm Q735 được sáng tác trong giai đoạn khoảng từ Q39844 đến Q35014, biểu thị các phong cách và Q5891 của nghệ thuật thời kỳ đó.[1] Nghệ thuật hiện đại gạt sang một bên các truyền thống quá khứ với một tinh thần thể nghiệm.[2] Các nghệ sĩ hiện đại thể nghiệm với những cách nhìn mới và với những ý tưởng mới về bản chất vật liệu và các chức năng của nghệ thuật. Nghệ thuật hiện đại có xu hướng xa rời Q1318295 vốn dĩ là đặc trưng của nền nghệ thuật truyền thống) để hướng tới tính Q673661. Tác phẩm nghệ thuật ra đời trong các năm gần đây thường được gọi là Q186030 hoặc Q3318563.

Nghệ thuật hiện đại bắt đầu với di sản hội họa của nhiều họa sĩ như Q5582, Q35548, Q37693, Q34013 và Q82445, tất cả mang tầm vóc quan trọng ảnh hưởng đến sự phát triển của nghệ thuật hiện đại. Vào đầu thế kỷ 20, Q5589 và một số nghệ sĩ trẻ khác

Bức *Tiếng thét* của Q41406, 1893

Images of English and Vietnamese article for Modern Art showing how content can be aligned across Wikipedia articles through mapping wikilinks to their respective Wikidata IDs. More information: https://arxiv.org/abs/2103.00068

How might we prioritize links to be added to Wikipedia articles?



Modern art (Q38166)

Other Genders	0 links	0%	Female	68 links	11%	Male	569 links	89%
			female	68 links	11%	male	569 links	89%

Artistic works produced during the period extending roughly from the 1860s to the 1970s. From Wikipedia, the free encyclopedia

For other uses, see Modern art (disambiguation). Not to be confused with contemporary art, nor art moderne.

Modern art includes artistic work produced during the period extending roughly from the 1860s to the 1970s, and denotes the styles and philosophies of the art produced during that era.[1] The term is usually associated with art in which the traditions of the past have been thrown aside in a spirit of experimentation.[2] Modern artists experimented with new ways of seeing and with fresh ideas about the nature of materials and functions of art. A tendency away from the narrative, which was characteristic for the traditional arts, toward abstraction is characteristic of much modern art. More recent artistic production is often called contemporary art or postmodern art.

Modern art

Modern art begins with the heritage of painters like Vincent van Gogh, Paul Cézanne, Paul Gauguin, Georges Seurat and Henri de Toulouse-Lautrec all of whom were essential for the development of modern art. At the beginning of the 20th century Henri Matisse and several other young artists including the pre-cubists Georges Braque, André Derain, Raoul Dufy, Jean Metzinger and Maurice de Vlaminck revolutionized the Paris art world with "wild", multi-colored, expressive landscapes and figure paintings that the critics called Fauvism. Matisse's two versions of The Dance signified a key point in his career

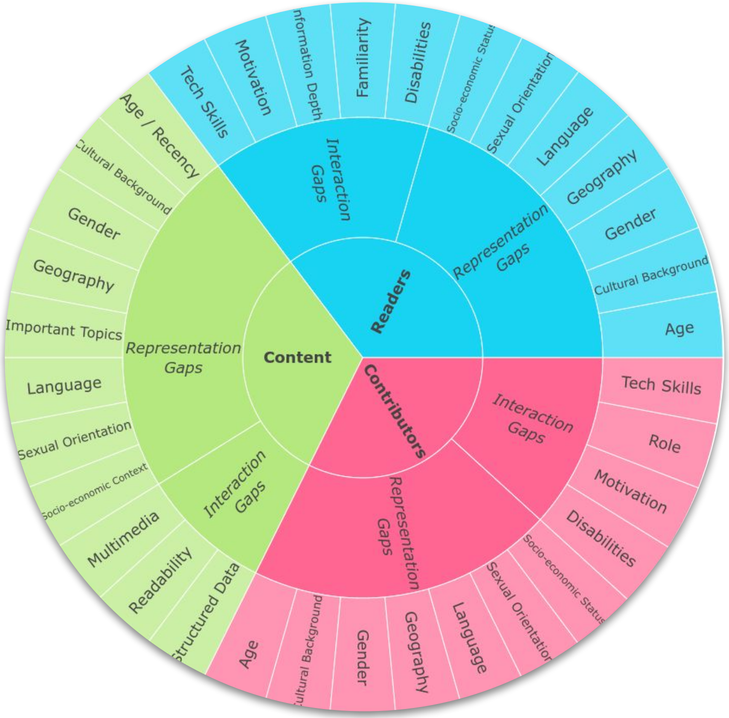
Snapshot of English Wikipedia Modern Art article with links highlighted based on gender of the person they are about. Links to articles about men are often more prominent than links to articles about women or non-binary individuals. More information: Youtube:WikiConference North America 2021

What special considerations do we make when considering the usage of recommendation or AI models for images?

Logo for VisibleWikiWomen, an annual campaign organized by Whose Knowledge? that focuses on adding images to articles about women.
CC-BY-SA 4.0. Whose Knowledge?

Next steps ...

We need more and more research to Address Knowledge Gaps!


Content
Representation Gaps
Age / Recency
Cultural Background
Gender
Geography
Important Topics
Language
Sexual Orientation
Socio-economic Context
Interaction Gaps
Multimedia
Readability
Structured Data
Readers
Interaction Gaps
Tech Skills
Motivation
Information Depth
Familiarity
Disabilities
Representation Gaps
Socio-economic Status
Sexual Orientation
Language
Geography
Gender
Cultural Background
Age
Contributors
Interaction Gaps
Tech Skills
Role
Motivation
Disabilities
Representation Gaps
Socio-economic Status
Sexual Orientation
Language
Geography
Gender
Cultural Background
Age