



- Main page
- Contents
- Current events
- Random article
- About Wikipedia
- Contact us
- Donate

- Contribute
- Help
- Learn to edit
- Community portal
- Recent changes
- Upload file

- Tools
- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Cite this page
- Wikidata item

- Print/export
- Download as PDF
- Printable version

- In other projects
- Wikimedia Commons
- Meta-Wiki
- Wikisource
- Wikiquote

Pitch

From Wikipedia, the free encyclopedia

This article is about Wikipedia. For Wikipedia's home page, see Main Page. For the English edition, see English Wikipedia. For other uses, see Wikipedia (disambiguation).

Wikipedia (/wɪkɪˈpiːdiə/ (listen) *wik-ih-PEE-dee-ə* or /wɪki-/ (listen) *wik-ee-*) is a free, multilingual open-collaborative online encyclopedia created and maintained by a community of volunteer contributors using a wiki-based editing system. Wikipedia is the largest general reference work on the Internet,^[3] and one of the 15 most popular websites as ranked by Alexa; in 2021, it was ranked as the 13th most visited.^{[4][note 3]} The project carries no advertisements and is hosted by the Wikimedia Foundation, an American non-profit organization funded mainly through donations, 80% of which are small donations from individual users.^[6]

Wikipedia was launched on January 15, 2001, by Jimmy Wales and Larry Sanger; its name was coined as a portmanteau of "wiki" and "encyclopedia". Initially available only in English, versions in other languages were quickly developed. The English Wikipedia, with 6.3 million articles as of March 2021, is the largest of the 319 language editions. Combined, Wikipedia's editions comprise more than 55 million articles, and attract more than 17 million edits and more than 1.7 billion unique visitors per month.^[7]

Wikipedia has been criticized for its open-door policy of allowing anyone to edit, its gender gap, and its editors being male.^[8] In 2006, *Time* magazine stated that the open-door policy of allowing anyone to edit had made Wikipedia the "biggest and perhaps the best encyclopedia in the world", and a testament to the vision of Jimmy Wales.^[9] The project's reputation improved further in the 2010s, as it received praise for its unique structure, culture, and absence of commercial bias.^{[3][8]} In 2018, Facebook and YouTube announced that they would help users detect fake news by suggesting links to related Wikipedia articles.^[10]

Contents [hide]
<div> <div><div>1</div><div>History</div></div> <div> <div></div> <div><div>1.1</div><div>Nupedia</div></div> <div><div>1.2</div><div>Launch and early growth</div></div> <div><div>1.3</div><div>Milestones</div></div> </div> </div>
<div> <div><div>2</div><div>Openness</div></div> <div> <div></div> <div><div>2.1</div><div>Restrictions</div></div> <div><div>2.2</div><div>Review of changes</div></div> <div><div>2.3</div><div>Vandalism</div></div> <div><div>2.4</div><div>Edit warring</div></div> </div> </div>
<div> <div><div>3</div><div>Policies and laws</div></div> <div> <div></div> <div><div>3.1</div><div>Content policies and guidelines</div></div> </div> </div>

Colin Salama

NAME

March 1st, 2021

DATE

Wikipedia



der

WIKIPEDIA

The logo of Wikipedia, a globe featuring glyphs from various writing systems

Screenshot

[show]

Type of site	Online encyclopedia
Available in	319 languages
Country of origin	United States
Owner	Wikimedia Foundation
Created by	Jimmy Wales Larry Sanger ^[1]
URL	wikipedia.org 
Commercial	No
Registration	Optional ^[note 1]
Users	<div>>353,150 active users^[note 2] and >94,591,047 registered users</div> <div>1,109 administrators (English)</div>
Launched	January 15, 2001; 20 years ago

5. Next Steps

Goal

The main goal of this project is to **build a simple recommendation system for Wikipedia articles** using NLP.

Recommendation systems are important for maximizing interaction with website and generally would improve the Wikipedia experience.

Frequently Bought Together



Price for all three: **\$74.20**

[Add all three to Cart](#) [Add all three to Wish List](#)


Show availability and shipping details

☒ **This item:** Beginning Ruby: From Novice to Professional (Expert's Voice in Open Source) by Peter Cooper Paperback **\$27.78**


☒ Learn to Program, Second Edition (The Facets of Ruby Series) by Chris Pine Paperback **\$16.94**

☒ Ruby on Rails Tutorial: Learn Web Development with Rails (2nd Edition) (Addison-Wesley Professional Ruby ... by Michael Hartl Paperback **\$29.48**


Customers Who Bought This Item Also Bought




Learn to Program, Second Edition (The Facets of...)
Chris Pine
★★★★★ 42
Paperback
\$16.94 ✓Prime




The Well-Grounded Rubyist
David A. Black
★★★★★ 39
Paperback
\$32.49 ✓Prime



Ruby on Rails Tutorial: Learn Web Development...
Michael Hartl
★★★★★ 70
Paperback
\$29.48 ✓Prime



The Ruby Programming Language
David Flanagan
★★★★★ 74
Paperback
\$26.35 ✓Prime



The Well-Grounded Rubyist
David A. Black
★★★★★ 19
#1 Best Seller in Ruby Programming Computer
Paperback
\$29.67 ✓Prime

YouTube

Search

Home

Trending

Subscriptions

LIBRARY

History

Watch Later

on.click - Click C...

Benchmark Sear...

Show more

SUBSCRIPTIONS

Adjustamatic

Browse channels

YouTube Movies

Settings

Help


Send feedback

About Press Copyright

Creators Advertise

Developers +YouTube


Recommended



The Ricky Gervais Show Season 2 Episode 09 Natural History

Tianshu Li • 384K views • 2 years ago

Recommended




The Martian - David Bowie - Starman

Acting is an Art • 724K views • 2 years ago

The Martian Movie Tribute

Recommended




Ed Sheeran - Perfect Duet with Beyonce (Live)

Ed • 7.1M views • 1 month ago

Ed Sheeran perfect official audio with beyoncé

Recommended



Karl's Complete Diary.


Marc Swyane • 4.2M views • 4 years ago

Questions for Karl 1: https://www.youtube.com/watch?v=hAaB2oeyA_4 Questions for Karl 2:

RECOMMENDED VIDEOS


1.

How Public Universities Became So Expensive



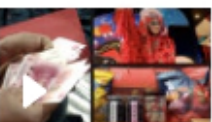
2.

Facebook Builds Out in NYC, Betting on In-Person Future




3.

China's New Tactic to Prevent the Spread of Covid-19? Cash and Gifts




4.

SpaceX's Starship Test Flight Ends in Explosion, Again



5.

Dozens of Boeing 777 Jets Grounded After Engine Failure



Pitch



WIKIPEDIA
 The Free Encyclopedia

Main page
 Contents
 Current events
 Random article
 About Wikipedia
 Contact us
 Donate

Contribute

Help
 Learn to edit
 Community portal
 Recent changes
 Upload file

Tools

What links here
 Related changes
 Special pages
 Permanent link
 Page information
 Cite this page
 Wikidata item

Print/export

Download as PDF
 Printable version

In other projects

Wikimedia Commons
 Meta-Wiki
 Wikibooks
 Wikiquote
 Wikiversity

Languages

Article
 Talk

Read

View source

View history

Search Wikipedia



Wikipedia

From Wikipedia, the free encyclopedia

This article is about Wikipedia. For Wikipedia's home page, see [Main Page](#). For the English edition, see [English Wikipedia](#). For other uses, see [Wikipedia \(disambiguation\)](#).

Wikipedia (/ˈwɪkɪˈpiːdiə/ (listen) *wik-ih-PEE-dee-ə* or /ˈwɪki-/ (listen) *wik-ee-*) is a free, multilingual open-collaborative online encyclopedia created and maintained by a community of volunteer contributors using a wiki-based editing system. Wikipedia is the largest general reference work on the Internet,^[3] and one of the 15 most popular websites as ranked by Alexa; in 2021, it was ranked as the 13th most visited.^{[4][note 3]} The project carries no advertisements and is hosted by the Wikimedia Foundation, an American non-profit organization funded mainly through donations, 80% of which are small donations from individual users.^[6]

Wikipedia was launched on January 15, 2001, by Jimmy Wales and Larry Sanger; its name was coined as a portmanteau of "wiki" and "encyclopedia". Initially available only in English, versions in other languages were quickly developed. The English Wikipedia, with 6.3 million articles as of March 2021, is the largest of the 319 language editions. Combined, Wikipedia's editions comprise more than 55 million articles, and attract more than 17 million edits and more than 1.7 billion unique visitors per month.^[7]

Wikipedia has been criticized for its uneven accuracy and for exhibiting systemic bias, particularly gender bias, with the majority of editors being male.^[8] In 2006, *Time* magazine stated that the open-door policy of allowing anyone to edit had made Wikipedia the "biggest and perhaps the best encyclopedia in the world", and a testament to the vision of Jimmy Wales.^[9] The project's reputation improved further in the 2010s, as it received praise for its unique structure, culture, and absence of commercial bias.^{[3][8]} In 2018, Facebook and YouTube announced that they would help users detect fake news by suggesting links to related Wikipedia articles.^[10]

Contents [hide]

- History
 - Nupedia
 - Launch and early growth
 - Milestones
- Openness
 - Restrictions
 - Review of changes
 - Vandalism
 - Edit warring
- Policies and laws
 - Content policies and guidelines
- Governance
 - Administrators

Wikipedia



WIKIPEDIA

The logo of Wikipedia, a globe featuring glyphs from various writing systems

Screenshot

[show]

Type of site	Online encyclopedia
Available in	319 languages
Country of origin	United States
Owner	Wikimedia Foundation
Created by	Jimmy Wales <div>Larry Sanger^[1]</div>
URL	wikipedia.org
Commercial	No
Registration	Optional ^[note 1]
Users	>353,150 active users ^[note 2] and >94,591,047 registered users <div>1,109 administrators (English)</div>
Launched	January 15, 2001; 20 years ago
Current status	Active
Content	CC Attribution / Share-Alike 3.0

Data



WIKIPEDIA
The Free Encyclopedia

WIKIPEDIA DATA

Only the English Wikipedia is considered for this project, including over 7 million articles. The text alone of these articles is ~15 GB compressed and ~58 GB uncompressed. It is all available for download on Wikipedia.

CLEANING THE DATA

An XML parser, Gensim's Wikipedia filtering function, and Gensim's preprocessor are used to clean the data into a usable format.

Additionally, only the introduction of Wikipedia articles is used in order to reduce the amount of data.

MULTIPROCESSING

Multiprocessing is used on partitioned Wiki files to reduce the estimated time for cleaning and tokenization from ~20 hours to ~3 hours. Then the tokens and article titles are aggregated into individual lists.



Method

- **Gensim's Corpus/Dictionary** format was used to examine this data, and the Corpus was streamed directly from memory due to prevent RAM overload.
- This Corpus was converted to a **TF-IDF Corpus** with basic term frequency, inverse document collection frequency, and cosine length normalization.
- Then the Gensim **docsim Similarity Matrix** is built from the TF-IDF Corpus and queried in order to find the N most similar articles by cosine similarity.

Wikipedia Recommender

Enter a Wikipedia article:

Brush Mountain East Wilderness

Your recommended articles are as follows:

-[Garden Mountain Wilderness](#)

-[Beartown Wilderness](#)

-[Rich Hole Wilderness](#)

-[Ramsey's Draft Wilderness](#)

-[Category:Wilderness Areas of New Hampshire](#)

Streamlit App

We **query our similarity matrix** through our document title to ID dictionary, and this returns the top 5 similar documents excluding the article itself.

Future Work

OTHER RECOMMENDATION SYSTEMS

Cosine distance and TF-IDF vectors were used to simplify the recommender due to the 7+ million documents. The next goal here is to implement LDA and Word2Vec recommenders as well.

SPEED UP RECOMMENDATION

Storing the top several recommendation systems into memory rather than querying the Similarity matrix would speed up the process.