Open access policy

The Wikimedia Foundation’s mission is to disseminate open knowledge effectively and globally. In keeping with this mission, the Wikimedia Foundation supports research in areas that benefit the Wikimedia community. We aim to make any work produced with our support openly available to the public and reusable on Wikimedia projects.
Minimalist user data collection

Revision histories are public data
No 3rd-party collection or sharing of user data
Private user data removed after 90 days
What can I do?

- NLP
  - Topic Models
  - Word Embeddings
  - Quality
- Graphs
  - Page links
  - Users graph
- Geolocation
- Socio-technical
  - User Influence
  - Cultural Differences
  - Online Deliberation
  - Articles Popularity
- Multimedia
  - Audio, Video, Image Analysis
- And more...
Research Resources

- Statistics
- Dumps
- MediaWiki Utilities
- Wikimedia API
- Page Views
- SQL Replicas / Quarry
- Clicks
- Event Stream
- Wikidata
- Commons
- ORES
- PAWS
New York University

From Wikipedia, the free encyclopedia

Coordinates: 40°43′48″N 73°59′42″W

This article is about the private institution founded in 1831. For other and similar uses, see University of New York.

"Nyu" redirects here. For the district in Japan, see Nyū District, Fukuī. For the Elfen Lied character, see List of Elfen Lied characters § Nyu.

New York University (NYU) is a private nonprofit research university based in New York City. Founded in 1831, NYU's primary campus is in Greenwich Village with other campuses throughout New York City. NYU students can also study abroad at its degree-granting campuses in NYU Abu Dhabi and NYU Shanghai, as well as its 11 academic centers in Accra, Berlin, Buenos Aires, Florence, London, Madrid, Paris, Prague, Sydney, Tel Aviv, and Washington, D.C.

In 2018, NYU was ranked amongst the top 30 universities internationally by the Academic Ranking of World Universities, Times Higher Education World University Rankings, and U.S. News & World Report. For the
<table>
<thead>
<tr>
<th>Date</th>
<th>User</th>
<th>Changes (Talk</th>
<th>Title</th>
<th>Files</th>
<th>Size</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>22:29, 12 November 2018</td>
<td>Tom Stevens (talk</td>
<td>+428</td>
<td>Athletics</td>
<td>Tag: Visual edit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04:58, 10 November 2018</td>
<td>GreenC bot (talk</td>
<td>-111</td>
<td>Wayback Med 2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07:25, 8 November 2018</td>
<td>Kiwi128 (talk</td>
<td></td>
<td>Source official website URL from Wikidata</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>04:51, 4 November 2018</td>
<td>Elkevbo (talk</td>
<td>+1,085</td>
<td>(no, we don't need all of this information in the lead; discuss in Talk)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Talk: New York University

This is the talk page for discussing improvements to the New York University article. This is not a forum for general discussion of the article’s subject.

Semi-protected edit request on 22 June 2018

Acid after first paragraph of article:

"With an undergraduate acceptance rate of 19 percent for the Class of 2022, NYU is becoming one of the most exclusive institutions in the United States."

Ssv289 (talk) 16:18, 22 June 2018 (UTC)

Not done: please provide reliable sources that support the change you want to be made.--QueerFilmNerd (talk) 17:38, 22 June 2018 (UTC)
Public Content

- XML dumps
- Mediawiki API
- SQL Replicas

https://dumps.wikimedia.org/

[[Artificial intelligence]] is a recurrent theme in [[science fiction]], whether [[utopia]]n, emphasising the potential benefits, or [[dystopian]], emphasising the dangers.

The notion of machines with human-like intelligence dates back at least to [[Samuel Butler (novelist)|Samuel Butler]]'s 1872 novel '[[Erewhon]]'.

==Background==
[[File:Didrachm Phaistos obverse CdM.jpg|thumb|A [[didrachm]] coin depicting the winged [[Talos]], an [[automaton]] or [[artificial being]] in [[ancient Greece|ancient Greek]] myth, c. 300 BC]]
Quarry / Sql Replicas

Quarry is a public querying interface for Wiki Replicas, a set of live replica SQL databases of public Wikimedia Wikis.

- [https://quarry.wmflabs.org](https://quarry.wmflabs.org)
- [https://upload.wikimedia.org/wikipedia/commons/9/94/MediaWiki_1.28.0_database_schema.svg](https://upload.wikimedia.org/wikipedia/commons/9/94/MediaWiki_1.28.0_database_schema.svg)
It's collaborative and searchable

SQL

```sql
use itwiki_p;
SELECT page_title, el_to
FROM externallinks
JOIN page ON page_id=el_from AND page_namespace=0
WHERE el_to LIKE "http://www.ted.com/talks/%" OR el_to LIKE "https://www.ted.com/talks/%" OR el_to LIKE "http://ted.com/talks/%" OR el_to LIKE "https://ted.com/talks/%"
```

Resultset (112 rows)

<table>
<thead>
<tr>
<th>page_title</th>
<th>el_to</th>
</tr>
</thead>
<tbody>
<tr>
<td>William_Kamkwamba</td>
<td><a href="http://www.ted.com/talks/william_kamkwamba_on_building_a_windmill.html">http://www.ted.com/talks/william_kamkwamba_on_building_a_windmill.html</a></td>
</tr>
<tr>
<td>WikiHouse</td>
<td><a href="http://www.ted.com/talks/alastair_parvin_architecture_for_the_people_by_the_people.html">http://www.ted.com/talks/alastair_parvin_architecture_for_the_people_by_the_people.html</a></td>
</tr>
<tr>
<td>Vita_personale_di_Leonardo_da_Vinci</td>
<td><a href="http://www.ted.com/talks/siegfried_woldhek_shows_how_he_found_the_true_face_of_leonardo.html">http://www.ted.com/talks/siegfried_woldhek_shows_how_he_found_the_true_face_of_leonardo.html</a></td>
</tr>
<tr>
<td>Vilayanur_S._Ramachandran</td>
<td><a href="http://www.ted.com/talks/lang/ita/vilayanur_ramachandran_on_your_mind.html">http://www.ted.com/talks/lang/ita/vilayanur_ramachandran_on_your_mind.html</a></td>
</tr>
<tr>
<td>Vilayanur_S._Ramachandran</td>
<td><a href="http://www.ted.com/talks/lang/ita/vs_ramachandran_the_neurons_that_shaped_civilization.html">http://www.ted.com/talks/lang/ita/vs_ramachandran_the_neurons_that_shaped_civilization.html</a></td>
</tr>
</tbody>
</table>
Wikidata

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wikisource, and others.

In October 2018, there 23 M of edits in Wikidata, compared with around 4 M edits in English Wikipedia.

- [https://www.wikidata.org/](https://www.wikidata.org/)
- [https://query.wikidata.org/](https://query.wikidata.org/)
New York University

From Wikipedia, the free encyclopedia

This article is about the private institution founded in 1831. For other and similar uses, see University of New York.

"Nyu" redirects here. For the district in Japan, see Nyū District, Fukuoka. For the Elfen Lied character, see List of Elfen Lied characters § Nyu.

New York University (NYU) is a private nonprofit research university based in New York City. Founded in 1831, NYU's primary campus is in Greenwich Village with other campuses throughout New York City.[12][13] NYU students can also study abroad at its degree-granting campuses in NYU Abu Dhabi and NYU Shanghai, as well as its 11 academic centers in Accra, Berlin, Buenos Aires, Florence, London, Madrid, Paris, Prague, Sydney, Tel Aviv, and Washington, D.C.[14][15]

In 2018, NYU was ranked amongst the top 30 universities Internationally by the Academic Ranking of World Universities, Times Higher Education World University Rankings, and U.S. News & World Report. For the class that matriculated in the fall of 2018, NYU received 75,037 applications for its undergraduate programs; this is more applications than any other private college or university in the United States.[19]

Alumni include heads of state, royalty, eminent scientists, inventors and entrepreneurs, media figures, founders and CEOs of Fortune 500 companies, and astronauts.[20][21][22] As of 2018, 37 Nobel Laureates, 7 Turing Award winners, 5 Fields Medalists, over 30 Academy Award winners, over 30 Pulitzer Prize winners, and hundreds of members of the National Academies of Sciences and United States Congress have been affiliated as faculty or alumni. Globally, NYU is ranked 7th by the Times Higher Education World University Rankings for producing alumni who are millionaires, and 4th by...
<table>
<thead>
<tr>
<th>founded by</th>
<th>Albert Gallatin</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>United States of America</td>
</tr>
<tr>
<td>located in the administrative territorial entity</td>
<td>Manhattan</td>
</tr>
<tr>
<td>coordinate location</td>
<td>40°43'48&quot;N, 73°59'42&quot;W</td>
</tr>
</tbody>
</table>
#Popular eye colors among humans

added before 2016-10

illustrates bubblechart view, count

defaultView: BubbleChart

```
SELECT ?eyeColorLabel (COUNT(?human) AS ?count)
WHERE
{
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]" }
}
GROUP BY ?eyeColorLabel
```
Wikimedia Commons (or simply Commons) is an online repository of free-use images, sound, and other media files.

As of 2018, the repository contains over 44 million free media files, managed and editable by registered volunteers. In July 2013, the number of edits on Commons reached 100,000,000.

https://commons.wikimedia.org/wiki/Main_Page
Wikimedia Statistics

Live dashboards about:
- Editing
- Reading
- Users

https://stats.wikimedia.org/v2
Page Views

This page documents the Pageview API (v1), a public API developed and maintained by the Wikimedia Foundation that serves analytical data about article pageviews of Wikipedia and its sister projects. With it, you can get pageview trends on specific articles or projects; filter by agent type or access method, among other things.

- https://wikitech.wikimedia.org/wiki/Analytics/AQS/Pageviews
Pageviews Analysis

Comparison of pageviews across multiple pages

Pages: Enter up to 10 pages

- Random forest
- Gradient boosting
- Deep learning

Chart type
Permalink
Download

Show values
Begin at zero
Logarithmic scale

Data chart showing pageviews analysis over time.
Pageviews Analysis
Comparison of pageviews across multiple pages

Pages: Enter up to 10 pages

Chart type
Permalink
Download

Show values
Begin at zero
Logarithmic scale
# Topviews Analysis

Most viewed pages of a project

<table>
<thead>
<tr>
<th>Rank</th>
<th>Page</th>
<th>Edits</th>
<th>Editors</th>
<th>Pageviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2018 FIFA World Cup</td>
<td>533</td>
<td>176</td>
<td>8,614,841</td>
</tr>
<tr>
<td>2</td>
<td>Exo (band)</td>
<td>70</td>
<td>14</td>
<td>5,786,931</td>
</tr>
<tr>
<td>3</td>
<td>Croatia</td>
<td>74</td>
<td>28</td>
<td>4,961,832</td>
</tr>
<tr>
<td>4</td>
<td>Kylian Mbappé</td>
<td>224</td>
<td>53</td>
<td>3,621,830</td>
</tr>
<tr>
<td>5</td>
<td>FIFA World Cup</td>
<td>178</td>
<td>74</td>
<td>3,591,619</td>
</tr>
<tr>
<td>6</td>
<td>List of FIFA World Cup finals</td>
<td>124</td>
<td>66</td>
<td>3,522,853</td>
</tr>
<tr>
<td>7</td>
<td>Cristiano Ronaldo</td>
<td>345</td>
<td>98</td>
<td>3,436,411</td>
</tr>
<tr>
<td>8</td>
<td>Deaths in 2018</td>
<td>1,986</td>
<td>197</td>
<td>3,297,083</td>
</tr>
<tr>
<td>9</td>
<td>Demi Lovato</td>
<td>40</td>
<td>20</td>
<td>2,943,525</td>
</tr>
<tr>
<td>10</td>
<td>Ant-Man and the Wasp</td>
<td>625</td>
<td>152</td>
<td>2,916,170</td>
</tr>
</tbody>
</table>

Some pages listed may be false positives. Known false positives: text.
Langviews Analysis
Pageviews of a page across all languages

Do another query

Jair Bolsonaro 10/22/2018 - 11/11/2018

<table>
<thead>
<tr>
<th>#</th>
<th>Language</th>
<th>Page title</th>
<th>Badges</th>
<th>Pageviews</th>
<th>Daily average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>num-languages</td>
<td>unique-titles</td>
<td></td>
<td>2,388,527</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>pt</td>
<td>Jair Bolsonaro</td>
<td></td>
<td>1,156,777</td>
<td>55,085</td>
</tr>
<tr>
<td>2</td>
<td>en</td>
<td>Jair Bolsonaro</td>
<td></td>
<td>583,581</td>
<td>27,790</td>
</tr>
<tr>
<td>3</td>
<td>es</td>
<td>Jair Bolsonaro</td>
<td></td>
<td>217,074</td>
<td>10,337</td>
</tr>
<tr>
<td>4</td>
<td>fr</td>
<td>Jair Bolsonaro</td>
<td></td>
<td>142,608</td>
<td>6,791</td>
</tr>
<tr>
<td>5</td>
<td>de</td>
<td>Jair Bolsonaro</td>
<td></td>
<td>89,745</td>
<td>4,274</td>
</tr>
</tbody>
</table>
https://stats.wikimedia.org/v2/#/es.wikipedia.org/reading/pageviews-by-country
Click dataset

The Wikipedia Clickstream dataset contains counts of (referer, resource) pairs extracted from the request logs of Wikipedia. A referer is an HTTP header field that identifies the address of the webpage that linked to the resource being requested. The data shows how people get to a Wikipedia article and what links they click on.

Automatic monthly releases of datasets for Wikipedia in English, Russian, German, Spanish and Japanese.

ORES (Objective Revision Evaluation Service) is a web service and API that provides machine learning as a service for Wikimedia projects maintained by the Scoring Platform team.

- [http://ores.wikimedia.org](http://ores.wikimedia.org)
Ores

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

http://ores.wmflabs.org/scores/enwiki/damaging/638307884

“damaging”: {
  “prediction”: false,
  “probability”: {
    “false”: 0.8683,
    “true”: 0.1317
  }
}

English Wikipedia revisions as seen by editors (left) and as scored by ORES’ “damage” prediction model (right).

https://blog.wikimedia.org/2015/11/30/artificial-intelligence-x-ray-specs/
Toolforge is a hosting environment for developers working on services that provide value to the Wikimedia movement. These services allow developers to easily do ad hoc analytics, administer bots, run webservices, and generally create tools to help editors and other volunteers in their work. The environment also includes access to a variety of data services.

● https://tools.wmflabs.org/
PAWS

A Web Shell (PAWS) is a Jupyter notebooks deployment that has been customized to make interacting with Wikimedia wikis easier.
Event Streams

EventStreams is a web service that exposes continuous streams of **structured event data**. It does so over HTTP using chunked transfer encoding following the Server-Sent Events protocol. EventStreams can be consumed directly via HTTP, but is more commonly used via a client library.

- [https://wikitech.wikimedia.org/wiki/EventStreams](https://wikitech.wikimedia.org/wiki/EventStreams)
Working Examples


https://github.com/digitalTranshumant/Wiki-examples