Wikidata

... 

Bringing structured data to Wikipedia with 16000 volunteers

Lydia Pintscher  @nighrose
Everyone needs access to knowledge!
$\frac{1}{2}$ Billion

Wikipedia readers/month
nearly 300 Wikipedia language editions
nearly 300 Wikipedia language editions

79 Wikipedia language editions with an article about the Zika virus
English Wikipedia
Russian Wikipedia
Japanese Wikipedia
Knowledge base
Community
**Mona Lisa** *(Q12418)*

- **painting by Leonardo da Vinci**
- **La Joconde** | **La Gioconda**

### In more languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Mona Lisa</td>
<td>painting by Leonardo da Vinci</td>
<td>La Joconde</td>
</tr>
<tr>
<td>German</td>
<td>Mona Lisa</td>
<td>Ölgemälde von Leonardo da Vinci</td>
<td>La Gioconda</td>
</tr>
<tr>
<td>Spanish</td>
<td>Mona Lisa</td>
<td>cuadro de Leonardo da Vinci</td>
<td>La Gioconda</td>
</tr>
<tr>
<td>Latin</td>
<td>lucunda</td>
<td>pictura quam Leonardus Vinciis inter annos 1503 et 1506 fecit</td>
<td>Mona Lisa</td>
</tr>
</tbody>
</table>

### More languages

**Statements**

- **instance of**
  - painting
  - 0 references
  - + add reference
  - + add

- **creator**
  - Leonardo da Vinci
  - 1 reference
  - + add
Current state of Wikidata

- 24 Million items
- 2700 properties
- 120 Million statements
Current state of Wikidata

16000 editors

3rd most active Wikimedia project
(after English Wikipedia and Commons)

1.5 Million SPARQL queries/day
What makes Wikidata special?
Community
People before data!
We are not trying to determine the truth but record what sources say
date of birth

2 BCE  Gregorian

- 1 reference

7 BCE  Gregorian

- 1 reference

+ add
<table>
<thead>
<tr>
<th>Instance of</th>
<th>1 reference</th>
<th>1 reference</th>
<th>1 reference</th>
<th>1 reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>deity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mythical character</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human biblical character</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>human</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We are not restricting the world
but acknowledge that the world is weird and complex
spouse

Eiffel Tower

1 reference
## February 30 (Q37096)

**date**

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>February 30</td>
<td>date</td>
<td>30 February</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>February 30th</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30th of February</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feb 30</td>
</tr>
<tr>
<td>German</td>
<td>30. Februar</td>
<td>Datum</td>
<td>30 February</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>February 30th</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30th of February</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Feb 30</td>
</tr>
<tr>
<td>Spanish</td>
<td>30 de febrero</td>
<td>fecha que ha sido utilizada esporádicamente en</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>diversos países</td>
<td></td>
</tr>
<tr>
<td>Latin</td>
<td>30 Februarii</td>
<td>No description defined</td>
<td></td>
</tr>
</tbody>
</table>
=> multilingual and multicultural project that acknowledges knowledge diversity
Difference to DBpedia?
Extraction from vs. injection of data into Wikipedia
Direct vs. indirect human curation of the data
Different vs. same data set for different languages
Release vs immediate availability
DBpedia and Wikidata have complementary strengths!

Wikidata: data collection, curation and maintenance -> up-to-date and with a community

DBpedia: skills and knowledge to do analysis of data and ontology, annotation, extraction and more
Lessons learned so far
The world is really really complicated

And Wikipedians are world champions in showing you where and how

There is no single truth. Deal with knowledge diversity!

The world is screwed up. Don’t try to force it into neat little boxes!
It is possible to build tools to deal with a really really complicated world

And we should build more of them!
Data must be used and maintained

When people hear knowledge base they think place to dump data. No!

Data maintenance is a huge undertaking.

It needs good (but not complex!) tools.
Even with hard-to-use and basic tools people can build amazing things.

Build flexible and open tools.
Being able to tinker is crucial.
Let people help themselves.
The South Pole Telescope (SPT) is a 10 meter (394 in) diameter telescope located at the Amundsen–Scott South Pole Station, Antarctica. The telescope is designed for observations in the microwave, millimeter-wave, and submillimeter-wave regions of the electromagnetic spectrum, with the particular design goal of measuring the faint, diffuse emission from the cosmic microwave background (CMB). The first major survey with the SPT—designed to find distant, massive, clusters of galaxies through their interaction with the CMB, with the goal of constraining the dark energy equation of state—was completed in October 2011. In early 2012, a new camera was installed on the SPT with even greater sensitivity and the capability to measure the polarization of incoming light. This camera is designed to measure the so-called “B-mode” or “curl” component of the polarized CMB, leading to constraints on the mass of the neutrino and the energy scale of inflation.

The SPT collaboration is made up of over a dozen (mostly North American) institutions, including the University of Chicago, the University of California-Berkeley, Case Western Reserve University, Harvard/Smithsonian Astrophysical Observatory, the University of Colorado-Boulder, McGill University, The University of Illinois at Urbana-Champaign, University of California at Davis, Ludwig Maximilian University of Munich, Argonne National Laboratory, and the National Institute for Standards and Technology. It is funded by the National Science Foundation.

Microwave/millimeter-wave observations at the South Pole

The South Pole is the premier observing site in the world for millimeter-wavelength observations. The Pole's high altitude (2.8 km/1.7 mi above sea level) means the atmosphere is thin, and the extreme cold keeps the amount of water vapor in the air low.[3] This is particularly important for observing at millimeter wavelengths, where incoming signals can be absorbed by water vapor, and where water vapor emits radiation that can be confused with astronomical signals. Because the sun does not rise and set daily, the atmosphere at the pole is particularly stable. Further, there is no interference from the sun in the millimeter range during the months of polar night.

The telescope

The telescope is a 10-meter (394 in) diameter off-axis Gregorian telescope in an altazimuth mount (at the poles, an altazimuth mount is effectively identical to an equatorial mount). It
<table>
<thead>
<tr>
<th>Catalog</th>
<th>Description</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>6DEG</td>
<td>Six degrees of Francis Bacon</td>
<td>Manually matched</td>
</tr>
<tr>
<td>AAT</td>
<td>Art &amp; Architecture Thesaurus by the Getty Research Institute</td>
<td>Automatically matched</td>
</tr>
<tr>
<td>ACAD</td>
<td>A Cambridge Alumni Database</td>
<td>Not on Wikidata</td>
</tr>
<tr>
<td>ADB</td>
<td>Australian Dictionary of Biography</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AGOHRERA</td>
<td>AGORHA event ID</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AGOHRHA work</td>
<td>AGORHA work ID</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AGORHA person</td>
<td>identifier for a person or institution in the Agorha databases...</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AUS-ORDER</td>
<td>General Division of the Order of Australia</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AWLD</td>
<td>UNESCO Atlas of the World’s Languages in Danger</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AcademiaNet</td>
<td>Database for excellent female scientists</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>AdS</td>
<td>Archives du Spectacle: people in theater and live performers</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>Appletons</td>
<td>Appletons' Cyclopaedia of American Biography</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>Architectes diocésains</td>
<td>French diocesan architects</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>Art UK venue</td>
<td>authority control identifier for venues housing art, in the ...</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>Artsy</td>
<td>Artsy artists</td>
<td>N/A Unmatched</td>
</tr>
<tr>
<td>Available games</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Author Language game</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Populate the Languages written and spoken (P1412) property by guessing from notable works or citizenship. Report problems to User:J6an</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kian game</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kian suggestions to add statements in items based on categories in Wikipedia articles. Contact Amir if a model has too much incorrect suggestions. 17 languages are supported. Source code can be found in <a href="#">here</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary sources</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check if a third-party site is a source for a statement.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Items without image</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>These items have no image, but there are some on their Wikipedia articles.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Source MetaData</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify a scientist on a publication. Is it the one suggested?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Les Misérables

Les Misérables (pronounced /lɛs ˈmizərəbəl/ or /lɛz ˈmizərəbəl/), a French historical novel by Victor Hugo, first published in 1862, that is considered one of the greatest novels.

author: Victor Hugo
1862

The Hunchback of Notre Dame

The Hunchback of Notre-Dame (French: Notre-Dame de Paris) is a French Romantic/Gothic novel by Victor Hugo published in 1831. The title refers to the Notre Dame Cathedral in Paris, on which the story is...

author: Victor Hugo
1831

Ninety-Three

Ninety-Three (Quatrevingt-treize) is the last novel by the French writer Victor Hugo. Published in 1874, shortly after the bloody upheaval of the Paris Commune, the novel concerns the Revolt in the Vendée and...

author: Victor Hugo

Cromwell

Cromwell is a play by Victor Hugo, written in 1827. It was influenced by Hugo's literary circle, which identified itself as Romanticist and chose as a model dramatist Shakespeare instead of the Classicists Jean Racine and...

author: Victor Hugo
1827

The Man Who Laughs

The Man Who Laughs (also published under the title By Order of the King) is a novel by Victor Hugo, originally published in April 1869 under the French title L'Homme qui rit. Although among Hugo's most obscure works...

author: Victor Hugo
Ask me anything

Who is the director of "Her"?

This query engine is the result of the Master's degree project of seven students.
If you are interested by some technical aspects, please have a look at the project website.
Who is the director of "Her"?

Spike Jonze  🍿  🏤

Spike Jonze (pronounced "Jones" /dʒoʊNZ/; born Adam Spiegel; October 22, 1969) is an American director, producer, screenwriter and actor, whose work includes music videos, commercials, film and television. He started his feature film directing career with Being John Malkovich (1999) and Adaptation (2002), both written by Charlie Kaufman, and then started movies with screenplays of his own with Where the Wild Things Are (2009) and Her (2013). Jonze is well known for his music video collaborations with Fatboy Slim, Weezer, Beastie Boys, Björk and Kanye West.  Wikipedia

This query engine is the result of the Master's degree project of seven students.

If you are interested by some technical aspects, please have a look at the project website.
Data is one of the most important resources of our time and we are an equalizing force
Important remaining struggles

Help, pretty please?

- Data curation and maintenance -> better tools and community growth
- Ensuring data quality -> references and more
- Data reusability
- Analysis of the data for blind spots and biases
- Large scale consistency
What’s next for Wikidata?

- Automated list generation
- Better connection in the linked open data web
- Support for multimedia meta data (Wikimedia Commons)
- Support for lexicographical data (Wiktionary)
Find more maps showing coverage of individual Wikipedias at https://ddll.inf.tu-dresden.de/web/Wikidata/Maps-06-2015/en

wikidata.org
@wikidata
#wikidata