

Wikimedia workshop

Wikipedia:
by the people,
for the people,
about the people.

Terminology

- Wikimedia –community and broader ‘movement’
- Wikipedia – the first project of many

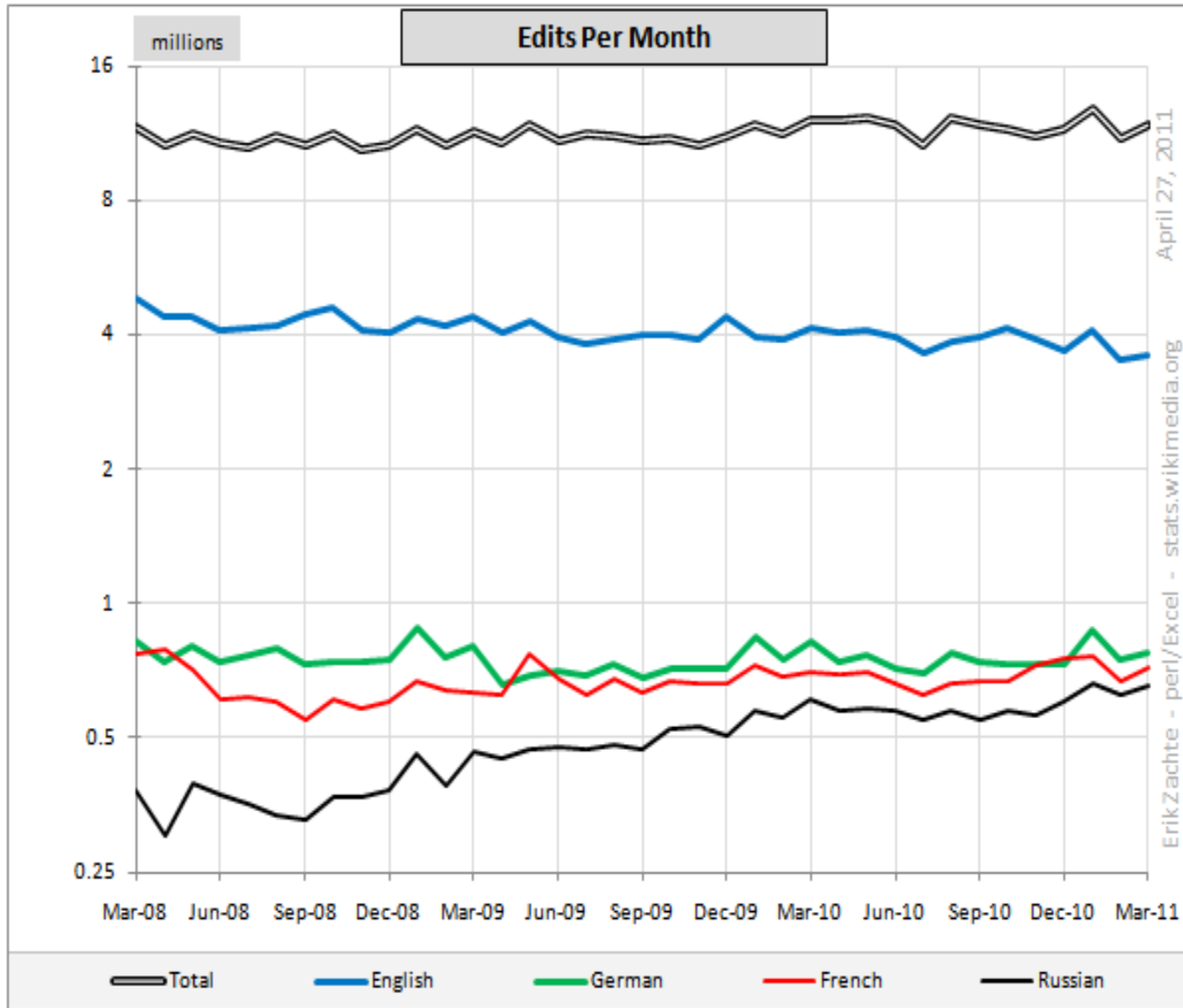
Other projects:

- Commons – images, sounds, video
- Wiktionary - dictionary
- Wikisource – library of public domain works, with volunteers transcribing scans, like Project Gutenberg
- Wikinews – citizen journalism
- Wikibooks – text books
- Wikiversity – academic work space
- Wikimedia Foundation – maintains the servers
- Wikimedia Australia
national non-profit user group

Wikipedia in a nutshell

- Free – libre – Creative Commons Share-alike
<http://freedomdefined.org/>
- Free – no cost to users
- Free – anyone can edit, no barrier to entry
- Free – citizen run – democratic - anarchy
- Free – freeform – no imposed structure
- Scope – the sum of all knowledge, in all languages
 - Library of Alexandria
 - Tower of Babel
- 10 years old

10 million edits/month



Source: Wikimedia Monthly Report Card

Wikipedia structure

- Namespaces
 - Main namespace – namespace “0” – the content
 - User: about participants
 - Wikipedia: about project (WP: is an alias)
 - Talk:, User_talk: Wikipedia_talk: – discussion page
- WikiProjects
 - <http://enwp.org/WP:Australia>
- Noticeboards
 - <http://enwp.org/WP:RSN> - reliable sources
 - <http://enwp.org/WP:COIN> - conflict of interest
 - <http://enwp.org/WP:BLPN> - living people
 - <http://enwp.org/WP:AUSN> - Australia

Data access methods

- HTML parsing – Internal reports
- XML Dumps
- API
- Microformats, RDFa
- Dbpedia

Data: XML Dumps

- http://meta.wikimedia.org/wiki/Data_dumps
- <http://dumps.wikimedia.org/>
- <http://dumps.wikimedia.org/enwiki/latest/>
- <http://dumps.wikimedia.org/enwiki/20110405/>
- Complete copy of every project
 - English Wikipedia: 350 Gb
 - German Wikipedia: 51 Gb
- No media
- Smaller subsets of each project
- APIs read the XML
- Import XML into MySQL database and use SQL

Data: Internal reports

- **Special Pages**
 - <http://enwp.org/Special:Specialpages>
- **Recent changes**
 - <http://enwp.org/Special:Recentchanges>
- **New pages**
 - <http://enwp.org/Special:Newpages>
- **Abuse filter**
 - <http://enwp.org/Special:AbuseLog>
 - <http://enwp.org/Special:AbuseFilter/410>

Data: API

- Can be used to edit content and perform actions (e.g. delete)
- Mostly used for querying metadata and lists
- Formats: JSON, XML, YAML, PHP, etc

http://www.mediawiki.org/wiki/API:Main_page

<http://www.mediawiki.org/wiki/API:Recentchanges>

[http://en.wikipedia.org/w/api.php?action=query&list=recentchanges
&rcprop=title|ids|sizes|flags|user&rclimit=100&rcnamespace=0](http://en.wikipedia.org/w/api.php?action=query&list=recentchanges&rcprop=title|ids|sizes|flags|user&rclimit=100&rcnamespace=0)

Data: microformats

- Microformats are semantic markup in HTML
- Example in Wikipedia –
 - `{{cite journal}}`
 - http://enwp.org/Template:cite_journal
 - <http://enwp.org/Wikipedia:COinS>
 - ContextObjects in Spans
 - bibliographic metadata in HTML
- Zotero – open source reference management software

Zotero

The screenshot shows the Zotero application interface. At the top, a Mozilla Firefox browser window displays the Wikipedia page for "King Arthur". A "Select Items" dialog box is open, prompting the user to "Select which items you'd like to add to your library". The dialog lists the following items:

- King Arthur in Antiquity
- The Discovery of King Arthur
- The New Arthurian Encyclopedia
- The Quest for Arthur's Britain
- The Mammoth Book of King Arthur
- King Arthur: Hero and Legend

Buttons for "Select All", "Deselect All", "OK", and "Cancel" are visible at the bottom of the dialog. Below the dialog, the Zotero interface is visible, showing a sidebar with "My Library", "King Arthur", and "Trash" folders. The main area displays a table with columns for "Title" and "Creator", and a message "No items selected".

Data: dbPedia

- Wikipedia is one of many data sources
- 3.5 million things described
- 1.67 million classified into a consistent Ontology
 - _ 462,000 places
 - _ 364,000 persons
 - _ 169,000 species
 - _ 148,000 organisations
 - _ 5,200 species
- 5.9 million links to external web pages
- 6.5 million external links into other RDF datasets
- 1.85 million links to images
- Creative Commons & Gnu Free Documentation License
- Web of Open Data : several billion RDF triples (assertions)
- <http://opendefinition.org/>

Data: dbPedia

<http://dbpedia.org/sparql>

```
SELECT DISTINCT ?politician WHERE { ?  
politician a <http://dbpedia.org/ontology/OfficeHolder> . ?  
politician ?related <http://dbpedia.org/resource/Canberra> }
```

Result:

- http://dbpedia.org/resource/Ben_Chifley
- http://dbpedia.org/resource/John_Dedman
- http://dbpedia.org/resource/Steve_Whan
- http://dbpedia.org/resource/Patricia_Hewitt
- http://dbpedia.org/resource/Satendra_Nandan
– Resides in Canberra

Frameworks

Python

PHP

Perl

.NET

Java

JavaScript

Ruby

Scheme

Lisp

Haskell

Tcl

Qt/C++

- http://en.wikipedia.org/wiki/Wikipedia:Creating_a_bot

Frameworks: Pywiki

- Originally called “pywikipediabot”
- Most mature
- Originally parsed the HTML internal reports
- Now uses the API
- Open source
- Python
- Includes many tools

Replace.py – search and replace

Pagefromfile.py – bulk page upload

<http://www.mediawiki.org/wiki/Pywikipediabot>

Tools: automated reports

- Article creation monitoring
 - Copyright infringement reports
- Edit monitoring
 - Vandalism detection and removal
- Account creation monitoring
 - Bad usernames

http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Au
Article_alerts

Tools: manual reports

- Article deletion monitoring
 - <http://enwp.org/WP:DSA>
 - Streamlined with JavaScript
- Article assessment
 - <http://enwp.org/Talk:Canberra>
 - “Featured article”!
- [http://enwp.org/Wikipedia:WikiProject Australia/Assessment](http://enwp.org/Wikipedia:WikiProject_Australia/Assessment)
- http://enwp.org/wiki/Category:Unknown-importance_Canberra_articles

Tools: Wiki checker

- <http://en.wikichecker.com/>
- Edit wars
- User contributions
- Article history

3,922 edits on article: [Canberra](#) 

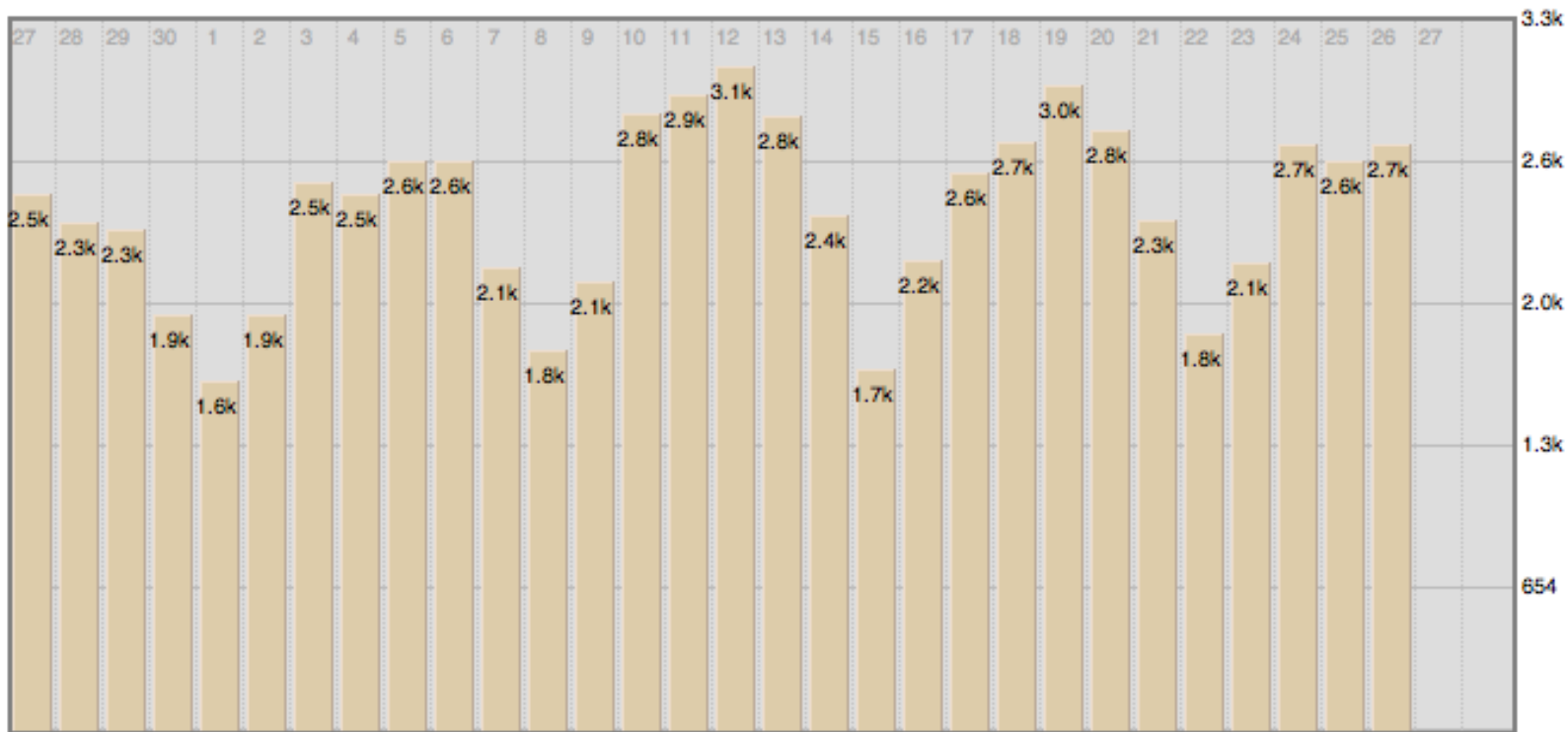
- The 1st edition of the actual version: 6:12:29 13/5/20
- Maintained 3,301 days
- One edit par 0.8 days (20.2 hours) on average.
- Number of editors: 1,471
- Anonymous user edited 1,195 times
- Edit count of the top 10% frequent users: 2,179

Tools: article information

1. <http://enwp.org/Canberra>
2. Click 'history'
 1. Page views per day
 - How important is this topic?
 2. Contributor overview
 - Who is responsible for this content
 3. Revision search
 - When was a fact added?
 4. Watchers
 - How many people are 'watching' this page

Page views

[Canberra](#) has been viewed 71916 times in the last 30 days. This article ranked 7807 in traffic on en.wikipedia.org.



Source: <http://stats.grok.se/en/latest/Canberra> (accessed May 26, 2011)

wikistats

- <http://stats.wikimedia.org/>
- Activity
 - Number of edits
 - Number of active accounts
 - normal – 10 edits per month
 - very active – 100 edits per month
- Growth stats
 - New files
 - New pages
 - New accounts
- <http://stats.wikimedia.org/EN/ChartsWikipediaEN.htm>

Example: SLQ

- State Library of Queensland
 - Desire to engage and participate in an online community
 - Bidirectional sync between library repository and Wikimedia Commons
 - dublin core <-> wiki
 - Want to provide training to Queenslanders
 - Including cross-training
 - Training in Miles - who is in Miles?
- <http://enwp.org/wiki/>
Category:Wikipedians_in_Queensland

Example: APC

- Australian Paralympic Committee
 - Want to write a history book
 - Economic benefit of ‘community engagement’ and ‘living book’
- http://en.wikiversity.org/wiki/The_History_of_the_Paralympic_Movement_in_Australia
- Low content and contributors
 - Need to build the community
 - Monitor and facilitate
- <http://en.wikipedia.org/w/index.php?title=Special:RecentChangesLinked&limit=50>
Australia_at_the_Paralympics

What next?

- Outreach
 - <http://outreach.wikimedia.org/>
- Research about the projects and community
 - <http://meta.wikimedia.org/wiki/Research>
- Improvements to the software
 - http://www.mediawiki.org/wiki/Research_Data_Proposals
 - <http://www.mediawiki.org/wiki/Manual:Wikistats/API>
- Collaboration with external organisations on objective-driven projects involving Wikipedia.

Questions

- How do you want to engage with Wikipedia?
- Do you know of a problem with Wikipedia content?