

Wikipedia:Database download

From Wikipedia, the free encyclopedia

For scheduling, related tools etc., see m:Data dumps.

"WP:DD" redirects here. For Duplication detector, see Wikipedia:Duplication detector.

Wikipedia offers free copies of all available content to interested users. These databases can be used for mirroring, personal use, informal backups, offline use or database queries (such as for Wikipedia:Maintenance). All text content is multi-licensed under the Creative Commons Attribution-ShareAlike 3.0 License (CC-BY-SA) and the GNU Free Documentation License (GFDL). Images and other files are available under different terms, as detailed on their description pages. For our advice about complying with these licenses, see Wikipedia:Copyrights.

Contents

- 1 Offline Wikipedia readers
 - 2 Where do I get it?
- 2.1 English-language Wikipedia
 - 2.2 Other languages
- 3 Where are the uploaded files (image, audio, video, etc., files)?
- 4 Dealing with compressed files
 - 5 Dealing with large files
- 5.1 File system limits
 - 5.2 Operating system limits
 - 5.3 Tips
 - 5.3.1 Detect corrupted files
 - 5.3.2 Reformatting external USB drives
 - 5.3.3 Linux and Unix
- 6 Why not just retrieve data from wikipedia.org at runtime?
- 6.1 Please do not use a web crawler
 - 6.1.1 Sample blocked crawler email
 - 6.2 Doing SQL queries on the current database dump
- 7 Database schema
- 7.1 SQL schema
 - 7.2 XML schema
- 8 Help to parse dumps for use in scripts
- 8.1 Doing Hadoop MapReduce on the Wikipedia current database dump
- 9 Help to import dumps into MySQL
- 10 Static HTML tree dumps for mirroring or CD distribution
- 10.1 Kiwix
 - 10.2 Aard Dictionary
 - 10.3 E-book
 - 10.4 Wikiviewer for Rockbox
 - 10.5 Old dumps
- 11 Dynamic HTML generation from a local XML database dump
- 11.1 XOWA
 - 11.1.1 Features
 - 11.2 Offline wikipedia reader

- 11.2.1 Main features
 - 11.3 WikiFilter
 - 11.3.1 WikiFilter system requirements
 - 11.3.2 How to set up WikiFilter
 - 11.4 WikiTaxi (for Windows)
 - 11.4.1 WikiTaxi system requirements
 - 11.4.2 WikiTaxi usage
 - 11.5 BzReader and MzReader (for Windows)
 - 11.6 EPWING
- 12 Mirror Building
 - 12.1 WP-MIRROR
- 13 See also
- 14 References
- 15 External links

Offline Wikipedia readers

Some of the many ways to read Wikipedia while offline:

- XOWA (#XOWA)
- Kiwix (§ Kiwix)
- WikiTaxi #WikiTaxi
- WikiReader
- aarddict #Aard Dictionary
- BzReader #BzReader and MzReader (for Windows)
- iPodLinux
- offline-wikipedia #Offline wikipedia reader
- Okawix <http://sourceforge.net/p/okawix/code/HEAD/tree/>
- Selected Wikipedia articles as a PDF, OpenDocument, etc. Wikipedia:Books
- Selected Wikipedia articles as a printed book Help:Books/Printed books
- Wiki as E-Book #E-book
- WikiFilter #WikiFilter
- Wikipedia Featured Articles as a Printed Book <http://www.brandnew.uk.com/wikipedia-as-a-printed-book/>
- Wikipedia on rockbox #Wikiviewer for Rockbox

Some of them are mobile applications -- see "list of Wikipedia mobile applications".

Where do I get it?

English-language Wikipedia

- Dumps from any Wikimedia Foundation project: dumps.wikimedia.org (<https://dumps.wikimedia.org/>) and the Internet Archive
- English Wikipedia dumps in SQL and XML: dumps.wikimedia.org/enwiki/ (<https://dumps.wikimedia.org/enwiki/>) and the Internet Archive (<https://archive.org/search.php?query=subject%3A%22enwiki%22%20AND%20subject%3A%22data%20dumps%22%20AND%20collection%3A%22wikimediadownloads%22>)
 - Download (https://meta.wikimedia.org/wiki/Data_dump_torrents#English_Wikipedia) the data dump using a BitTorrent client (torrenting has many benefits and reduces server load, saving bandwidth costs).
 - [pages-articles.xml.bz2](#) - Current revisions only, no talk or user pages; this is probably what you want, and is approximately 13 GB compressed (expands to over 58 GB when uncompressed).
 - [pages-meta-current.xml.bz2](#) - Current revisions only, all pages (including talk)
 - [abstract.xml.gz](#) - page abstracts
 - [all-titles-in-ns0.gz](#) - Article titles only (with redirects)
 - SQL files for the pages and links are also available
 - All revisions, all pages: **These files expand to multiple terabytes of text. Please only download these if you know you can cope with this quantity of data.** Go to Latest Dumps (<https://dumps.wikimedia.org/enwiki/latest/>) and look out for all the files that have 'pages-meta-history' in their name.
- To download a subset of the database in XML format, such as a specific category or a list of articles see: [Special:Export](#), usage of which is described at [Help:Export](#).
- Wiki front-end software: MediaWiki [1] (<https://www.mediawiki.org/>).
- Database backend software: You want to download MySQL.
- Image dumps: See below.

Other languages

In the dumps.wikimedia.org (<https://dumps.wikimedia.org/>) directory you will find the latest SQL and XML dumps for the projects, not just English. The sub-directories are named for the language code and the appropriate project. Some other directories (e.g. *simple*, *nostalgia*) exist, with the same structure. These dumps are also available from the Internet Archive.

Where are the uploaded files (image, audio, video, etc., files)?

Images and other uploaded media are available from mirrors in addition to being served directly from Wikimedia servers. Bulk download is (as of September 2013) available from mirrors but not offered directly from Wikimedia servers. See the list of current mirrors. You should `rsync` from the mirror, then fill in the missing images from upload.wikimedia.org (https://upload.wikimedia.org); when downloading from upload.wikimedia.org you should throttle yourself to 1 cache miss per second (you can check headers on a response to see if was a hit or miss and then back off when you get a miss) and you shouldn't use more than one or two simultaneous HTTP connections. In any case, make sure you have an accurate user agent string with contact info (email address) so ops can contact you if there's an issue. You should be getting checksums from the mediawiki API and verifying them. The API Etiquette page contains some guidelines, although not all of them apply (for example, because upload.wikimedia.org isn't MediaWiki, there is no `maxlag` parameter).

Unlike most article text, images are not necessarily licensed under the GFDL & CC-BY-SA-3.0. They may be under one of many free licenses, in the public domain, believed to be fair use, or even copyright infringements (which should be deleted). In particular, use of fair use images outside the context of Wikipedia or similar works may be illegal. Images under most licenses require a credit, and possibly other attached copyright information. This information is included in image description pages, which are part of the text dumps available from dumps.wikimedia.org (<https://dumps.wikimedia.org/>). In conclusion, download these images at your own risk (Legal (<https://dumps.wikimedia.org/legal.html>))

Dealing with compressed files

Compressed dump files are significantly compressed, thus after being uncompressed will take up **large** amounts of drive space. A large list of decompression programs are described in *Comparison of file archivers*. The following programs in particular can be used to uncompress `bzip2` `.bz2` `.zip` and `.7z` files.

Windows

Beginning with Windows XP, a basic decompression program enables decompression of zip files.^{[1][2]} Among others, the following can be used to decompress `bzip2` files.

- `bzip2` (command-line) (<ftp://sources.redhat.com/pub/bzip2/v102/bzip2-102-x86-win32.exe>) (from here (<http://sources.redhat.com/bzip2/>)) is available for free under a BSD license.
- 7-Zip is available for free under an LGPL license.
- WinRAR
- WinZip

Macintosh (Mac)

- OS X ships with the command-line `bzip2` tool.

GNU/Linux

- Most GNU/Linux distributions ship with the command-line `bzip2` tool.

Berkeley Software Distribution (BSD)

- Some BSD systems ship with the command-line `bzip2` tool as part of the operating system. Others, such as OpenBSD, provide it as a package which must first be installed.

Notes

1. Some older versions of `bzip2` may not be able to handle files larger than 2 GB, so make sure you have the latest version if you experience any problems.
2. Some older archives are compressed with `gzip`, which is compatible with PKZIP (the most common Windows format).

Dealing with large files

As files grow in size, so does the likelihood they will exceed some limit of a computing device. Each operating system, file system, hard storage device, and software (application) has a maximum file size limit. Each one of these will likely have a different maximum, and the lowest limit of all of them will become the file size limit for a storage device.

The older the software in a computing device, the more likely it will have a 2 GB file limit somewhere in the system. This is due to older software using 32-bit integers for file indexing, which limits file sizes to 2^{31} bytes (2 GB) (for signed integers), or 2^{32} (4 GB) (for unsigned integers). Older C programming libraries have this 2 or 4 GB limit, but

the newer file libraries have been converted to 64-bit integers thus supporting file sizes up to 2⁶³ or 2⁶⁴ bytes (8 or 16 EB).

Before starting a download of a large file, check the storage device to ensure its file system can support files of such a large size, and check the amount of free space to ensure that it can hold the downloaded file.

File system limits

There are two limits for a file system: the file system size limit, and the file system limit. In general, since the file size limit is less than the file system limit, the larger file system limits are a moot point. A large percentage of users assume they can create files up to the size of their storage device, but are wrong in their assumption. For example, a 16 GB storage device formatted as FAT32 file system has a file limit of 4 GB for any single file. The following is a list of the most common file systems, and see Comparison of file systems for additional detailed information.

Windows

- FAT16 supports files up to 4 GB. FAT16 is the factory format of smaller USB drives and all SD cards that are 2 GB or smaller.
- FAT32 supports files up to 4 GB. FAT32 is the factory format of larger USB drives and all SDHC cards that are 4 GB or larger.
- exFAT supports files up to 127 PB. exFAT is the factory format of all SDXC cards, but is incompatible with most flavors of UNIX due to licensing problems.
- NTFS supports files up to 16 TB. NTFS is the default file system for Windows computers, including Windows 2000, Windows XP, and all their successors to date.
- ReFS supports files up to 16 EB.

Macintosh (Mac)

- HFS Plus (HFS+) supports files up to 8 EB on Mac OS X 10.2+ and iOS. HFS+ is the default file system for OS X computers.

Linux

- ext2 and ext3 supports files up to 16 GB, but up to 2 TB with larger block sizes. See http://www.suse.com/~aj/linux_lfs.html for more information.
- ext4 supports files up to 16 TB, using 4 KB block size. (limit removed in e2fsprogs-1.42 (2012) (<https://fedoraproject.org/wiki/Features/F17Ext4Above16T>))
- XFS supports files up to 8 EB.
- ReiserFS supports files up to 1 EB, 8 TB on 32-bit systems.
- JFS supports files up to 4 PB.
- Btrfs supports files up to 16 EB.
- NILFS supports files up to 8 EB.
- YAFFS2 supports files up to 2 GB

FreeBSD

- ZFS supports files up to 16 EB.

FreeBSD and other BSDs

- Unix File System (UFS) supports files up to 8 ZiB.

Operating system limits

Each operating system has internal file system limits for file size and drive size, which is independent of the file system or physical media. If the operating system has any limits lower than the file system or physical media, then the OS limits will be the real limit.

Windows

- Windows 95, 98, ME have a 4 GB limit for all file sizes.
- Windows XP has a 16 TB limit for all file sizes.
- Windows 7 has a 16 TB limit for all file sizes.
- Windows 8, Server 2012 have a 256 TB limit for all file sizes.

Linux

- 32-bit kernel 2.4.x systems have a 2 TB limit for all file systems.
- 64-bit kernel 2.4.x systems have an 8 EB limit for all file systems.
- 32-bit kernel 2.6.x systems without option CONFIG_LBD have a 2 TB limit for all file systems.
- 32-bit kernel 2.6.x systems with option CONFIG_LBD and all 64-bit kernel 2.6.x systems have an 8 ZB limit for all file systems.^[3]

Google Android

Google Android is based on Linux, which determines its base limits.

- Internal storage:
 - Android 2.3 and later uses the ext4 file system.^[4]
 - Android 2.2 and earlier uses the YAFFS2 file system.
- External storage slots:
 - All Android devices should support FAT16, FAT32, ext2 file systems.
 - Android 2.3 and later supports ext4 file system.

Apple iOS (see List of iOS devices)

- All devices support HFS Plus (HFS+) for internal storage. No devices have external storage slots.

Tips

Detect corrupted files

It is useful to check the MD5 sums (provided in a file in the download directory) to make sure the download was complete and accurate. This can be checked by running the "md5sum" command on the files downloaded. Given their sizes, this may take some time to calculate. Due to the technical details of how files are stored, *file sizes* may be reported differently on different filesystems, and so are not necessarily reliable. Also, corruption may have occurred during the download, though this is unlikely.

Reformatting external USB drives

If you plan to download Wikipedia Dump files to one computer and use an external USB flash drive or hard drive to copy them to other computers, then you will run into the 4 GB FAT32 file size limit. To work around this limit, reformat the >4 GB USB drive to a file system that supports larger file sizes. If working exclusively with Windows XP-Vista-7 computers, then reformat the USB drive to NTFS file system.

Linux and Unix

If you seem to be hitting the 2 GB limit, try using wget version 1.10 or greater, cURL version 7.11.1-1 or greater, or a recent version of lynx (using -dump). Also, you can resume downloads (for example wget -c).

Why not just retrieve data from wikipedia.org at runtime?

Suppose you are building a piece of software that at certain points displays information that came from Wikipedia. If you want your program to display the information in a different way than can be seen in the live version, you'll probably need the wikicode that is used to enter it, instead of the finished HTML.

Also, if you want to get all the data, you'll probably want to transfer it in the most efficient way that's possible. The wikipedia.org servers need to do quite a bit of work to convert the wikicode into HTML. That's time consuming both for you and for the wikipedia.org servers, so simply spidering all pages is not the way to go.

To access any article in XML, one at a time, access Special:Export/Title of the article.

Read more about this at Special:Export.

Please be aware that live mirrors of Wikipedia that are dynamically loaded from the Wikimedia servers are prohibited. Please see Wikipedia:Mirrors and forks.

Please do not use a web crawler

Please do not use a web crawler to download large numbers of articles. Aggressive crawling of the server can cause a dramatic slow-down of Wikipedia.

Sample blocked crawler email

IP address *nnn.nnn.nnn.nnn* was retrieving up to 50 pages per second from wikipedia.org addresses. robots.txt (<https://en.wikipedia.org/robots.txt>) has a rate limit of one per second set using the Crawl-delay setting. Please respect that setting. If you must exceed it a little, do so only during the least busy times shown in our site load graphs at **stats.wikimedia.org/EN/ChartsWikipediaZZ.htm** (**<https://stats.wikimedia.org/EN/ChartsWikipediaZZ.htm>**). It's worth noting that to crawl the whole site at one hit per second will take several weeks. The originating IP is now blocked or will be shortly. Please contact us if you want it unblocked. Please don't try to circumvent it – we'll just block your whole IP range.

If you want information on how to get our content more efficiently, we offer a variety of methods, including weekly database dumps which you can load into MySQL and crawl locally at any rate you find convenient. Tools are also available which will do that for you as often as you like once you have the infrastructure in place.

Instead of an email reply you may prefer to visit #mediawiki [connect](https://webchat.freenode.net/?channels=#mediawiki) (<https://webchat.freenode.net/?channels=#mediawiki>) at [irc.freenode.net](https://webchat.freenode.net/?channels=#mediawiki) to discuss your options with our team.

Note that the robots.txt currently has a commented out Crawl-delay:

```
## *at least* 1 second please. preferably more :D
## we're disabling this experimentally 11-09-2006
#Crawl-delay: 1
```

Please be sure to use an intelligent non-zero delay regardless.

Doing SQL queries on the current database dump

You can do SQL queries on the current database dump (as a replacement for the disabled Special:Asksql page).

Database schema

SQL schema

See also: mw:Manual:Database layout

The sql file used to initialize a MediaWiki database can be found here (<https://phabricator.wikimedia.org/diffusion/MW/browse/master/maintenance/tables.sql>).

XML schema

The XML schema for each dump is defined at the top of the file. And also described in the MediaWiki export help page.

Help to parse dumps for use in scripts

- [Wikipedia:Computer help desk/ParseMediaWikiDump](#) describes the Perl `Parse::MediaWikiDump` library, which can parse XML dumps.
- [Wikipedia preprocessor \(wikiprep.pl\)](#) (<http://www.cs.technion.ac.il/~gabr/resources/code/wikiprep>) is a Perl script that preprocesses raw XML dumps and builds link tables, category hierarchies, collects anchor text for each article etc.
- [Wikipedia SQL dump parser](#) (<https://github.com/svick/Wikipedia-Sql-dump-parser>) is a .NET library to read MySQL dumps without the need to use MySQL database
- [Dictionary Builder](#) (<https://github.com/newca12/dictionary-builder>) is a Java program that can parse XML dumps and extract entries in files

Doing Hadoop MapReduce on the Wikipedia current database dump

You can do Hadoop MapReduce queries on the current database dump, but you will need an extension to the `InputRecordFormat` to have each `<page>` `</page>` be a single mapper input. A working set of java methods (`jobControl`, `mapper`, `reducer`, and `XmlInputRecordFormat`) is available at Hadoop on the Wikipedia (<https://tpmoyer-gallery.appspot.com/hadoopWikipedia>)

Help to import dumps into MySQL

See:

- [mw:Manual:Importing XML dumps](#)
- [m:Data_dumps](#)

Static HTML tree dumps for mirroring or CD distribution

MediaWiki 1.5 includes routines to dump a wiki to HTML, rendering the HTML with the same parser used on a live wiki. As the following page states, putting one of these dumps on the web unmodified will constitute a trademark violation. They are intended for private viewing in an intranet or desktop installation.

- If you want to draft a traditional website in Mediawiki and dump it to HTML format, you might want to try `mw2html` (<http://barnesc.blogspot.com/2005/10/mw2html-export-mediawiki-to-static.html>) by User:Connelly.
- If you'd like to help develop dump-to-static HTML tools, please drop us a note on the developers' mailing list.
- Static HTML dumps are now available here (https://dumps.wikimedia.org/other/static_html_dumps/), but are not current.

See also:

- [mw:Alternative parsers](#) lists some other not working options for getting static HTML dumps
- [Wikipedia:Snapshots](#)
- [Wikipedia:TomeRaider database](#)
- <http://sdict.com> hosts a January 2007 snapshot in the open source Sdictionary .dct format
- <http://ahuv.net/wikipedia> hosts October 2010 processed snapshot in the freeware MDict .mdx format

Kiwix

Kiwix (http://wiki.kiwix.org/wiki/Main_Page) for Windows, OSX, GNU/Linux and Android.

- Here for example the English Wikipedia (http://download.kiwix.org/portable/wikipedia_en_all.zip.torrent).

Aard Dictionary

Aard Dictionary (<http://aarddict.org/>) is an Offline Wikipedia reader. No images. Cross-Platform for Windows, Mac, Linux, Android, Maemo. Runs on rooted Nook and Sony PRS-T1 eBooks readers. <https://github.com/aarddict>



E-book

The wiki-as-ebook (<http://wiki-as-ebook.sourceforge.net/>) store provides ebooks created from a large set of Wikipedia articles with grayscale images for e-book-readers (2013).

Wikiviewer for Rockbox

The wikiviewer plugin for rockbox permits viewing converted Wikipedia dumps on many Rockbox devices. It needs a custom build and conversion of the wiki dumps using the instructions available at <http://www.rockbox.org/tracker/4755>. The conversion recompresses the file and splits it into 1 GB files and an index file which all need to be in the same folder on the device or micro sd card.

Old dumps

- The static version of Wikipedia created by Wikimedia: <http://static.wikipedia.org/> Feb. 11, 2013 - This is apparently offline now. There was no content.
- Wiki2static (<http://www.tommasoconforti.com/>) (site down as of October 2005) **was** an experimental program set up by User:Alfio to generate html dumps, inclusive of images, search function and alphabetical index. At the linked site experimental dumps and the script itself can be downloaded. As an example it was used to generate these copies of English Wikipedia 24 April 04 (http://fixedreference.org/en/20040424/wikipedia/Main_Page), Simple Wikipedia 1 May 04 (http://fixedreference.org/simple/20040501/wikipedia/Main_Page)(old database) format and English Wikipedia 24 July 04 (http://july.fixedreference.org/en/20040724/wikipedia/Main_Page)Simple Wikipedia 24 July 04 (http://july.fixedreference.org/simple/20040724/wikipedia/Main_Page), Wikipedia Francais 27 Juillet 2004 (<http://july.fixedreference.org/fr/20040727/wikipedia/Accueil>) (new format). BozMo uses a version to generate periodic static copies at fixed reference (<http://fixedreference.org/>).

Dynamic HTML generation from a local XML database dump

Instead of converting a database dump file to many pieces of static HTML, one can also use a dynamic HTML generator. Browsing a wiki page is just like browsing a Wiki site, but the content is fetched and converted from a local dump file on request from the browser.

XOWA

XOWA is a free, open-source application that helps download Wikipedia to a computer. Access all of Wikipedia offline, without an internet connection! It is currently in the beta stage of development, but is functional. It is available for download here (http://xowa.org/home/wiki/Help/Download_XOWA.html).

Features

- Displays all articles from Wikipedia without an internet connection.
- Download a complete, recent copy of English Wikipedia.
- Display 5.2+ million articles in full HTML formatting.
- Show images within an article. Access 3.7+ million images using the offline image databases.
- Works with any Wikimedia wiki, including Wikipedia, Wiktionary, Wikisource, Wikiquote, Wikivoyage (also some non-wmf dumps)
- Works with any non-English language wiki such as French Wikipedia, German Wikisource, Dutch Wikivoyage, etc.
- Works with other specialized wikis such as Wikidata, Wikimedia Commons, Wikispecies, or any other MediaWiki generated dump
- Set up over 660+ other wikis including:
 - English Wiktionary
 - English Wikisource
 - English Wikiquote
 - English Wikivoyage
 - Non-English wikis, such as French Wiktionary, German Wikisource, Dutch Wikivoyage
 - Wikidata
 - Wikimedia Commons
 - Wikispecies
 - ... and many more!
- Update your wiki whenever you want, using Wikimedia's database backups.

- Navigate between offline wikis. Click on "Look up this word in Wiktionary" and instantly view the page in Wiktionary.
- Edit articles to remove vandalism or errors.
- Install to a flash memory card for portability to other machines.
- Run on Windows, Linux and Mac OS X.
- View the HTML for any wiki page.
- Search for any page by title using a Wikipedia-like Search box.
- Browse pages by alphabetical order using Special:AllPages.
- Find a word on a page.
- Access a history of viewed pages.
- Bookmark your favorite pages.
- Downloads images and other files on demand (when connected to the internet)
- Sets up Simple Wikipedia in less than 5 minutes
- Can be customized at many levels: from keyboard shortcuts to HTML layouts to internal options

Offline wikipedia reader

(for Mac OS X, GNU/Linux, FreeBSD/OpenBSD/NetBSD, and other Unices)

The offline-wikipedia project (<http://owi.sourceforge.net/>) provides a very effective way to get an offline version of Wikipedia. It uses entirely free software. Packages are available for Ubuntu and soon for other Linux distributions.

Main features

1. Very fast searching
2. Keyword (actually, title words) based searching
3. Search produces multiple possible articles: you can choose amongst them
4. LaTeX based rendering for mathematical formulae
5. Minimal space requirements: the original .bz2 file plus the index
6. Very fast installation (a matter of hours) compared to loading the dump into MySQL

WikiFilter

WikiFilter (<http://wikifilter.sourceforge.net/>) is a program which allows you to browse over 100 dump files without visiting a Wiki site.

WikiFilter system requirements

- A recent Windows version (WinXP is fine; Win98 and WinME won't work because they don't have NTFS support)
- A fair bit of hard drive space (To install you will need about 12 - 15 Gigabytes; afterwards you will only need about 10 Gigabytes)

How to set up WikiFilter

1. Start downloading a Wikipedia database dump file such as an English Wikipedia dump (<https://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles.xml.bz2>). It is best to use a download manager such as GetRight so you can resume downloading the file even if your computer crashes or is shut down during the download.
2. Download XAMPP from [2] (http://sourceforge.net/project/showfiles.php?group_id=61776&package_id=89552) (you must get the 1.5.0 version for it to work). Make sure to pick the file whose filename ends with .exe
3. Install/extract it to C:\XAMPP\.
4. Download WikiFilter 2.3 from this site: <http://sourceforge.net/projects/wikifilter>. You will have a choice of files to download, so make sure that you pick the 2.3 version. Extract it to C:\WIKIFILTER.
5. Copy the WikiFilter.so into your C:\XAMPP\apache\modules folder.
6. Edit your C:\xampp\apache\conf\httpd.conf file, and add the following line:
 - LoadModule WikiFilter_module "C:/XAMPP/ apache/modules/WikiFilter.so"
7. When your Wikipedia file has finished downloading, uncompress it into your C:\WIKIFILTER folder. (I used WinRAR <http://www.rarlab.com/> demo version - BitZipper <http://www.bitzipper.com/winrar.html> works well too.)
8. Run WikiFilter (WikiIndex.exe), and go to your C:\WIKIFILTER folder, and drag and drop the XML file into the window, click Load, then Start.
9. After it finishes, exit the window, and go to your C:\XAMPP\ folder. Run the setup_xampp.bat file to configure xampp.
10. When you finish with that, run the Xampp-Control.exe file, and start Apache.
11. Browse to <http://localhost/wiki> and see if it works
 - If it doesn't work, see the forums (http://sourceforge.net/forum/forum.php?forum_id=495411).

WikiTaxi (for Windows)

WikiTaxi (<http://www.wikitaxi.org>) is an offline-reader for wikis in MediaWiki format. It enables users to search and browse popular wikis like Wikipedia, Wikiquote, or WikiNews, without being connected to the Internet. WikiTaxi works well with different languages like English, German, Turkish, and others but has a problem with right-to-left language scripts. WikiTaxi does not display images.

WikiTaxi system requirements

- Any Windows version starting from Windows 95 or later. Large File support (greater than 4 GB which requires an exFAT filesystem) for the huge wikis (English only at the time of this writing).
- It also works on Linux with Wine.
- 16 MB RAM minimum for the WikiTaxi reader, 128 MB recommended for the importer (more for speed).
- Storage space for the WikiTaxi database. This requires about 11.7 GiB for the English Wikipedia (as of 5 April 2011), 2 GB for German, less for other Wikis. These figures are likely to grow in the future.

WikiTaxi usage

1. Download WikiTaxi and extract to an empty folder. No installation is otherwise required.
2. Download the XML database dump (*.xml.bz2) of your favorite wiki.
3. Run WikiTaxi_Importer.exe to import the database dump into a WikiTaxi database. The importer takes care to uncompress the dump as it imports, so make sure to save your drive space and do not uncompress beforehand.
4. When the import is finished, start up WikiTaxi.exe and open the generated database file. You can start searching, browsing, and reading immediately.
5. After a successful import, the XML dump file is no longer needed and can be deleted to reclaim disk space.
6. To update an offline Wiki for WikiTaxi, download and import a more recent database dump.

For WikiTaxi reading, only two files are required: WikiTaxi.exe and the .taxi database. Copy them to any storage device (memory stick or memory card) or burn them to a CD or DVD and take your Wikipedia with you wherever you go!

BzReader and MzReader (for Windows)

BzReader (<https://code.google.com/p/bzreader/>) is an offline Wikipedia reader with fast search capabilities. It renders the Wiki text into HTML and doesn't need to decompress the database. Requires Microsoft .NET framework 2.0.

MzReader (<http://homepage.ntlworld.com/bharat.vadera/MzReader/>) by Mun206 works with (though is not affiliated with) BzReader, and allows further rendering of wikicode into better HTML, including an interpretation of the monobook skin. It aims to make pages more readable. Requires Microsoft Visual Basic 6.0 Runtime, which is not supplied with the download. Also requires Inet Control and Internet Controls (Internet Explorer 6 ActiveX), which are packaged with the download.

EPWING

Offline Wikipedia database in EPWING dictionary format, which is common and an out-dated Japanese Industrial Standards (JIS) in Japan, can be read including thumbnail images and tables with some rendering limits, on any systems where a reader is available (Booookends (<https://sites.google.com/site/booookends>)). There are many free and commercial readers for Windows (including Mobile), Mac OS X, iOS (iPhone, iPad), Android, Unix-Linux-BSD, DOS, and Java-based browser applications (EPWING Viewers (<http://maximilk.web.fc2.com/viewers.htm>)).

Mirror Building

WP-MIRROR

WP-MIRROR is a free utility for mirroring any desired set of WMF wikis. That is, it builds a wiki farm that the user can browse locally. WP-MIRROR builds a complete mirror with original size media files. WP-MIRROR is available for download (<http://www.nongnu.org/wp-mirror/>).

See also

- DBpedia
- WikiReader
- m:Export
- m:Help:Downloading pages
- m:Import
- Meta:Data dumps#Other tools, for related tools, e.g. extractors and "dump readers"
- Wikipedia:Wikipedia CD Selection
- Wikipedia:Size of Wikipedia
- meta:Mirroring Wikimedia project XML dumps
- meta:Static version tools

References

1. "Benchmarked: What's the Best File Compression Format?" (<http://www.howtogeek.com/200698/benchmarked-whats-the-best-file-compression-format/>). *How To Geek*. How-To Geek, LLC. Retrieved 18 January 2017.
2. "Zip and unzip files" (<https://support.microsoft.com/en-us/help/14200/windows-compress-uncompress-zip-files>). *Microsoft*. Microsoft. Retrieved 18 January 2017.
3. Large File Support in Linux (http://www.suse.com/~aj/linux_lfs.html)
4. Android 2.2 and before used YAFFS file system; December 14, 2010. (<http://www.h-online.com/open/news/item/Android-2-3-Gingerbread-to-use-Ext4-file-system-1152775.html>)

External links

- Wikimedia downloads (<https://dumps.wikimedia.org/>).
- Domas visits logs (<http://dammit.lt/wikistats/>) (read this! (<http://infodisiac.com/blog/2010/07/wikimedia-page-view-s-some-good-and-bad-news/>)). Also, old data (https://archive.org/details/wikipedia_visitor_stats_200712) in the Internet Archive.
- Wikimedia mailing lists archives.
- User:Emijrp/Wikipedia Archive. An effort to find all the Wiki[mp]edia available data, and to encourage people to download it and save it around the globe.
- Script to download all Wikipedia 7z dumps (<https://github.com/WikiTeam/wikiteam/blob/master/wikipediadownloader.py>).

Retrieved from "https://en.wikipedia.org/w/index.php?title=Wikipedia:Database_download&oldid=782277350"

Categories: Wikipedia downloads | Wikipedia database reports

-
- This page was last edited on 25 May 2017, at 22:01.
 - Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.