Manual:Importing XML dumps

From MediaWiki.org

This page describes methods to **import XML dumps**.

MediaWiki uses an abstract XML based format for content dumps. This is what Special:Export generates, and also what is used for XML dumps of Wikipedia and other Wikimedia sites, as described in meta:Data dumps. The format is explained in meta:Help:Export in some detail.

There are several methods for importing such XML dumps:

Contents

Using Special:Import

Special:Import can be used by wiki users with the `import` permission (per default, users in the `sysop` group) to import a small number of pages (about 100 should be safe). Trying to import large dumps this way may result in timeouts or connection failures. See meta:Help:Import for a detailed description.

See Manual:XML Import file manipulation in CSharp for a C# code sample that manipulates an XML import file.

Possible Problems

For using Transwiki-Import PHP safe_mode must be off and open_basedir must be empty. Otherwhise the import fails.

Using importDump.php, if you have shell access

Recommended method for general use, but slow for very big data sets. For very large amounts of data, such as a dump of a big Wikipedia, use mwdumper, and import the links tables as separate SQL dumps.

importDump.php is a command line script located in the `maintenance` folder of your MediaWiki installation. If you have shell access, you can call importdump.php like this:

```
php importDump.php <dumpfile>
```

where `<dumpfile>` is the name of the XML dump file. If the file is compressed and that has a `.gz` or `.bz2` file extension, it is decompressed automatically.

💡 **Note:** to run importDump.php (or any other tool from the maintenance directory), you need to set up your `AdminSettings.php` file.

💡 **Note:** running importDump.php can take quite a long time. For a large Wikipedia dump with millions of pages, it may take *days*, even on a fast server. Also note that the information in meta:Help:Import about merging histories, etc. also applies.

After running importDump.php, you may want to run rebuildrecentchanges.php in order to update the content of your Special:Recentchanges page.

FAQ

How do we setup debug mode?
In the **importDump.php** file there is a variable **$debug**. Set that variable to **true** to display debug messages. Set that variable to **false** to turn off debug messages.

How do we do a dryrun (no data added to the database)?
In the **importDump.php** file there is a variable **$dryrun**. Set that variable to **true** to run a dryrun. Set that variable to **false** to add the data to the database.

Which file controls the main functionality of importing the data?
Import.php in the **includes** directory. The main class is **WikiImporter**.

*Does SpecialImport.php (Importer from the special pages) use **WikiImporter** class?*
Yes, **importDump.php** and **SpecialImport.php** both use **WikiImporter**.

Using mwdumper

mwdumper is a Java application that can be used to read, write and convert MediaWiki XML dumps. It can be used to generate a SQL dump from the XML file (for later use with `mysql` or `phpmyadmin`) as well as for importing into the database directly. It is a lot faster than importDump.php, however, it only imports the revisions (page contents), and does not update the internal link tables accordingly -- that means that category pages and many special pages will show incomplete or incorrect information unless you update those tables.

If available, you can fill the link tables by importing separate SQL dumps of these tables using the `mysql` command line client directly. For Wikimedia wikis, this data is available along with the XML dumps.

Otherwise, you can run `rebuildall.php`, which will take a *long* time, because it has to parse all pages. This is not recommended for large data sets.

Using xml2sql

Xml2sql (a multiplatform ANSI C program) converts a MediaWiki XML file into an SQL dump for use with `mysql` or `phpmyadmin`. Just like using mwdumper (see above), importing this way is fast, but does not update secondary data like link tables, so you need to run `rebuildall.php`, which nullifies that advantage.

⚠ **Warning**: xml2sql is **not an official tool** and not maintained by MediaWiki developers. It may become outdated and **incompatible** with the latest version of MediaWiki!

It already did. It's reporting "unexpected element <redirect>" just as the original 1.15 Import.php does.

What to import?

- /Importing from mediawiki.org

See also

- meta:Data dumps#Tools
- Manual:System administration#Getting data

Language:	
	English • Русский