

Tutorial: Technical documentation

Statistics Explained

This documentation has been established for organisations interested to install a similar system for similar purposes.

Introduction

Statistics Explained website is based on Mediawiki software.

Statistics Explained is Eurostat's new way of publishing statistics on the internet, making full use of all linking and layering possibilities.

Its main purpose is to explain European statistics, by presenting data and pointing out what is interesting or surprising about them, with all the background needed for understanding them. The data discussed are recent, but not necessarily the very latest available. Statistics Explained offers, as a second high-value feature, deep and specific links to the most recent figures on Eurostat's website, as well as to metadata, additional information about the data such as definitions, methodological explanations, legal texts, etc. In this way, it can also serve as a portal to European data on any topic, even for specialists.

Statistics Explained is made up mainly of Statistical articles explaining all themes and subthemes, from many different points of view, in a language which is readily understandable by all. Road transport, for instance, is looked at and analyzed in articles on infrastructure, environmental impact, energy consumption, economic importance, congestion problems, traffic accidents, regional patterns and many more. Together they offer a comprehensive and tightly interconnected view on any given topic.

The software Mediawiki was customised in order to meet the needs of Eurostat, that's why some core files or extensions were modified.

The version of mediawiki we use is the 1.14.0.

In order to install the same system, there is several steps. The first one is the hardware configuration, then the software configuration of the webserver, and the last one the installation of the Mediawiki software.

Hardware configuration

Mediawiki can be hosted in a Virtual Machine, or in a dedicated webserver.

Statistics Explained website is hosted in both Test and Production environment in Virtual Machines.

Here is the minimal configuration for a webserver:

- 2 GB of RAM
- 10 GB of Hard disk
- Intel XEON @ 2.4GHz (at least)

Software configuration

Red Hat Enterprise Linux

The operating system of the webserver of Statistics Explained is **Red Hat Enterprise Linux 5.3** . Any other Linux distribution can be easily used (such as Debian, Mandriva, Ubuntu, etc...). This link will explain all the installation steps for each operating system:

http://www.mediawiki.org/wiki/Manual:Installation_guide

Apache

In the link below, you can find the way to install Apache in a Web Server.

http://www.mediawiki.org/wiki/Apache_configuration

The module **libphp5** must be enabled, this will allow to use PHP in the Web Server.

MySQL

MediaWiki stores all the text and data (content pages, user details, system messages, etc.) in a database. 3 database engines can be chosen:

MySQL 4.0 or later

SQLite3

PostgreSQL 8.1

For reasons of simplicity MySQL is chosen.

The MySQL database engine is the most commonly-used database backend for MediaWiki. Since it is the relational database management system used by the Wikimedia Foundation wiki farm in its own websites, it is well-supported in MediaWiki.

PHP

PHP is the programming language in which MediaWiki is written, and is required in order to run the software.

PHP version 5.2 is recommended to run the last stable version of Mediawiki 1.16.0 for a better use .

The value of `memory_limit` in the `php.ini` file should be changed to 512, in order to make piwik working correctly.

Compile time options

Perl Compatible Regular Expressions, Session and Standard PHP Library are required by Mediawiki, and at least one database driver must be enabled (MySQL, PostgreSQL or SQLite (through PDO)).

Useful extensions

The installation of the following extensions would be an asset:

libxml

ldap Support

oci8

eAccelerator

MySQL

Third-party tools

Sendmail

MediaWiki by default uses sendmail to send email notifications to users if they are enabled in the wiki and in the users preferences. Sendmail comes installed by default on Unix and Linux systems.

Once sendmail is installed, you must modify your `php.ini` file to point PHP at the correct sendmail executable.

ImageMagick / GD

ImageMagick or GD are both PHP libraries which can manipulate images, to resize them to create thumbnails

for example. The best of this two libraries is ImageMagick, because it produces better quality thumbnails.

MediaWiki can be configured to use ImageMagick to do dynamic resizing and thumbnailing of images.

Once ImageMagick is installed, you must enable ImageMagick and point MediaWiki to the convert program on the computer in the configuration file of Mediawiki (LocalSettings.php) like this:

```
wgUseImageMagick = true;  
wgImageMagickConvertCommand = '/usr/bin/convert'; for linux
```

TeX

MediaWiki uses a subset of TeX markup, including some extensions from LaTeX and AMSLaTeX, for mathematical formulae. It generates either PNG images or simple HTML markup, depending on user preferences and the complexity of the expression. In the future, as more browsers are smarter, it will be able to generate enhanced HTML or even MathML in many cases.

Here is the installation guide (CS 205 WP1 Task 5B) of TeX for Mediawiki.

Mediawiki installation

Custom functionalities

Sending a mail after different actions:

When an action is performed for an article (not editing, but sighting, validating, deleting or moving), a mail is sent to the watchlist of the article. Here is the tutorial to perform this task.

Forbidding the access for draft articles to non-logged users and also for specific pages

The main problem for mediawiki is that draft pages can be displayed when a user is not logged in. A simple function which forbids the access to draft pages is written within the extension FlaggedRevs, and when a non-logged user tries to display a draft, he is redirected to an error page.

The same behaviour exists for specific pages. A wiki article is created within a namespace, and contains pages which will not be displayed to public users. A new extension is created to meet this need.

Tutorial to perform this task can be found [here](#).

Logging the failed authentication attempts

After 5 wrong login attempts (configurable value), user account is blocked. Failed attempts can be logged by applying this tutorial.

Correcting a bug for the sortkeys

When adding a category link with the FCKEditor or with the basic text editor, if a sortkey is inserted within the category link, it will be ignored when saving the article. The only possible fix is to create an extension that adds a sortkey before saving the article.

Tutorial to perform this task can be found [here](#).

Removing temporary PDF files

The extension Collection creates temporary PDF files on the web server. When a user does not download it, the pdf is not deleted from the server and file system is not freed.

A cleanup script can be written to delete files that were not downloaded. Tutorial to perform this task can be found [here](#).

Displaying the stable version instead of the draft page

By default, Mediawiki displays the draft version instead of the stable version of an article. This can be a little bit disappointing for logged in users. A small patch will change this behavior. You can find [here](#) the tutorial to perform this task.

Resolving a problem with the user rights management

This is a bug reported by Eurostat. When searching a username in the User Rights Management screen, some users were not recognized in the system. The reason is that the username is transformed, the first letter is a capital letter, and the rest of the word lost the formatting.

Tutorial to perform this task can be found [here](#).

Fixing a problem with AddThis

This is a bug reported by Eurostat. When hovering over the "share" link in Statistics Explained, the window freezes in Internet Explorer, and then continues to load. Desactivating the onmouseover event solved the problem.

Tutorial to perform this task can be found [here](#).

Skin

Modification of the Skin (TASK 3C)

Adaptation of the Skin (Task 1)

Templates

The different templates used in Statistics Explained are described in the links below. Other used templates are described mainly in the FCKEditor customization.

Featured Article (TASK 2A1)

New Articles (TASK 2A2)

Extensions

Task7-extensions

This page will contain all the installed extensions.

- Bullet Feed
- Category Tree
- Cite
- CloseWikis
- Collection
- Contact
- EditArticle
- EditUser
- EditWarning
- EurostatTranslate
- FCKEditor
- FlaggedRevisions
- ForbidAccess
- LabeledSectionTransclusion
- LDAP Authentication Plugin
- Lucene-Search
- MagpieRSS
- MultiBoilerPlate
- MWSearch
- ParserFunctions
- PDFBook
- RSS Feed
- send2friend
- Random Area
- User merge and delete
- WikiArticleFeeds
- WhoiswatchingExternal applications

Piwik