



WIKIMEDIA

HACKATHON

Yobot, AutoWikiBrowser, CHECKWIKI

Marios Magioladitis



WIKIMEDIA
Community User Group
GREECE

21.05.2022

AutoWikiBrowser

The semi-automated wiki editor

Version 5.9.0.1



Yobot, **AutoWikiBrowser**, CHECKWIKI

Marios Magioladitis



05.11.2017

Wikipedia:

enwiki:

5,500,000 articles



- **Big Data:** Manual data processing is inadequate to deal with that amount of data
- Great need to apply same rules in all pages
- Many editors, not all familiar with Mediawiki and wikisyntax
- Vandalism
- Typos
- Mass re-categorisations etc.

AutoWikiBrowser



AutoWikiBrowser – Default.xml – Empire Professional Baseball League

File View List Plugins Options Tools Help

Double click on a line to undo all changes on that line, or single click to focus the edit box to that line.

Current revision Your text

Line 1:

```

+ {{Multiple issues}}
+ {{Orphan|date={{subst:CURRENTMONTHNAME}} {{subst:CURRENTYEAR}}}}
+ {{hoax|date=December 2015}}
+ }}
+
+ {{Infobox sports league
|title= Empire League

```

Make list

Source: New pages

1423 pages Make list

Avonleigh, Rockhampton

- Automotive industry in South Africa
- 1928 Saint Mary's Gaels football tea
- A New Place z Drown
- Friedrich Loods
- 1933 Auburn Tigers football team
- Sailors Run
- Sir James Whyllie, 1st Baronet
- Reasoners Run
- 1929 Auburn Tigers football team
- 1930 Auburn Tigers football team
- Kaimi Fairbairn
- 1927 Saint Mary's Gaels football tea
- Dyes Fork
- 1 Peter 5
- John Bremner (disambiguation)
- Kress Dunsburn-Said
- Empire Professional Baseball League**

Remove Filter

Options More... Disambig Skip Start

Default Summary clean up

Lock Minor edit Start

Page statistics

Words: 124 Stop

Links: 18 Preview Off

Images: 1

Categories: 1 Watch Move

Intervis links: 0

Dates O/I/A: 0/0/0 Protect Delete

Alerts

Skip

Save

Find

Regex Find

Case sensitive

Edit box History What links here Logs Page Logs

clean up, added [[CAT:Orphan]] tag

```

{{Multiple issues}}
{{Orphan|date={{subst:CURRENTMONTHNAME}} {{subst:CURRENTYEAR}}}}
{{hoax|date=December 2015}}
}}
+
+ {{Infobox sports league
|title= Empire League
|logo = File:EMPIRELEAGUELOGOdghj.png
|caption =
|sport = [[Baseball]]
|founded = 2016
|class =
|folded =
|teams = 4
|country = USA
|website = [http://www.empireproleague.com Official]
}}
The Empire League is a planned [[independent baseball league]] scheduled to start in 2016, replacing

```

Pages/min: 1 Edits/min: 0 Edits: 1 Skipped: 61 New: 0 en.wikipedia CAPTAIN ROU 0

AutoWikiBrowser



- Created in 2006
- Open-source
- Semi-automated MediaWiki editor for **Windows**
- > 100 code syntax fixes with the use of regular expressions
- Fully compatible with all Wikipedia and their sister projects
- Allows plugins
- Written in C#

AWB is powerful



- Page edits since English Wikipedia was set up (2001): 852,605,210 (Special:Statistics)
- AWB edits in en.wiki since 2009: 108,633,087 (<https://tools.wmflabs.org/awb/stats/>)
- AWB edits in en.wiki 2008-2009: 4,904,034 (<https://tools.wmflabs.org/awb/stats/>)
- Vi.wiki edits with AWB: 193,854,162
- Lietuval.it edits with AWB: 875,640,318

The code at a glance



- > 3 millions lines of code [including libraries]
- 633K comments
- 1.6Mb code in C#
- 957 years of effort (COCOMO model)
- Many plugins
- Edit box supports the Microsoft Text Services Framework for use with speech recognition/handwriting applications
- Licence: GPL-2.0+
- <https://www.openhub.net/p/AutoWikiBrowser>



Code is available at Sourceforge

AutoWikiBrowser

Brought to you by: [magioladitis](#), [maxsem](#), [reedy_boy](#), [rjwilmsi](#)

[Summary](#) [Files](#) [Reviews](#) [Support](#) [Wiki](#) [Code](#)

★ 4.0 Stars (4)

↓ 80 Downloads (This Week)

📅 Last Update: 2015-10-01

🐦 Tweet 5

➕ 3

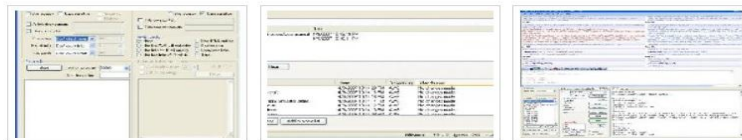
👍 Like 20



Download

AutoWikiBrowserS600.zip

[Browse All Files](#)



Description

AutoWikiBrowser is a semi-automated Wikipedia editor, designed to make tedious, repetitive tasks quicker and easier. For more information, see the project homepage at <http://en.wikipedia.org/wiki/Wikipedia:AutoWikiBrowser>

[AutoWikiBrowser Web Site](#)

Categories

Browsers

License

GNU General Public License version 2.0 (GPLv2)



Code is available at Sourceforge

sourceforge Search Browse Enterprise Blog Jobs Deals Help Log In or Join

SOLUTION CENTERS Go Parallel Resources Newsletters

AutoWikiBrowser

Brought to you by: magioladitis, maxsem, reedy_boy, rjwilmsi

Summary Files Reviews Support Wiki Code

<https://sourceforge.net/projects/autowikibrowser/>

★ 4 Stars (4)
Downloaded 101
Tweet 5 Like +20



Description

AutoWikiBrowser is a semi-automated Wikipedia editor, designed to make tedious, repetitive tasks quicker and easier. For more information, see the project homepage at <http://en.wikipedia.org/wiki/Wikipedia:AutoWikiBrowser>

[AutoWikiBrowser Web Site](#)

Categories
Browsers

License
GNU General Public License version 2.0 (GPLv2)



The code

```
objects
  TemplateParameters.cs
  Templating.cs
  NoLimitsPlugin
  TheTemplator
  TypoScan
  UnitTests
  WikiFunctions
    References
    API
    Article
    Background
    Controls
    CustomModules
    DBScanner
    Diff
    Disambiguation
    Lists
    Logging
    Parse
      BoldTitle.cs
      CategoryDefaultSc
      CiteTemplates.cs
      FindandReplace.cs
      FixDates.cs
      FixHeadings.cs
      FixSyntax.cs
      HideText.cs
      MetaDataSorter.cs
      MultipleIssuesTem
      Parsers.cs
      PersonData.cs
      References.cs
  WikiFunctions.Parse.Parsers
    335
    336
    337 // needs to be applied after SquareBracketsInExternalLinks
    338 if (orphanedSingleBrackets && !SyntaxRegexFileWithHTTP.IsMatch(articleText))
    339 {
    340     articleText = SyntaxRegexWikilinkMissingClosingBracket.Replace(articleText, "[[{$1}]");
    341     articleText = SyntaxRegexWikilinkMissingOpeningBracket.Replace(articleText, "[[{$1}]");
    342 }
    343
    344 // adds missing http:// to bare url references lacking it - CHECKWIKI error 62
    345 articleText = RefURLMissingHttp.Replace(articleText, @"{$1}http://www.");
    346
    347 // repair bad Image/external links, ssb check for performance
    348 if (ssb.Any(m => m.Contains(":") && m.ToLower().Contains(":http")))
    349     articleText = SyntaxRegexExternalLinkToImageURL.Replace(articleText, "{$1}");
    350
    351 // CHECKWIKI error 69
    352 bool isbnDash = articleText.Contains("ISBN-");
    353 if (isbnDash || articleText.Contains("ISBN:") || articleText.Contains("ISBN\t") || ssb.Contains("[[ISBN]]"))
    354     articleText = SyntaxRegexISBN.Replace(articleText, "ISBN $1");
    355
    356 if (isbnDash)
    357     articleText = SyntaxRegexISBN2.Replace(articleText, "ISBN ");
    358
    359 if (ssb.Contains("[[ISBN]]"))
    360     articleText = SyntaxRegexISBN3.Replace(articleText, "ISBN $1");
    361
    362 if (ssb.Contains("[[International Standard Book Number|ISBN]]"))
    363     articleText = SyntaxRegexISBN4.Replace(articleText, "ISBN $1");
    364
    365 if (articleText.Contains("PMID:"))
    366     articleText = SyntaxRegexPMID.Replace(articleText, "{$1} {$2}");
    367
    368 // Remove sup tags from ordinals per [[WP:ORDINAL]].
```



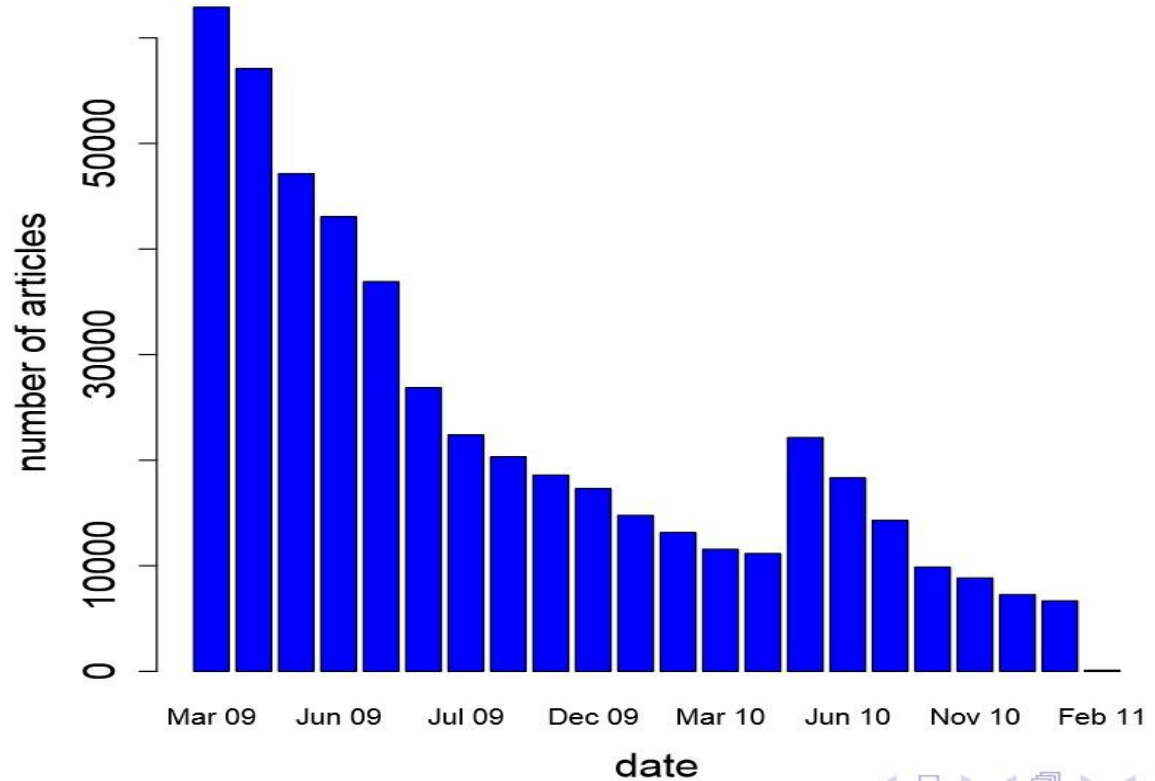
The code

```
WikiFunctions.Parse.Parsers | FixSyntax(string articleText)
250     bool badHttpLinks = Tools.DeduplicateList((from Match m in HttpLinks.Matches(articleText.ToLower()) select m.Value).ToList()).Any(s =>
251
252     if (badHttpLinks)
253         articleText = MultipleHttpInLink.Replace(articleText, "$1");
254
255     articleText = MultipleFtpInLink.Replace(articleText, "$1");
256
257     if (badHttpLinks && TemplateExists(alltemplates, WikiRegexes.UrlTemplate))
258         articleText = WikiRegexes.UrlTemplate.Replace(articleText, m => m.Value.Replace("http://http://", "http://"));
259
260     if (badHttpLinks && !SyntaxRegexHTTPNumber.IsMatch(articleText))
261     {
262         articleText = MissingColonInHttpLink.Replace(articleText, "$1://$2");
263         articleText = SingleTripleSlashInHttpLink.Replace(articleText, "$1://$2");
264         articleText = articleText.Replace("https://http://", "https://");
265     }
266
267     if (CellpaddingTypoQuick.IsMatch(articleText))
268         articleText = CellpaddingTypo.Replace(articleText, "$1cellpadding");
269
270     if (SimpleTagsList.Any(s => s.Contains("font")))
271         articleText = RemoveNoPropertyFontTags.Replace(articleText, "$1");
272
273     // ...

```

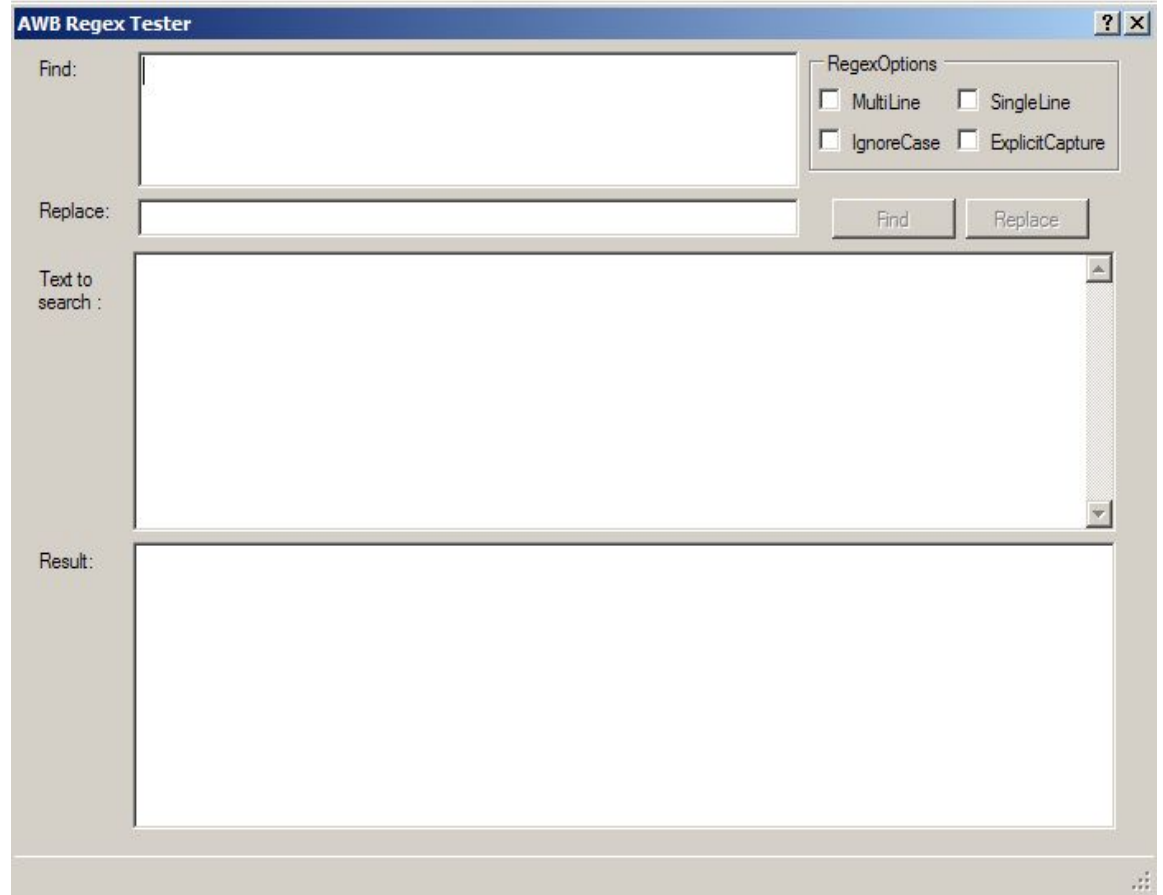


An example: BLP tags





Advanced tools





Advanced tools

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Text.RegularExpressions;
using WikiFunctions;

namespace AutoWikiBrowser.CustomModules
{
    class CustomModule : WikiFunctions.Plugin.IModule
    {
        WikiFunctions.Plugin.IAutoWikiBrowser awb;

        public CustomModule(WikiFunctions.Plugin.IAutoWikiBrowser mAWB)
        {
            awb = mAWB;
        }

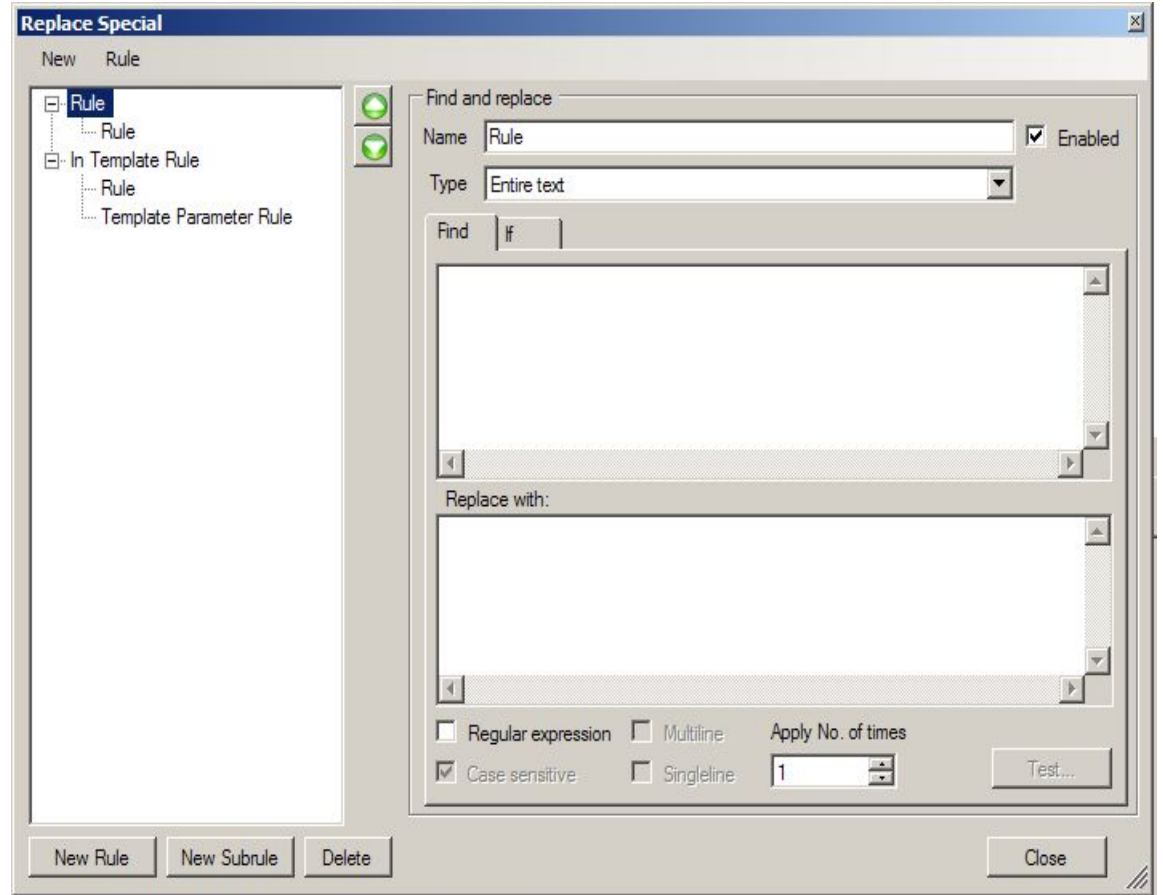
        public string ProcessArticle(string ArticleText, string ArticleTitle, int wikiNa
        {
            Skip = false;
            Summary = "test";

            ArticleText = "test \r\n\r\n" + ArticleText;

            return ArticleText;
        }
    }
}
```



Advanced tools



CHECKWIKI

en.wikipedia.org/wiki/Wikipedia:WikiProject_Check_Wikipedia

Magioladitis 10:25 Talk Sandbox

Project page **Talk** Read Edit View history More TW Search

Wikipedia:WikiProject Check Wikipedia

From Wikipedia, the free encyclopedia

Check Wikipedia **Toolforge** **List of Errors**

Checkwiki helps clean up syntax and other errors in the source code of Wikipedia by:

- Eliminating errors in the [wiki syntax](#), such as missing close tags or brackets;
- Checking for [accessibility](#) issues, such as small print or heading problems;
- Correcting, deleting or moving code that does not follow conventions, such as the [position of references](#) with respect to punctuation;
- Finding articles that have [invalid ISBNs](#) ([list available here](#))

Operation [edit]

- The program scans every new [dump](#) for over 30 languages and creates a list of articles with errors.
- For a few Wikipedias, the program scans newly revised articles daily, to create a new list for users, omitting already-corrected articles.
- A specific interface - [here](#) for the English project - facilitates correcting problems detected by the program in the following three levels of ranking priorities errors. It helps keep the error lists updated in near real time by providing some statistics on corrections made recently.
- You can download the Checkwiki programs at [GitHub](#). The programs are licensed under [GPLv3](#).
- Interested users might also like to review the regularly-updated lists of wiki syntax issues at [Special:LintErrors](#).
- An [analysis](#) is also performed by [WikiCleanerBot](#) twice a month on the dumps for some of the errors, including some additional errors compared to the project interface.

Tools [edit]

https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Check_Wikipedia

WIKIPROJECT REPORT

Checking that everything's all right

By Resprinter123

[Contribute](#) — [Share this](#) [show]

This week, the *Signpost* decided to have a look around with **WikiProject Check Wikipedia**, a maintenance project not concerned so much with articles' content, but in all the tiny errors that are to be found scattered within them. Their front page gives a list of things they mainly focus on:

- “ Checkwiki helps clean up syntax and other errors in the source code of the Wikipedia by finding problems such as:
- Eliminate errors in [wiki syntax](#), such as a missing close tags or brackets;
 - Check for [accessibility](#) issues, such as small print or heading problems;
 - Correct, delete or move code that does not follow conventions, such as the [position of references](#) with respect to punctuation;
 - Find articles that have [invalid ISBNs](#) (list available [here](#));
- ”

CHECKWIKI helps clean up syntax and other errors in the source code of Wikipedia

CHECKWIKI

Check Wikipedia

→ Homepage

More information at the [projectpage](#).

Choose your project!

This table will update every 15 minutes.

Project [↑] _↓	To-do [↑] _↓	Done [↑] _↓	Last dump [↑] _↓	Last update [↑] _↓	Page at Wikipedia	Translation
afwiki	2173	0	2022-05-20	2022-05-20	here	here
alswiki	230	0	2022-05-20	2022-05-20	here	here
arwiki	33020	0	2022-05-20	2022-05-21	here	here
arywiki	4287	0	2022-05-20	2022-05-20	here	here
arwiki	379944	0	2022-05-20	2022-05-20	here	here
barwiki	2481	0	2022-05-20	2022-05-20	here	here
bewiki	191686	0	2022-05-20	2022-05-20	here	here
bnwiki	13332	0	2022-05-20	2022-05-21	here	here
bnwikisource	101	0	2022-05-20	2022-05-20	here	here
bnwikivoyage	14	0	2022-05-20	2022-05-20	here	here
bnwiktionary	15	0	2022-05-20	2022-05-20	here	here
bswiki	46243	0	2022-05-20	2022-05-21	here	here
cawiki	10613	0	2022-05-20	2022-05-21	here	here
cswiki	66162	1	2022-05-20	2022-05-21	here	here
cywiki	78027	0	2022-05-20	2022-05-21	here	here
dawiki	6378	0	2022-05-20	2022-05-21	here	here
dewiki	5300	108	2018-11-01	2022-05-21	here	here
dewikisource	25803	0	2022-05-20	2022-05-21	here	here
elwiki	35100	0	2022-05-20	2022-05-21	here	here
enwiki	98162	50	2017-01-01	2022-05-21	here	here
enwiktionary	113987	0	2022-05-20	2022-05-21	here	here
eowiki	71397	0	2022-05-20	2022-05-21	here	here
eswiki	31723	50	2022-05-01	2022-05-21	here	here
fawiki	30749	0	2022-05-20	2022-05-21	here	here
fiwiki	47118	0	2022-05-20	2022-05-21	here	here
frwiki	9039	0	2022-05-20	2022-05-21	here	here
frwikiversity	5569	930	2022-05-20	2022-05-21	here	here
frwikivoyage	988	398	2022-05-20	2022-05-21	here	here
frwiktionary	56079	326	2022-05-20	2022-05-21	here	here
fywiki	12446	0	2022-05-20	2022-05-21	here	here
gdwiki	3994	0	2022-05-20	2022-05-21	here	here
hewiki	27881	118	2022-05-20	2022-05-21	here	here
hrwiki	42811	0	2022-05-20	2022-05-21	here	here
huwiki	13634	0	2022-05-20	2022-05-21	here	here
idwiki	78964	0	2022-05-20	2022-05-21	here	here

<https://checkwiki.toolforge.org/cgi-bin/checkwiki.cgi>



The team





The team



Phabricator



- It is a suite of web-based software development collaboration tools, including the Differential code review tool, the Diffusion repository browser, the Herald change monitoring tool, the Maniphest bug tracker and the Phriction wiki.
- Phabricator integrates with Git, Mercurial, and Subversion. It is available as free software under the Apache License, version 2.
- Phabricator was originally developed as an internal tool at Facebook. Phabricator's principal developer is Evan Priestley. Priestley left Facebook to continue Phabricator's development in a new company called Phacility.

Report bugs and ideas!

https://phabricator.wikimedia.org/project/view/1012/

PHABRICATOR

AutoWikiBrowser Public

Sort: Natural Filter: Open Tasks Manage Board

T111663 Tag AWB edits	T112219 Allow CTRL + C, V and X to work in the More tab	T99277 [Migrated] NullReferenceException in MainForm.SetProject	T109522 Have AWB fix some table ending issues	T99420 [Migrated] Remove stub templates from redirect pages
T111310 Ability for crosswiki login (propagating session), rather than to login to each and every wiki	T112202 AutoWikiBrowser exception error	T108817 The result returned by server was blank	T112078 Replace För with {{För}}-templates on sv.wp	T99714 [Migrated] Change {{(Expansion section)}} with {{(Empty section)}} in section
T99314 [Migrated] Sometimes upon clicking save AWB restarts same article without hand edited changes	T99715 [Migrated] Remove {{Stub}} when a more fine-grained stub tag already exists on the page	T107644 ArgumentException in MainForm.MainForm_FormClosing	T100695 [Migrated] Localization in Turkish	
T99303 [Migrated] AWB cannot save letter + combining diacritic when a precomposed Unicode glyph is available	T111660 Enable general fixes in Draft namespace	T106817 AccessViolationException in UnsafeNativeMethods.CallWindowProc	T99327 [Migrated] Hyphen rule applied inconsistently	
T37654 API edit of translate page gives Unknown error: "tpt-target-page"	T111693 Convert selected external links to HTTPS	T102218 ArgumentException in ApiEdit.Save	T99307 [Migrated] Typo fixing should be before SimplifyLinks	
T100288 [Migrated] Can't save default prefs	T109254 Fix double equals signs in template parameters	T100286 [Migrated] ThreadStateException in AsyncApiEdit.Abort (Main form Stop button, AsyncApiEdit.Abort)	T99290 [Migrated] AWB does not merge group references	
T41492 prop=info doesn't behave consistently when given a Special Page, vs when given a non special page which redirects to a special page (in comparison to another redirect)	T105498 Remove empty "See also"-section	T92352 ConfigurationErrorsException - Configuration system failed to initialize - Root element is missing	T105806 Swedish Wikipedia (sv.wp) specific header changes	
T91080 Document/annotate ListProviders whether they return BIG or SMALL query results	T99317 [Migrated] ExplicitCapture	T101152 [Migrated] AWB Updater: ZipException in ZipFile.ReadEntries	T105348 Multiple line breaks in {{cite web}}	
T109757 During run, Stop after redirected file drops the file from subsequent run	T100583 [Migrated] Title case for citations	T101040 InvalidOperationException in UserPrefs.LoadPrefs	T100608 [Migrated] Better support for links to web.archive.org in cite templates	
	T99310 [Migrated] Unable to undo changes via diff double click	T99334 [Migrated] AWB does not start on Windows XP with .NET Framework 2.0 or 3.0 (Custom Module error)	T100582 [Migrated] Do not include unimportant prefixes in {{DEFAULTSORT}}	
	T99273 [Migrated] External processing error		T100581 [Migrated] When maintenance templates immediately follow section header, change to section template	
	T99275 [Migrated] Wikia: This page is			

Acknowledgments



- The speaker is member of Wikimedia Community User Group Greece
- Slides are available on Wikimedia Commons