



FOSSCOMM 2017

Mediawiki and AutoWikiBrowser

Marios Magioladitis

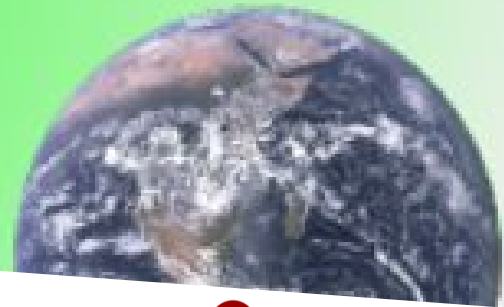


05.11.2017

AutoWikiBrowser

The semi-automated wiki editor

Version 5.9.0.1



Mediawiki and AutoWikiBrowser

Marios Magioladitis



WIKIMEDIA
Community User Group
GREECE

05.11.2017

Wikipedia:



- Created in 2001 by Larry Sanger and Jimmy Wales
- Owned by non-profit Wikimedia Foundation
- Anyone can edit
- 40 million articles in 299 languages
- 25k-60k pages requests per second
- It is accessed via website, mobile apps, search engines, etc.

Mediawiki



- Free and open-source wiki software.
- Originally developed by Magnus Manske and improved by Lee Daniel Crocker.
- Initial release: 2002
- It runs on many websites, including Wikipedia, Wiktionary and Wikimedia Commons.
- It is written in the PHP programming language and stores the contents into a database.
- The software is optimized to efficiently handle large projects, which can have terabytes of content and hundreds of thousands of hits per second.

Mediawiki



MediaWiki has had 79.849 commits made by 776 contributors, representing 540.379 lines of code

```
958 /**
959  * Make URL indexes, appropriate for the e1_index field of externallinks.
960  *
961  * @param string $url
962  * @return array
963  */
964 function wfMakeUrlIndexes( $url ) {
965     $bits = wfParseUrl( $url );
966
967     // Reverse the labels in the hostname, convert to lower case
968     // For emails reverse domainpart only
969     if ( $bits['scheme'] == 'mailto' ) {
970         $mailparts = explode( '@', $bits['host'], 2 );
971         if ( count( $mailparts ) === 2 ) {
972             $domainpart = strtolower( implode( '.', array_reverse( explode( '.', $mailparts[1] ) ) ) );
973         } else {
974             // No domain specified, don't mangle it
975             $domainpart = '';
976         }
977         $reversedHost = $domainpart . '@' . $mailparts[0];
978     } else {
979         $reversedHost = strtolower( implode( '.', array_reverse( explode( '.', $bits['host'] ) ) ) );
980     }
981     // Add an extra dot to the end
982     // why? Is it in wrong place in mailto links?
983     if ( substr( $reversedHost, -1, 1 ) !== '.' ) {
984         $reversedHost .= '.';
985     }
986     // Reconstruct the pseudo-URL
987     $prot = $bits['scheme'];
988     $index = $prot . $bits['delimiter'] . $reversedHost;
```

Mediawiki



- <https://www.mediawiki.org>
- Extensions and themes

Mediawiki



- Achieving scalability through multiple layers of caching and database replication has been a major concern for developers.
- On Wikipedia more than 1000 automated and semi-automated bots and other tools have been developed to assist in editing.

Mediawiki



Editing Pride and Prejudice

B I [Advanced](#) [Special characters](#) [Help](#)

`''''''Pride and Prejudice''''''` is a romance novel by `[[Jane Austen]]`, first published in 1813. The story charts the emotional development of the protagonist, `[[Elizabeth Bennet]]`, who learns the error of making hasty judgements and comes to appreciate the difference between the superficial and the essential. The comedy of the writing lies in the depiction of `[[manners]]`, `[[education]]`, and `[[marriage]]` and `[[money]]` in the `[[Regency era|British Regency]]`.

Mr Bennet of the Longbourn estate has five daughters, but his property is entailed, meaning that none of the girls can inherit it. Since his wife had no fortune, it is imperative that one of the girls marries well in order to support the others on his death. However, Jane Austen's opening line, "It is a truth universally acknowledged that a single man in possession of a good fortune must be in want of a wife" is a sentence filled with irony and playfulness. The novel revolves around the necessity of marrying for love, not simply for monetary reasons, despite the social pressures to make a good (i.e. wealthy) match.

`''Pride and Prejudice''` retains the fascination of modern readers, consistently appearing near the top of lists of "most-loved books" among both `[[Literary critics|literary scholars]]` and the general public. It has become one of the most popular novels in `[[English literature]]`, with over 20 million copies sold, and paved the way`{{specify|date=December 2016}}` for many archetypes that abound in modern literature. Continuing interest in the book has resulted in a number of dramatic adaptations and an abundance of novels and stories imitating Austen's memorable characters or themes.`<ref>{{cite web|url=http://www.monstersandcritics.com/dvd/reviews/article_1475660.php/Pride_and_Prejudice_%E2%80%93_Blu-ray_Review|title=Monstersandcritics.com|date=7 May 2009|publisher=Monstersandcritics.com|archive-url=https://web.archive.org/web/20091026154330/http://www.monstersandcritics.com:80/dvd/reviews/article_1475660.php/Pride_and_Prejudice_%E2%80%93_Blu-ray_Review|archive-date=26 October 2009|dead-url=yes|accessdate=27 January 2012}}</ref>` The 2005 film, `''[[Pride & Prejudice (2005 film)|Pride and Prejudice]]''`, starring `[[Keira Knightley]]` and `[[Matthew Macfadyen]]` is the most recent Hollywood adaption of

Mediawiki



MediaWiki

visual editor



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikimedia Shop

- Interaction
 - Help
 - About Wikipedia
 - Community portal
 - Recent changes
 - Contact Wikipedia

- Toolbox
 - What links here
 - Related changes
 - Upload file
 - Special pages
 - Permanent link
 - Page information
 - Cite this page
 - Rate this page

- Print/export
 - Create a book
 - Download as PDF
 - Printable version

- Languages
 - Dansk
 - Deutsch
 - Eesti
 - Esperanto
 - Français
 - עברית
 - فارسی

Editing Wikipedia...

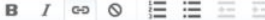
Jdforrester (WMF) Talk Preferences Watchlist Contributions Log out

Article Talk

Read Edit VisualEditor View history

Search

Paragraph



Leave feedback 2 notices Cancel Review and save

Islington

From Wikipedia, the free encyclopedia

For other uses, see Islington (disambiguation).

Islington (/ˈɪzlɪntɒn/) is a neighbourhood in Greater London, England, and forms the central district of the London Borough of Islington. It is a desirable residential district of Inner London, extending from Islington High Street to Highbury Fields, encompassing the area around the busy Upper Street. The area is bounded to the north by the Regent's Canal, to the east by the Regent's Canal, to the south by Upper Street such as St Mary's, St Peter's Church and the Regent's Canal, and to the west by the Regent's Canal.

Coordinates: 51°54′0″N 0°10′27″W﻿ / ﻿51.90000°N 0.17139°W﻿ / 51.90000; -0.17139

Modern definition

Islington grew as a separate district in 1965, when it was merged with the City of Islington to form the London Borough of Islington. This gave rise to the modern definition of the area. The area is bounded by Liverpool Road to the north, Upper Street to the east, Canonbury, The main road to the east, and the Regent's Canal to the west. The area around Angel, Islington, is a separate district in its own right: The Angel, Islington. The northern part of this area (from the Liverpool Road to the east) is the district of Islington, while the southern half is in neighbouring Finsbury. The area bounded by Upper Street to the east and the Regent's Canal to the west is the adjoining district of Pentonville.

History

Etymology

Islington was originally named by the Saxons *Giseldone* (1005), then *Gislandune* (1062). The name means "Gisla's hill" from the Old English personal name *Gisla* and *dun* ("hill", "down"). The name later mutated to *Isledon*, which remained in use well into the 17th century when the modern form arose.[1] In medieval times, Islington was just one of many small manors thereabouts, along with *Bernersbury*, *Neweton Berewe* or *Hey-bury* and *Canonesbury* (Barnsbury, Highbury and Canonbury – names first recorded in the 13th and 14th centuries).

Origins



Buildings on Islington High Street



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



- 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

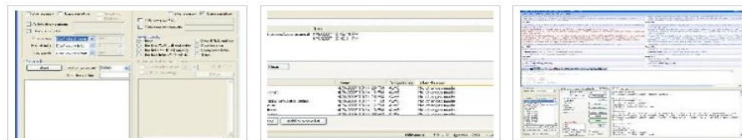


Download

AutoWikiBrowser5600.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

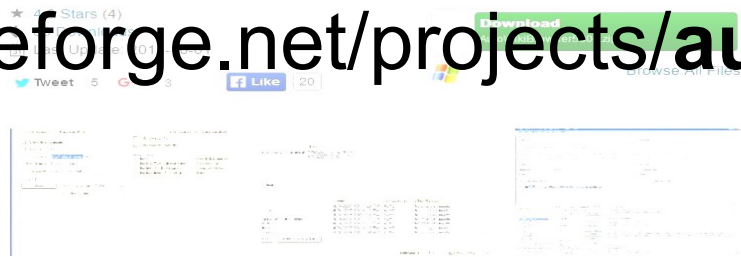


AutoWikiBrowser

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

Summary Files Reviews Support Wiki Code

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



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]].
```

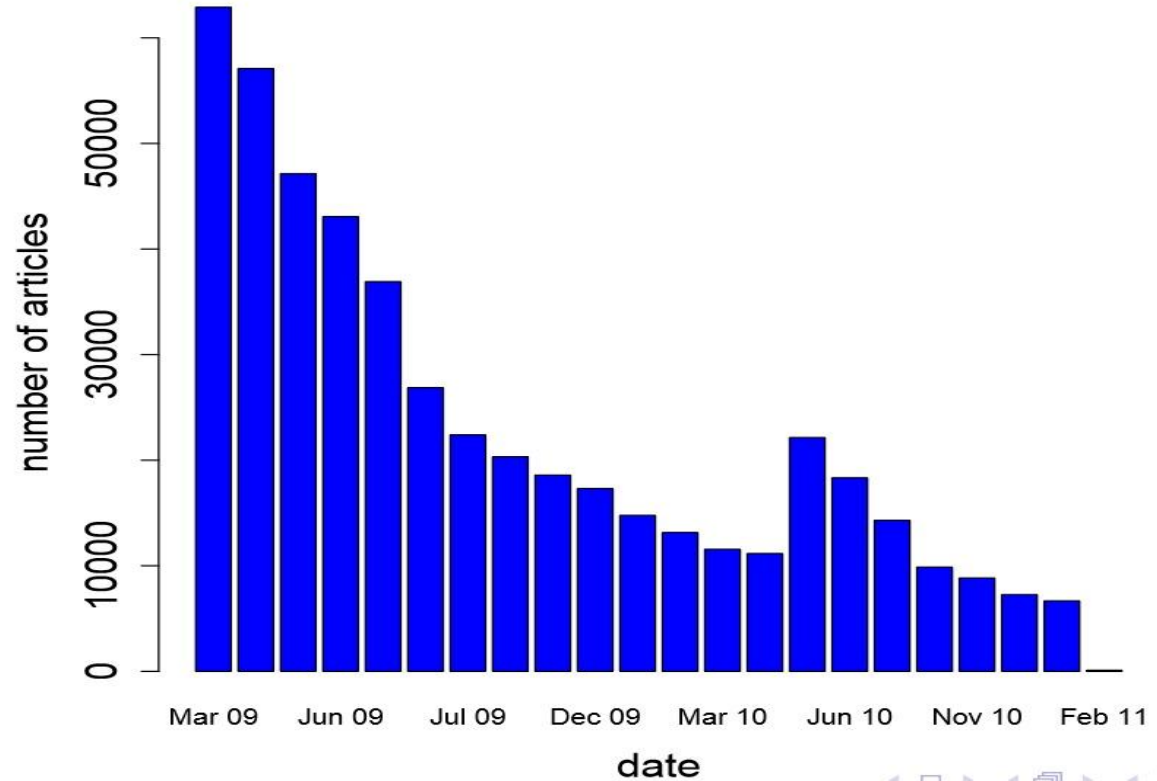


The code

```
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     // ...
```

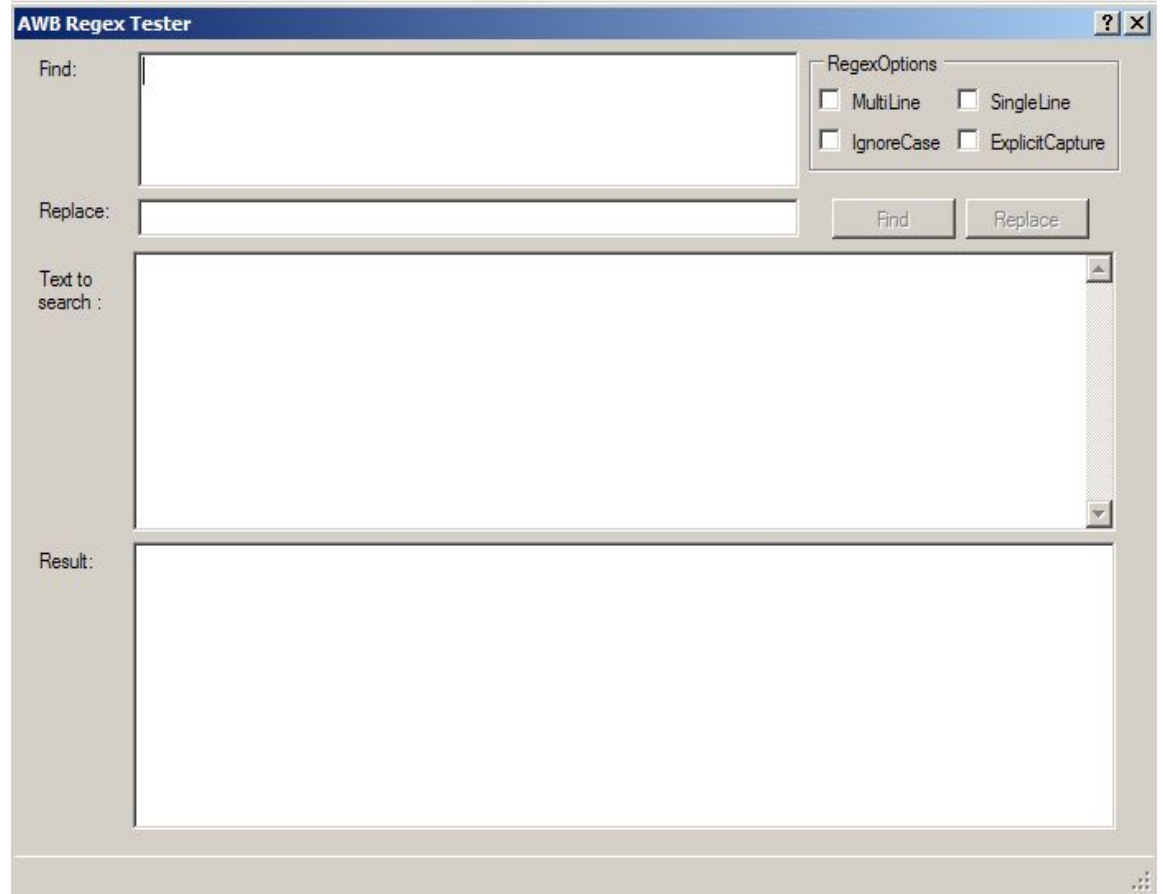


An example: BLP tags





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;
        }

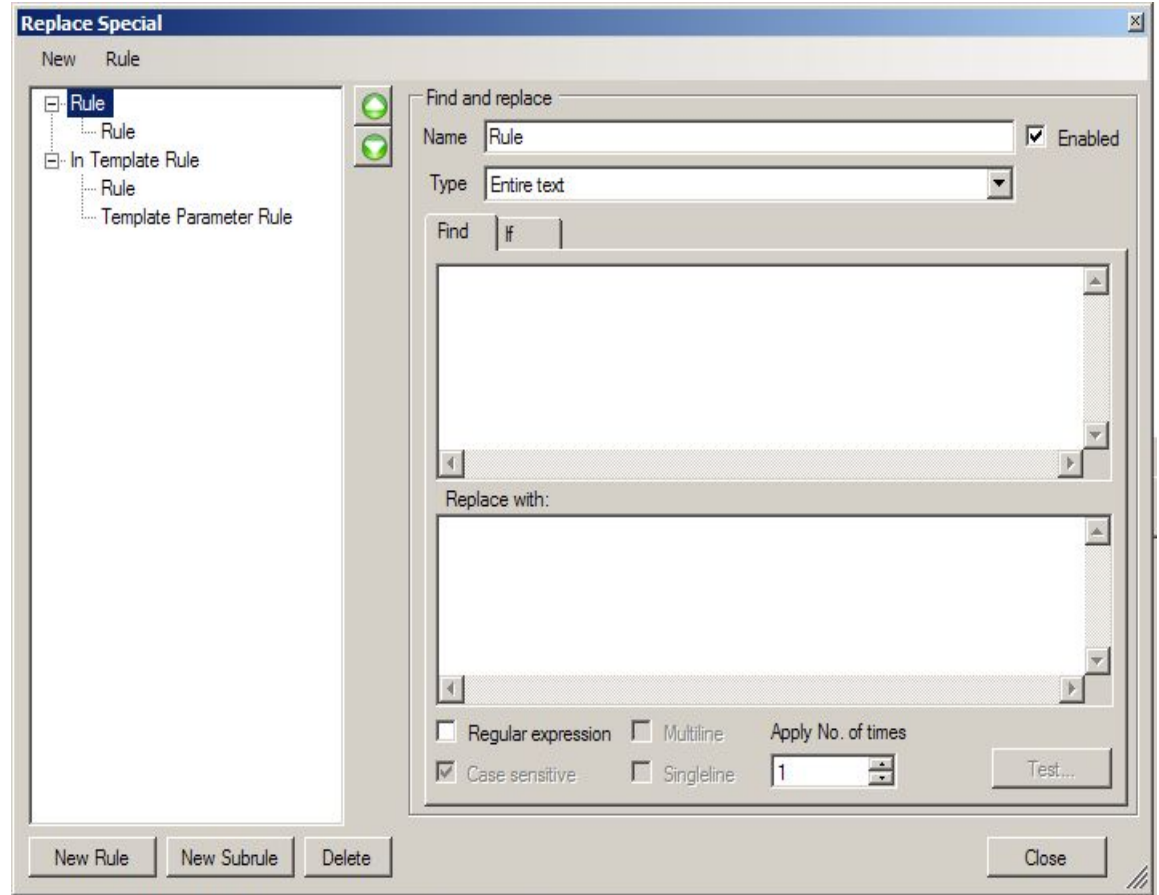
        public string ProcessArticle(string ArticleText, string ArticleTitle, int wikiNa
        {
            Skip = false;
            Summary = "test";

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

            return ArticleText;
        }
    }
}
```



Advanced tools





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 and received a WMF grant to travel and participate in this Conference
- Slides are available on Wikimedia Commons