
WikiTeam Documentation

Release 0.3

WikiTeam developers

July 30, 2016

1	Indices and tables	3
	Python Module Index	5

Contents:

`dumpgenerator.avoidWikimediaProjects (config={}, other={})`
Skip Wikimedia projects and redirect to the dumps website

`dumpgenerator.bye ()`
Closing message

`dumpgenerator.checkAPI (api=None, session=None)`
Checking API availability

`dumpgenerator.checkIndex (index=None, cookies=None, session=None)`
Checking index.php availability

`dumpgenerator.checkXMLIntegrity (config={}, titles=[], session=None)`
Check XML dump integrity, to detect broken XML chunks

`dumpgenerator.cleanHTML (raw='')`
Extract only the real wiki content and remove rubbish

`dumpgenerator.cleanXML (xml='')`
Trim redundant info

`dumpgenerator.curateImageURL (config={}, url='')`
Returns an absolute URL for an image, adding the domain if missing

`dumpgenerator.delay (config={}, session=None)`
Add a delay if configured for that

`dumpgenerator.domain2prefix (config={}, session=None)`
Convert domain name to a valid prefix filename.

`dumpgenerator.fixBOM (request)`
Strip Unicode BOM

`dumpgenerator.generateImageDump (config={}, other={}, images=[], start='', session=None)`
Save files and descriptions using a file list

`dumpgenerator.generateXMLDump (config={}, titles=[], start=None, session=None)`
Generates a XML dump for a list of titles

`dumpgenerator.getImageNames (config={}, session=None)`
Get list of image names

`dumpgenerator.getImageNamesAPI (config={}, session=None)`
Retrieve file list: filename, url, uploader

`dumpgenerator.getImageNamesScraper (config={}, session=None)`
Retrieve file list: filename, url, uploader

`dumpgenerator.getJSON (request)`
Strip Unicode BOM

`dumpgenerator.getNamespacesAPI (config={}, session=None)`
Uses the API to get the list of namespaces names and ids

`dumpgenerator.getNamespacesScraper (config={}, session=None)`
Hackishly gets the list of namespaces names and ids from the dropdown in the HTML of Special:AllPages

`dumpgenerator.getPageTitles (config={}, session=None)`
Get list of page titles

`dumpgenerator.getPageTitlesAPI (config={}, session=None)`
Uses the API to get the list of page titles

`dumpgenerator.getPageTitlesScraper (config={}, session=None)`
Scrape the list of page titles from Special:Allpages

`dumpgenerator.getUserAgent ()`
Return a cool user-agent to hide Python user-agent

`dumpgenerator.getWikiEngine (url='')`
Returns the wiki engine of a URL, if known

`dumpgenerator.getXMLFileDesc (config={}, title='', session=None)`
Get XML for image description page

`dumpgenerator.getXMLHeader (config={}, session=None)`
Retrieve a random page to extract XML headers (namespace info, etc)

`dumpgenerator.getXMLPage (config={}, title='', verbose=True, session=None)`
Get the full history (or current only) of a page

`dumpgenerator.getXMLPageCore (headers={}, params={}, config={}, session=None)`

`dumpgenerator.loadConfig (config={}, configfilename='')`
Load config file

`dumpgenerator.logerror (config={}, text='')`
Log error in file

`dumpgenerator.main (params=[])`
Main function

`dumpgenerator.mwGetAPIAndIndex (url='')`
Returns the MediaWiki API and Index.php

`dumpgenerator.readTitles (config={}, start=None)`
Read title list from a file, from the title "start"

`dumpgenerator.removeIP (raw='')`
Remove IP from HTML comments <!-- -->

`dumpgenerator.reverse_readline (filename, buf_size=8192, truncate=False)`
a generator that returns the lines of a file in reverse order

`dumpgenerator.saveConfig (config={}, configfilename='')`
Save config file

`dumpgenerator.saveImageNames (config={}, images=[], session=None)`
Save image list in a file, including filename, url and uploader

`dumpgenerator.saveIndexPHP (config={}, session=None)`
Save index.php as .html, to preserve license details available at the botom of the page

`dumpgenerator.saveLogs (config={}, session=None)`
Save Special:Log

`dumpgenerator.saveSiteInfo (config={}, session=None)`
Save a file with site info

`dumpgenerator.saveSpecialVersion (config={}, session=None)`
Save Special:Version as .html, to preserve extensions details

`dumpgenerator.truncateFilename (other={}, filename='')`
Truncate filenames when downloading images with large filenames

`dumpgenerator.undoHTMLEntities (text='')`
Undo some HTML codes

Indices and tables

- `genindex`
- `modindex`
- `search`

d

`dumpgenerator`, [1](#)

A

avoidWikimediaProjects() (in module dumpgenerator), 1

B

bye() (in module dumpgenerator), 1

C

checkAPI() (in module dumpgenerator), 1
checkIndex() (in module dumpgenerator), 1
checkXMLIntegrity() (in module dumpgenerator), 1
cleanHTML() (in module dumpgenerator), 1
cleanXML() (in module dumpgenerator), 1
curateImageURL() (in module dumpgenerator), 1

D

delay() (in module dumpgenerator), 1
domain2prefix() (in module dumpgenerator), 1
dumpgenerator (module), 1

F

fixBOM() (in module dumpgenerator), 1

G

generateImageDump() (in module dumpgenerator), 1
generateXMLDump() (in module dumpgenerator), 1
getImageNames() (in module dumpgenerator), 1
getImageNamesAPI() (in module dumpgenerator), 1
getImageNamesScraper() (in module dumpgenerator), 1
getJSON() (in module dumpgenerator), 1
getNamespacesAPI() (in module dumpgenerator), 1
getNamespacesScraper() (in module dumpgenerator), 1
getPageTitles() (in module dumpgenerator), 1
getPageTitlesAPI() (in module dumpgenerator), 1
getPageTitlesScraper() (in module dumpgenerator), 1
getUserAgent() (in module dumpgenerator), 2
getWikiEngine() (in module dumpgenerator), 2
getXMLFileDesc() (in module dumpgenerator), 2
getXMLHeader() (in module dumpgenerator), 2
getXMLPage() (in module dumpgenerator), 2
getXMLPageCore() (in module dumpgenerator), 2

L

loadConfig() (in module dumpgenerator), 2
logerror() (in module dumpgenerator), 2

M

main() (in module dumpgenerator), 2
mwGetAPIAndIndex() (in module dumpgenerator), 2

R

readTitles() (in module dumpgenerator), 2
removeIP() (in module dumpgenerator), 2
reverse_readline() (in module dumpgenerator), 2

S

saveConfig() (in module dumpgenerator), 2
saveImageNames() (in module dumpgenerator), 2
saveIndexPHP() (in module dumpgenerator), 2
saveLogs() (in module dumpgenerator), 2
saveSiteInfo() (in module dumpgenerator), 2
saveSpecialVersion() (in module dumpgenerator), 2

T

truncateFilename() (in module dumpgenerator), 2

U

undoHTMLEntities() (in module dumpgenerator), 2