
idutils Documentation

Release 1.1.1dev20181118

Invenio Software

Nov 28, 2018

Contents

1	Installation	3
2	API	5
3	Changes	9
4	Contributing	11
5	License	13
6	Authors	15
	Python Module Index	17

Small library for validating and normalising persistent identifiers used in scholarly communication.

- Free software: Revised BSD license
- Documentation: <https://idutils.readthedocs.io>.
- Validation and normalization of persistent identifiers.
- Detection of persistent identifier scheme.
- Generation of resolving links for persistent identifiers.
- Supported schemes: ISBN10, ISBN13, ISSN, ISTC, DOI, Handle, EAN8, EAN13, ISNI ORCID, ARK, PURL, LSID, URN, Bibcode, arXiv, PubMed ID, PubMed Central ID, GND, SRA, BioProject, BioSample, Ensembl, UniProt, RefSeq, Genome Assembly.

CHAPTER 1

Installation

The IDUtils package is on PyPI so all you need is:

```
$ pip install idutils
```


Small library for persistent identifiers used in scholarly communication.

`idutils.is_isbn10 (val)`

Test if argument is an ISBN-10 number.

Courtesy Wikipedia: http://en.wikipedia.org/wiki/International_Standard_Book_Number

`idutils.is_isbn13 (val)`

Test if argument is an ISBN-13 number.

Courtesy Wikipedia: http://en.wikipedia.org/wiki/International_Standard_Book_Number

`idutils.is_isbn (val)`

Test if argument is an ISBN-10 or ISBN-13 number.

`idutils.is_issn (val)`

Test if argument is an ISSN number.

`idutils.is_istc (val)`

Test if argument is a International Standard Text Code.

See http://www.istc-international.org/html/about_structure_syntax.aspx

`idutils.is_doi (val)`

Test if argument is a DOI.

`idutils.is_handle (val)`

Test if argument is a Handle.

Note, DOIs are also handles, and handle are very generic so they will also match e.g. any URL your parse.

`idutils.is_ean8 (val)`

Test if argument is a International Article Number (EAN-8).

`idutils.is_ean13 (val)`

Test if argument is a International Article Number (EAN-13).

`idutils.is_ean (val)`

Test if argument is a International Article Number (EAN-13 or EAN-8).

See [http://en.wikipedia.org/wiki/International_Article_Number_\(EAN\)](http://en.wikipedia.org/wiki/International_Article_Number_(EAN)).

`idutils.is_isni(val)`

Test if argument is an International Standard Name Identifier.

`idutils.is_orcid(val)`

Test if argument is an ORCID ID.

See <http://support.orcid.org/knowledgebase/articles/116780-structure-of-the-orcid-identifier>

`idutils.is_purl(val)`

Test if argument is a PURL.

`idutils.is_url(val)`

Test if argument is a URL.

`idutils.is_lsid(val)`

Test if argument is a LSID.

`idutils.is_urn(val)`

Test if argument is an URN.

`idutils.is_ads(val)`

Test if argument is an ADS bibliographic code.

`idutils.is_arxiv_post_2007(val)`

Test if argument is a post-2007 arXiv ID.

`idutils.is_arxiv_pre_2007(val)`

Test if argument is a pre-2007 arXiv ID.

`idutils.is_arxiv(val)`

Test if argument is an arXiv ID.

See http://arxiv.org/help/arxiv_identifier and http://arxiv.org/help/arxiv_identifier_for_services.

`idutils.is_pmid(val)`

Test if argument is a PubMed ID.

Warning: PMID are just integers, with no structure, so this function will say any integer is a PubMed ID

`idutils.is_pmcid(val)`

Test if argument is a PubMed Central ID.

`idutils.is_gnd(val)`

Test if argument is a GND Identifier.

`idutils.is_sra(val)`

Test if argument is an SRA accession.

`idutils.is_bioproject(val)`

Test if argument is a BioProject accession.

`idutils.is_biosample(val)`

Test if argument is a BioSample accession.

`idutils.is_ensembl(val)`

Test if argument is an Ensembl accession.

`idutils.is_uniprot(val)`

Test if argument is a UniProt accession.

`idutils.is_refseq(val)`

Test if argument is a RefSeq accession.

`idutils.is_genome(val)`

Test if argument is a GenBank or RefSeq genome assembly accession.

`idutils.detect_identifier_schemes(val)`

Detect persistent identifier scheme for a given value.

Note: Some schemes like PMID are very generic.

`idutils.normalize_doi(val)`

Normalize a DOI.

`idutils.normalize_handle(val)`

Normalize a Handle identifier.

`idutils.normalize_ads(val)`

Normalize an ADS bibliographic code.

`idutils.normalize_orcid(val)`

Normalize an ORCID identifier.

`idutils.normalize_gnd(val)`

Normalize a GND identifier.

`idutils.normalize_pmid(val)`

Normalize an PubMed ID.

`idutils.normalize_arxiv(val)`

Normalize an arXiv identifier.

`idutils.normalize_pid(val, scheme)`

Normalize an identifier.

E.g. doi:10.1234/foo and http://dx.doi.org/10.1234/foo and 10.1234/foo will all be normalized to 10.1234/foo.

`idutils.to_url(val, scheme, url_scheme='http')`

Convert a resolvable identifier into a URL for a landing page.

Parameters

- **val** – The identifier's value.
- **scheme** – The identifier's scheme.
- **url_scheme** – Scheme to use for URL generation, 'http' or 'https'.

Returns URL for the identifier.

New in version 0.3.0: `url_scheme` used for URL generation.

CHAPTER 3

Changes

Version 1.1.1 (2018-11-18)

- Changes URL resolution for bibcodes to use <https://ui.adsabs.harvard> instead of <https://adsabs.harvard.edu/abs/>.
- Allows choosing HTTP/HTTPS for any generated URL by `idutils.to_url`.

Version 1.1.0 (2018-08-17)

- Adds support for genomic identifiers: SRA, BioProject, BioSample, Ensembl, UniProt, RefSeq, GenBank/RefSeq.
- Fixes bug in bibcode detection for non-capitalized journals.

Version 1.0.1 (2018-05-02)

- Fixes bug causing invalid DOIs to be accepted.

Version 1.0.0 (2017-12-07)

- Fixes handling of unicode characters in DOIs.
- Adds support for APS style arXiv identifiers.

Version 0.2.4 (2017-01-30)

- Removes *Python 3.3* from a list of supported Python versions and adds *Python 3.6*
- Moves from *isbnid* (v0.3.4) to *isbnid_fork* (v0.4.4) library.

Version 0.2.3 (2016-09-21)

- Adds an optional parameter in `idutils.to_url` to use HTTPS scheme for PID providers that support it.
- Detects and parses Handles and DOIs without the “http(s)://”, and ignores whitespace after scheme tags (eg. “doi: 10.123/456”).

Version 0.2.2 (2016-09-16)

- Fixes issue where a valid ISBN with dashes and spaces could not be normalized.

Version 0.2.1 (2016-06-17)

- Changes ISBN normalization to use *isbnid* instead of *isbnlib*. Now, importing this library will not change the default socket timeout, resulting in unwanted side effects.

Version 0.2.0 (2016-04-07)

- Changes URL resolution for DOIs to use <https://doi.org> instead of <http://dx.doi.org> according to https://www.doi.org/doi_handbook/3_Resolution.html#3.8

Version 0.1.1 (2015-07-22)

- Fixes GND validation and normalization.
- Replaces invalid package name in *run-tests.sh* and makes *run-tests.sh* file executable. One can now use *docker-compose run --rm web /code/run-tests.sh* to run all the CI tests (pep257, sphinx, test suite).
- Initial release of Docker configuration suitable for local developments. *docker-compose build* rebuilds the image, *docker-compose run --rm web /code/run-tests.sh* runs the test suite.

Version 0.1.0 (2015-07-02)

- First public release.

CHAPTER 4

Contributing

Bug reports, feature requests, and other contributions are welcome. If you find a demonstrable problem that is caused by the code of this library, please:

1. Search for [already reported problems](#).
2. Check if the issue has been fixed or is still reproducible on the latest *master* branch.
3. Create an issue with **a test case**.

If you create a feature branch, you can run the tests to ensure everything is operating correctly:

```
$ ./run-tests.sh
```


CHAPTER 5

License

IDUtils is free software; you can redistribute it and/or modify it under the terms of the Revised BSD License quoted below.

Copyright (C) 2015-2018 CERN. Copyright (C) 2018 Alan Rubin.

All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
- Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDERS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

In applying this license, CERN does not waive the privileges and immunities granted to it by virtue of its status as an Intergovernmental Organization or submit itself to any jurisdiction.

CHAPTER 6

Authors

- Adrian Pawel Baran
- Alan Rubin
- Alexander Ioannidis
- Jiri Kuncar
- Lars Holm Nielsen
- Pedro Gaudencio
- Tibor Simko

i

idutils, 5

D

`detect_identifier_schemes()` (in module `idutils`), 7

I

`idutils` (module), 5

`is_ads()` (in module `idutils`), 6

`is_arxiv()` (in module `idutils`), 6

`is_arxiv_post_2007()` (in module `idutils`), 6

`is_arxiv_pre_2007()` (in module `idutils`), 6

`is_bioproject()` (in module `idutils`), 6

`is_biosample()` (in module `idutils`), 6

`is_doi()` (in module `idutils`), 5

`is_ean()` (in module `idutils`), 5

`is_ean13()` (in module `idutils`), 5

`is_ean8()` (in module `idutils`), 5

`is_ensembl()` (in module `idutils`), 6

`is_genome()` (in module `idutils`), 6

`is_gnd()` (in module `idutils`), 6

`is_handle()` (in module `idutils`), 5

`is_isbn()` (in module `idutils`), 5

`is_isbn10()` (in module `idutils`), 5

`is_isbn13()` (in module `idutils`), 5

`is_isni()` (in module `idutils`), 6

`is_issn()` (in module `idutils`), 5

`is_istc()` (in module `idutils`), 5

`is_lsid()` (in module `idutils`), 6

`is_orcid()` (in module `idutils`), 6

`is_pmcid()` (in module `idutils`), 6

`is_pmid()` (in module `idutils`), 6

`is_purl()` (in module `idutils`), 6

`is_refseq()` (in module `idutils`), 6

`is_sra()` (in module `idutils`), 6

`is_uniprot()` (in module `idutils`), 6

`is_url()` (in module `idutils`), 6

`is_urn()` (in module `idutils`), 6

N

`normalize_ads()` (in module `idutils`), 7

`normalize_arxiv()` (in module `idutils`), 7

`normalize_doi()` (in module `idutils`), 7

`normalize_gnd()` (in module `idutils`), 7

`normalize_handle()` (in module `idutils`), 7

`normalize_orcid()` (in module `idutils`), 7

`normalize_pid()` (in module `idutils`), 7

`normalize_pmid()` (in module `idutils`), 7

T

`to_url()` (in module `idutils`), 7