wn

Release 0.9.3

Michael Wayne Goodman

Nov 14, 2022

CONTENTS

CHAPTER

ONE

OVERVIEW

This package provides an interface to wordnet data, from simple lookup queries, to graph traversals, to more sophisticated algorithms and metrics. Features include:

- Support for wordnets in the WN-LMF format
- A SQLite database backend for data consistency and efficient queries
- Accurate modeling of Words, Senses, and Synsets

CHAPTER

TWO

QUICK START

```
$ pip install wn
```

```
>>> import wn
>>> wn.download('ewn:2020')
>>> wn.synsets('coffee')
[Synset('ewn-04979718-n'), Synset('ewn-07945591-n'), Synset('ewn-07945759-n'), Synset(
↪'ewn-12683533-n')]
```

CHAPTER

THREE

CONTENTS

3.1 Installation and Configuration

See also:

This guide is for installing and configuring the Wn software. For adding lexicons to the database, see *Working with Lexicons*.

3.1.1 Installing from PyPI

Install the latest release from PyPI:

```
pip install wn
```

To get the dependencies for the `wn.web` module, use the `web` installation extra:

```
pip install wn[web]
```

3.1.2 The Data Directory

By default, Wn stores its data (such as downloaded LMF files and the database file) in a `.wn_data/` directory under the user's home directory. This directory can be changed (see *Configuration* below). Whenever Wn attempts to download a resource or access its database, it will check for the existence of, and create if necessary, this directory, the `.wn_data/downloads/` subdirectory, and the `.wn_data/wn.db` database file. The file system will look like this:

```
.wn_data/
├── downloads
│   ├── ...
│   └── ...
└── wn.db
```

The `...` entries in the `downloads/` subdirectory represent the files of resources downloaded from the web. Their filename is a hash of the URL so that Wn can avoid downloading the same file twice.

3.1.3 Configuration

The *wn.config* object contains the paths Wn uses for local storage and information about resources available on the web. To change the directory Wn uses for storing data locally, modify the `wn.config.data_directory` member:

```
import wn
wn.config.data_directory = '~/Projects/wn_data'
```

There are some things to note:

- The downloads directory and database path are always relative to the data directory and cannot be changed directly.
- This change only affects subsequent operations, so any data in the previous location will not be moved nor deleted.
- This change only affects the current session. If you want a script or application to always use the new location, it must reset the data directory each time it is initialized.

You can also add project information for remote resources. First you add a project, with a project ID, full name, and language code. Then you create one or more versions for that project with a version ID, resource URL, and license information. This may be done either through the *wn.config* object's *add_project()* and *add_project_version()* methods, or loaded from a TOML file via the *wn.config* object's *load_index()* method.

```
wn.config.add_project('ewn', 'English WordNet', 'en')
wn.config.add_project_version(
    'ewn', '2020',
    'https://en-word.net/static/english-wordnet-2020.xml.gz',
    'https://creativecommons.org/licenses/by/4.0/',
)
```

3.1.4 Installing From Source

If you wish to install the code from the source repository (e.g., to get an unreleased feature or to contribute toward Wn's development), clone the repository and use Flit to install:

```
$ git clone https://github.com/goodmami/wn.git
$ cd wn
$ flit install
```

Developers of Wn may want to use the `--symlink` option which makes the install "editable" (subsequent edits to the source code will be reflected without having to reinstall):

```
$ flit install --symlink
```

3.2 Command Line Interface

Some of Wn's functionality is exposed via the command line.

3.2.1 Global Options

-d DIR, **--dir** DIR

Change to use DIR as the data directory prior to invoking any commands.

3.2.2 Subcommands

3.2.3 download

Download and add projects to the database given one or more project specifiers or URLs.

```
$ python -m wn download oewn:2021 omw:1.4 cili
$ python -m wn download https://en-word.net/static/english-wordnet-2021.xml.gz
```

--index FILE

Use the index at FILE to resolve project specifiers.

```
$ python -m wn download --index my-index.toml mywn
```

--no-add

Download and cache the remote file, but don't add it to the database.

3.2.4 lexicons

The lexicons subcommand lets you quickly see what is installed:

```
$ python -m wn lexicons
omw-en       1.4     [en]    OMW English Wordnet based on WordNet 3.0
omw-sk       1.4     [sk]    Slovak WordNet
omw-pl       1.4     [pl]    plWordNet
omw-is       1.4     [is]    IceWordNet
omw-zsm      1.4     [zsm]   Wordnet Bahasa (Malaysian)
omw-sl       1.4     [sl]    sloWNet
omw-ja       1.4     [ja]    Japanese Wordnet
...
```

-l LG, **--lang** LG

--lexicon SPEC

The --lang or --lexicon option can help you narrow down the results:

```
$ python -m wn lexicons --lang en
oewn      2021    [en]    Open English WordNet
omw-en    1.4     [en]    OMW English Wordnet based on WordNet 3.0
$ python -m wn lexicons --lexicon "omw-*"
omw-en    1.4     [en]    OMW English Wordnet based on WordNet 3.0
omw-sk    1.4     [sk]    Slovak WordNet
omw-pl    1.4     [pl]    plWordNet
omw-is    1.4     [is]    IceWordNet
omw-zsm   1.4     [zsm]   Wordnet Bahasa (Malaysian)
```

3.2.5 projects

The `projects` subcommand lists all known projects in Wn's index. This is helpful to see what is available for downloading.

```
$ python -m wn projects
ic   cili    1.0     [---]   Collaborative Interlingual Index
ic   oewn    2021    [en]    Open English WordNet
ic   ewn     2020    [en]    Open English WordNet
ic   ewn     2019    [en]    Open English WordNet
i-   odenet  1.4     [de]    Open German WordNet
ic   odenet  1.3     [de]    Open German WordNet
ic   omw     1.4     [mul]   Open Multilingual Wordnet
ic   omw-en  1.4     [en]    OMW English Wordnet based on WordNet 3.0
...
```

3.2.6 validate

Given a path to a WN-LMF XML file, check the file for structural problems and print a report.

```
$ python -m wn validate english-wordnet-2021.xml
```

--select CHECKS

Run the checks with the given comma-separated list of check codes or categories.

```
$ python -m wn validate --select E W201 W204 deWordNet.xml
```

--output-file FILE

Write the report to FILE as a JSON object instead of printing the report to stdout.

3.3 FAQ

3.3.1 Is Wn related to the NLTK's *nltk.corpus.wordnet* module?

Only in spirit. There was an effort to develop the NLTK's module as a standalone package (see https://github.com/nltk/wordnet/), but development had slowed. Wn has the same broad goals and a similar API as that standalone package, but fundamental architectural differences demanded a complete rewrite, so Wn was created as a separate project. With approval from the other package's maintainer, Wn acquired the wn project on PyPI and can be seen as its successor.

3.3.2 Is Wn compatible with the NLTK's module?

The API is intentionally similar, but not exactly the same (for instance see the next question), and there are differences in the ways that results are retrieved, particularly for non-English wordnets. See *Migrating from the NLTK* for more information. Also see *Where is the Princeton WordNet data?*.

3.3.3 Where are the `Lemma` objects? What are `Word` and `Sense` objects?

Unlike the original WNDB data format of the original WordNet, the WN-LMF XML format grants words (called *lexical entries* in WN-LMF and a *`Word`* object in Wn) and word senses (*`Sense`* in Wn) explicit, first-class status alongside synsets. While senses are essentially links between words and synsets, they may contain metadata and be the source or target of sense relations, so in some ways they are more like nodes than edges when the wordnet is viewed as a graph. The NLTK's module, using the WNDB format, combines the information of a word and a sense into a single object called a `Lemmas`. Wn also has an unrelated concept called a *`lemma()`*, but it is merely the canonical form of a word.

3.3.4 Where is the Princeton WordNet data?

The original English wordnet, named simply *WordNet* but often referred to as the *Princeton WordNet* to better distinguish it from other projects, is specifically the data distributed by Princeton in the WNDB format. The Open Multilingual Wordnet (OMW) packages an export of the WordNet data as the *OMW English Wordnet based on WordNet 3.0* which is used by Wn (with the lexicon ID `omw-en`). It also has a similar export for WordNet 3.1 data (`omw-en31`). Both of these are highly compatible with the original data and can be used as drop-in replacements.

Prior to Wn version 0.9 (and, correspondingly, prior to the OMW data version 1.4), the `pwn:3.0` and `pwn:3.1` English wordnets distributed by OMW were incorrectly called the *Princeton WordNet* (for WordNet 3.0 and 3.1, respectively). From Wn version 0.9 (and from version 1.4 of the OMW data), these are called the *OMW English Wordnet based on WordNet 3.0/3.1* (`omw-en:1.4` and `omw-en31:1.4`, respectively). These lexicons are intentionally compatible with the original WordNet data, and the 1.4 versions are even more compatible than the previous `pwn:3.0` and `pwn:3.1` lexicons, so it is strongly recommended to use them over the previous versions.

3.3.5 Why don't all wordnets share the same synsets?

The Open Multilingual Wordnet (OMW) contains wordnets for many languages created using the *expand* methodology [VOSSEN1998], where non-English wordnets provide words on top of the English wordnet's synset structure. This allows new wordnets to be built in much less time than starting from scratch, but with a few drawbacks, such as that words cannot be added if they do not have a synset in the English wordnet, and that it is difficult to version the wordnets independently (e.g., for reproducibility of experiments involving wordnet data) as all are interconnected. Wn, therefore, creates new synsets for each wordnet added to its database, and synsets then specify which resource they belong to. Queries can specify which resources may be examined. Also see *Interlingual Queries*.

3.3.6 Why does Wn's database get so big?

The *OMW English Wordnet based on WordNet 3.0* takes about 114 MiB of disk space in Wn's database, which is only about 8 MiB more than it takes as a WN-LMF XML file. The NLTK, however, uses the obsolete WNDB format which is more compact, requiring only 35 MiB of disk space. The difference with the Open Multilingual Wordnet 1.4 is more striking: it takes about 659 MiB of disk space in the database, but only 49 MiB in the NLTK. Part of the difference here is that the OMW files in the NLTK are simple tab-separated-value files listing only the words added to each synset for each language. In addition, Wn creates new synsets for each wordnet added (see the previous question). One more reason is that Wn creates various indexes in the database for efficient lookup.

3.4 Working with Lexicons

3.4.1 Terminology

In Wn, the following terminology is used:

lexicon An inventory of words, senses, synsets, relations, etc. that share a namespace (i.e., that can refer to each other).

wordnet A group of lexicons (but usually just one).

resource A file containing lexicons.

package A directory containing a resource and optionally some metadata files.

collection A directory containing packages and optionally some metadata files.

project A general term for a resource, package, or collection, particularly pertaining to its creation, maintenance, and distribution.

In general, each resource contains one lexicon. For large projects like the Open English WordNet, that lexicon is also a wordnet on its own. For a collection like the Open Multilingual Wordnet, most lexicons do not include relations as they are instead expected to use those from the OMW's included English wordnet, which is derived from the Princeton WordNet. As such, a wordnet for these sub-projects is best thought of as the grouping of the lexicon with the lexicon providing the relations.

3.4.2 Lexicon and Project Specifiers

Wn uses *lexicon specifiers* to deal with the possibility of having multiple lexicons and multiple versions of lexicons loaded in the same database. The specifiers are the joining of a lexicon's name (ID) and version, delimited by `:`. Here are the possible forms:

```
*           -- any/all lexicons
id          -- the most recently added lexicon with the given id
id:*        -- all lexicons with the given id
id:version  -- the lexicon with the given id and version
*:version   -- all lexicons with the given version
```

For example, if `ewn:2020` was installed followed by `ewn:2019`, then `ewn` would specify the `2019` version, `ewn:*` would specify both versions, and `ewn:2020` would specify the `2020` version.

The same format is used for *project specifiers*, which refer to projects as defined in Wn's index. In most cases the project specifier is the same as the lexicon specifier (e.g., `ewn:2020` refers both to the project to be downloaded and the lexicon that is installed), but sometimes it is not. The 1.4 release of the Open Multilingual Wordnet, for instance, has the project specifier `omw:1.4` but it installs a number of lexicons with their own lexicon specifiers (`omw-zsm:1.4`, `omw-cmn:1.4`, etc.). When only an id is given (e.g., `ewn`), a project specifier gets the *first* version listed in the index (in the default index, conventionally, the first version is the latest release).

3.4.3 Downloading Lexicons

Use *wn.download()* to download lexicons from the web given either an indexed project specifier or the URL of a resource, package, or collection.

```
>>> import wn
>>> wn.download('odenet')  # get the latest Open German WordNet
>>> wn.download('odenet:1.3')  # get the 1.3 version
>>> # download from a URL
>>> wn.download('https://github.com/omwn/omw-data/releases/download/v1.4/omw-1.4.tar.xz')
```

The project specifier is only used to retrieve information from Wn's index. The lexicon IDs of the corresponding resource files are what is stored in the database.

3.4.4 Adding Local Lexicons

Lexicons can be added from local files with *wn.add()*:

```
>>> wn.add('~/data/omw-1.4/omw-nb/omw-nb.xml')
```

Or with the parent directory as a package:

```
>>> wn.add('~/data/omw-1.4/omw-nb/')
```

Or with the grandparent directory as a collection (installing all packages contained by the collection):

```
>>> wn.add('~/data/omw-1.4/')
```

Or from a compressed archive of one of the above:

```
>>> wn.add('~/data/omw-1.4/omw-nb/omw-nb.xml.xz')
>>> wn.add('~/data/omw-1.4/omw-nb.tar.xz')
>>> wn.add('~/data/omw-1.4.tar.xz')
```

3.4.5 Listing Installed Lexicons

If you wish to see which lexicons have been added to the database, *wn.lexicons()* returns the list of *wn.Lexicon* objects that describe each one.

```
>>> for lex in wn.lexicons():
...     print(f'{lex.id}:{lex.version}\t{lex.label}')
...
omw-en:1.4      OMW English Wordnet based on WordNet 3.0
omw-nb:1.4      Norwegian Wordnet (Bokmål)
odenet:1.3      Offenes Deutsches WordNet
ewn:2020        English WordNet
ewn:2019        English WordNet
```

3.4.6 Removing Lexicons

Lexicons can be removed from the database with *wn.remove()*:

```
>>> wn.remove('omw-nb:1.4')
```

Note that this removes a single lexicon and not a project, so if, for instance, you've installed a multi-lexicon project like omw, you will need to remove each lexicon individually or use a star specifier:

```
>>> wn.remove('omw-*:1.4')
```

3.4.7 WN-LMF Files, Packages, and Collections

Wn can handle projects with 3 levels of structure:

- WN-LMF XML files
- WN-LMF packages
- WN-LMF collections

WN-LMF XML Files

A WN-LMF XML file is a file with a `.xml` extension that is valid according to the WN-LMF specification.

WN-LMF Packages

If one needs to distribute metadata or additional files along with WN-LMF XML file, a WN-LMF package allows them to include the files in a directory. The directory should contain exactly one `.xml` file, which is the WN-LMF XML file. In addition, it may contain additional files and Wn will recognize three of them:

LICENSE (**.txt** | **.md** | **.rst**) the full text of the license

README (**.txt** | **.md** | **.rst**) the project README

citation.bib a BibTeX file containing academic citations for the project

```
omw-sq/
├── omw-sq.xml
├── LICENSE.txt
└── README.md
```

WN-LMF Collections

In some cases a project may manage multiple resources and distribute them as a collection. A collection is a directory containing subdirectories which are WN-LMF packages. The collection may contain its own README, LICENSE, and citation files which describe the project as a whole.

```
omw-1.4/
├── omw-sq
│   ├── oms-sq.xml
│   ├── LICENSE.txt
│   └── README.md
```

(continues on next page)

(continued from previous page)

```
├── omw-lt
│   ├── citation.bib
│   ├── LICENSE
│   └── omw-lt.xml
├── ...
├── citation.bib
├── LICENSE
└── README.md
```

3.5 Basic Usage

See also:

This document covers the basics of querying wordnets, filtering results, and performing secondary queries on the results. For adding, removing, or inspecting lexicons, see *Working with Lexicons*. For more information about interlingual queries, see *Interlingual Queries*.

For the most basic queries, Wn provides several module functions for retrieving words, senses, and synsets:

```
>>> import wn
>>> wn.words('pike')
[Word('ewn-pike-n')]
>>> wn.senses('pike')
[Sense('ewn-pike-n-03311555-04'), Sense('ewn-pike-n-07795351-01'), Sense('ewn-pike-n-
↪03941974-01'), Sense('ewn-pike-n-03941726-01'), Sense('ewn-pike-n-02563739-01')]
>>> wn.synsets('pike')
[Synset('ewn-03311555-n'), Synset('ewn-07795351-n'), Synset('ewn-03941974-n'), Synset(
↪'ewn-03941726-n'), Synset('ewn-02563739-n')]
```

Once you start working with multiple wordnets, these simple queries may return more than desired:

```
>>> wn.words('pike')
[Word('ewn-pike-n'), Word('wnja-n-66614')]
>>> wn.words('chat')
[Word('ewn-chat-n'), Word('ewn-chat-v'), Word('frawn-lex14803'), Word('frawn-lex21897')]
```

You can specify which language or lexicon you wish to query:

```
>>> wn.words('pike', lang='ja')
[Word('wnja-n-66614')]
>>> wn.words('chat', lexicon='frawn')
[Word('frawn-lex14803'), Word('frawn-lex21897')]
```

But it might be easier to create a *Wordnet* object and use it for queries:

```
>>> wnja = wn.Wordnet(lang='ja')
>>> wnja.words('pike')
[Word('wnja-n-66614')]
>>> frawn = wn.Wordnet(lexicon='frawn')
>>> frawn.words('chat')
[Word('frawn-lex14803'), Word('frawn-lex21897')]
```

In fact, the simple queries above implicitly create such a *Wordnet* object, but one that includes all installed lexicons.

3.5.1 Primary Queries

The queries shown above are "primary" queries, meaning they are the first step in a user's interaction with a wordnet. Operations performed on the resulting objects are then *secondary queries*. Primary queries optionally take several fields for filtering the results, namely the word form and part of speech. Synsets may also be filtered by an interlingual index (ILI).

Searching for Words

The *wn.words()* function returns a list of *Word* objects that match the given word form or part of speech:

```
>>> wn.words('pencil')
[Word('ewn-pencil-n'), Word('ewn-pencil-v')]
>>> wn.words('pencil', pos='v')
[Word('ewn-pencil-v')]
```

Calling the function without a word form will return all words in the database:

```
>>> len(wn.words())
311711
>>> len(wn.words(pos='v'))
29419
>>> len(wn.words(pos='v', lexicon='ewn'))
11595
```

If you know the word identifier used by a lexicon, you can retrieve the word directly with the *wn.word()* function. Identifiers are guaranteed to be unique within a single lexicon, but not across lexicons, so it's best to call this function from an instantiated *Wordnet* object or with the `lexicon` parameter specified. If multiple words are found when querying multiple lexicons, only the first is returned.

```
>>> wn.word('ewn-pencil-n', lexicon='ewn')
Word('ewn-pencil-n')
```

Searching for Senses

The *wn.senses()* and *wn.sense()* functions behave similarly to *wn.words()* and *wn.word()*, except that they return matching *Sense* objects.

```
>>> wn.senses('plow', pos='n')
[Sense('ewn-plow-n-03973894-01')]
>>> wn.sense('ewn-plow-v-01745745-01')
Sense('ewn-plow-v-01745745-01')
```

Senses represent a relationship between a *Word* and a *Synset*. Seen as an edge between nodes, senses are often given less prominence than words or synsets, but they are the natural locus of several interesting features such as sense relations (e.g., for derived words) and the natural level of representation for translations to other languages.

Searching for Synsets

The *wn.synsets()* and *wn.synset()* functions are like those above but allow the `ili` parameter for filtering by interlingual index, which is useful in interlingual queries:

```
>>> wn.synsets('scepter')
[Synset('ewn-14467142-n'), Synset('ewn-07282278-n')]
>>> wn.synset('ewn-07282278-n').ili
'i74874'
>>> wn.synsets(ili='i74874')
[Synset('ewn-07282278-n'), Synset('wnja-07267573-n'), Synset('frawn-07267573-n')]
```

3.5.2 Secondary Queries

Once you have gotten some results from a primary query, you can perform operations on the *Word*, *Sense*, or *Synset* objects to get at further information in the wordnet.

Exploring Words

Here are some of the things you can do with *Word* objects:

```
>>> w = wn.words('goose')[0]
>>> w.pos  # part of speech
'n'
>>> w.forms()  # other word forms (e.g., irregular inflections)
['goose', 'geese']
>>> w.lemma()  # canonical form
'goose'
>>> w.derived_words()
[Word('ewn-gosling-n'), Word('ewn-goosy-s'), Word('ewn-goosey-s')]
>>> w.senses()
[Sense('ewn-goose-n-01858313-01'), Sense('ewn-goose-n-10177319-06'), Sense('ewn-goose-n-
↪07662430-01')]
>>> w.synsets()
[Synset('ewn-01858313-n'), Synset('ewn-10177319-n'), Synset('ewn-07662430-n')]
```

Since translations of a word into another language depend on the sense used, *Word.translate* returns a dictionary mapping each sense to words in the target language:

```
>>> for sense, ja_words in w.translate(lang='ja').items():
...     print(sense, ja_words)
...
Sense('ewn-goose-n-01858313-01') [Word('wnja-n-1254'), Word('wnja-n-33090'), Word('wnja-
↪n-38995')]
Sense('ewn-goose-n-10177319-06') []
Sense('ewn-goose-n-07662430-01') [Word('wnja-n-1254')]
```

Exploring Senses

Compared to *Word* and *Synset* objects, there are relatively few operations available on *Sense* objects. Sense relations and translations, however, are important operations on senses.

```
>>> s = wn.senses('dark', pos='n')[0]
>>> s.word()    # each sense links to a single word
Word('ewn-dark-n')
>>> s.synset()  # each sense links to a single synset
Synset('ewn-14007000-n')
>>> s.get_related('antonym')
[Sense('ewn-light-n-14006789-01')]
>>> s.get_related('derivation')
[Sense('ewn-dark-a-00273948-01')]
>>> s.translate(lang='fr')  # translation returns a list of senses
[Sense('frawn-lex52992--13983515-n')]
>>> s.translate(lang='fr')[0].word().lemma()
'obscurité'
```

Exploring Synsets

Many of the operations people care about happen on synsets, such as hierarchical relations and metrics.

```
>>> ss = wn.synsets('hound', pos='n')[0]
>>> ss.senses()
[Sense('ewn-hound-n-02090203-01'), Sense('ewn-hound_dog-n-02090203-02')]
>>> ss.words()
[Word('ewn-hound-n'), Word('ewn-hound_dog-n')]
>>> ss.lemmas()
['hound', 'hound dog']
>>> ss.definition()
'any of several breeds of dog used for hunting typically having large drooping ears'
>>> ss.hypernyms()
[Synset('ewn-02089774-n')]
>>> ss.hypernyms()[0].lemmas()
['hunting dog']
>>> len(ss.hyponyms())
20
>>> ss.hyponyms()[0].lemmas()
['Afghan', 'Afghan hound']
>>> ss.max_depth()
15
>>> ss.shortest_path(wn.synsets('dog', pos='n')[0])
[Synset('ewn-02090203-n'), Synset('ewn-02089774-n'), Synset('ewn-02086723-n')]
>>> ss.translate(lang='fr')  # translation returns a list of synsets
[Synset('frawn-02087551-n')]
>>> ss.translate(lang='fr')[0].lemmas()
['chien', 'chien de chasse']
```

3.5.3 Filtering by Language

The `lang` parameter of *wn.words()*, *wn.senses()*, *wn.synsets()*, and *Wordnet* allows a single BCP 47 language code. When this parameter is used, only entries in the specified language will be returned.

```
>>> import wn
>>> wn.words('chat')
[Word('ewn-chat-n'), Word('ewn-chat-v'), Word('frawn-lex14803'), Word('frawn-lex21897')]
>>> wn.words('chat', lang='fr')
[Word('frawn-lex14803'), Word('frawn-lex21897')]
```

If a language code not used by any lexicon is specified, a *wn.Error* is raised.

3.5.4 Filtering by Lexicon

The `lexicon` parameter of *wn.words()*, *wn.senses()*, *wn.synsets()*, and *Wordnet* take a string of space-delimited *lexicon specifiers*. Entries in a lexicon whose ID matches one of the lexicon specifiers will be returned. For these, the following rules are used:

- A full `id:version` string (e.g., `ewn:2020`) selects a specific lexicon
- Only a lexicon `id` (e.g., `ewn`) selects the most recently added lexicon with that ID
- A star * may be used to match any lexicon; a star may not include a version

```
>>> wn.words('chat', lexicon='ewn:2020')
[Word('ewn-chat-n'), Word('ewn-chat-v')]
>>> wn.words('chat', lexicon='wnja')
[]
>>> wn.words('chat', lexicon='wnja frawn')
[Word('frawn-lex14803'), Word('frawn-lex21897')]
```

3.6 Interlingual Queries

This guide explains how interlingual queries work within Wn. To get started, you'll need at least two lexicons that use interlingual indices (ILIs). For this guide, we'll use the Open English WordNet (`oewn:2021`), the Open German WordNet (`odenet:1.4`), also known as OdeNet, and the Japanese wordnet (`omw-ja:1.4`).

```
>>> import wn
>>> wn.download('oewn:2021')
>>> wn.download('odenet:1.4')
>>> wn.download('omw-ja:1.4')
```

We will query these wordnets with the following *Wordnet* objects:

```
>>> en = wn.Wordnet('oewn:2021')
>>> de = wn.Wordnet('odenet:1.4')
```

The object for the Japanese wordnet will be discussed and created below, in *Cross-lingual Relation Traversal*.

3.6.1 What are Interlingual Indices?

It is common for users of the Princeton WordNet to refer to synsets by their WNDB offset and type, but this is problematic because the offset is a byte-offset in the wordnet data files and it will differ for wordnets in other languages and even between versions of the same wordnet. Interlingual indices (ILIs) address this issue by providing stable identifiers for concepts, whether for a synset across versions of a wordnet or across languages.

The idea of ILIs was proposed by [Vossen99] and it came to fruition with the release of the Collaborative Interlingual Index (CILI; [Bond16]). CILI therefore represents an instance of, and a namespace for, ILIs. There could, in theory, be alternative indexes for particular domains (e.g., names of people or places), but currently there is only the one.

As an example, the synset for *apricot* (fruit) in WordNet 3.0 is `07750872-n`, but it is `07766848-n` in WordNet 3.1. In OdeNet 1.4, which is not released in the WNDB format and therefore doesn't use offsets at all, it is `13235-n` for the equivalent word (*Aprikose*). However, all three use the same ILI: `i77784`.

Not every synset is guaranteed to be associated with an ILI, and some have the special value `in` indicates that the project is proposing that a new ILI be created in the CILI project for the concept, but until that happens it cannot be used in interlingual queries.

3.6.2 Using Interlingual Indices

For synsets that have an associated ILI, you can retrieve it via the *wn.Synset.ili* accessor:

```
>>> apricot = en.synsets('apricot')[1]
>>> apricot.ili
ILI('i77784')
```

From this object you can get various properties of the ILI, such as the ID as a string, its status, and its definition, but if you have not added CILI to Wn's database it will not be very informative:

```
>>> apricot.ili.id
'i77784'
>>> apricot.ili.status
'presupposed'
>>> apricot.ili.definition() is None
True
```

The `presupposed` status means that the ILI was in use by a lexicon, but there is no other source of truth for the index. CILI can be downloaded just like a lexicon:

```
>>> wn.download('cili:1.0')
```

Now the status and definition should be more useful:

```
>>> apricot.ili.status
'active'
>>> apricot.ili.definition()
'downy yellow to rosy-colored fruit resembling a small peach'
```

ILI IDs may be used to lookup synsets:

```
>>> Aprikose = de.synsets(ili=apricot.ili.id)[0]
>>> Aprikose.lemmas()
['Marille', 'Aprikose']
```

3.6.3 Translating Words, Senses, and Synsets

Rather than manually inserting the ILI IDs into Wn's lookup functions as shown above, Wn provides the *wn.Synset.translate()* method to make it easier:

```
>>> apricot.translate(lexicon='odenet:1.4')
[Synset('odenet-13235-n')]
```

The method returns a list for two reasons: first, it's not guaranteed that the target lexicon has only one synset with the ILI and, second, you can translate to more than one lexicon at a time.

Sense objects also have a *translate()* method, returning a list of senses instead of synsets:

```
>>> de_senses = apricot.senses()[0].translate(lexicon='odenet:1.4')
>>> [s.word().lemma() for s in de_senses]
['Marille', 'Aprikose']
```

Word have a *translate()* method, too, but it works a bit differently. Since each word may be part of multiple synsets, the method returns a mapping of each word sense to the list of translated words:

```
>>> result = en.words('apricot')[0].translate(lexicon='odenet:1.4')
>>> for sense, de_words in result.items():
...     print(sense, [w.lemma() for w in de_words])
...
Sense('oewn-apricot__1.20.00..') []
Sense('oewn-apricot__1.13.00..') ['Marille', 'Aprikose']
Sense('oewn-apricot__1.07.00..') ['lachsrosa', 'lachsfarbig', 'in Lachs', 'lachsfarben',
↪'lachsrot', 'lachs']
```

The three senses above are for *apricot* as a tree, a fruit, and a color. OdeNet does not have a synset for apricot trees, or it has one not associated with the appropriate ILI, and therefore it could not translate any words for that sense.

3.6.4 Cross-lingual Relation Traversal

ILIs have a second use in Wn, which is relation traversal for wordnets that depend on other lexicons, i.e., those created with the *expand* methodology. These wordnets, such as many of those in the Open Multilingual Wordnet, do not include synset relations on their own as they were built using the English WordNet as their taxonomic scaffolding. Trying to load such a lexicon when the lexicon it requires is not added to the database presents a warning to the user:

```
>>> ja = wn.Wordnet('omw-ja:1.4')
[...] WnWarning: lexicon dependencies not available: omw-en:1.4
>>> ja.expanded_lexicons()
[]
```

Warning: Do not rely on the presence of a warning to determine if the lexicon has its expand lexicon loaded. Python's default warning filter may only show the warning the first time it is encountered. Instead, inspect *wn.Wordnet.expanded_lexicons()* to see if it is non-empty.

When a dependency is unmet, Wn only issues a warning, not an error, and you can continue to use the lexicon as it is, but it won't be useful for exploring relations such as hypernyms and hyponyms:

```
>>> anzu = ja.synsets(ili='i77784')[0]
>>> anzu.lemmas()
['', '', '']
>>> anzu.hypernyms()
[]
```

One way to resolve this issue is to install the lexicon it requires:

```
>>> wn.download('omw-en:1.4')
>>> ja = wn.Wordnet('omw-ja:1.4')  # no warning
>>> ja.expanded_lexicons()
[<Lexicon omw-en:1.4 [en]>]
```

Wn will detect the dependency and load `omw-en:1.4` as the *expand* lexicon for `omw-ja:1.4` when the former is in the database. You may also specify an expand lexicon manually, even one that isn't the specified dependency:

```
>>> ja = wn.Wordnet('omw-ja:1.4', expand='oewn:2021')  # no warning
>>> ja.expanded_lexicons()
[<Lexicon oewn:2021 [en]>]
```

In this case, the Open English WordNet is an actively-developed fork of the lexicon that `omw-ja:1.4` depends on, and it should contain all the relations, so you'll see little difference between using it and `omw-en:1.4`. This works because the relations are found using ILIs and not synset offsets. You may still prefer to use the specified dependency if you have strict compatibility needs, such as for experiment reproducibility and/or compatibility with the NLTK. Using some other lexicon as the expand lexicon may yield very different results. For instance, `odenet:1.4` is much smaller than the English wordnets and has fewer relations, so it would not be a good substitute for `omw-ja:1.4`'s expand lexicon.

When an appropriate expand lexicon is loaded, relations between synsets, such as hypernyms, are more likely to be present:

```
>>> anzu = ja.synsets(ili='i77784')[0]  # recreate the synset object
>>> anzu.hypernyms()
[Synset('omw-ja-07705931-n')]
>>> anzu.hypernyms()[0].lemmas()
['']
>>> anzu.hypernyms()[0].translate(lexicon='oewn:2021')[0].lemmas()
['edible fruit']
```

3.7 The Structure of a Wordnet

A **wordnet** is an online lexicon which is organized by concepts.

The basic unit of a wordnet is the synonym set (**synset**), a group of words that all refer to the same concept. Words and synsets are linked by means of conceptual-semantic relations to form the structure of wordnet.

3.7.1 Words, Senses, and Synsets

We all know that **words** are the basic building blocks of languages, a word is built up with two parts, its form and its meaning, but in natural languages, the word form and word meaning are not in an elegant one-to-one match, one word form may connect to many different meanings, so hereforth, we need **senses**, to work as the unit of word meanings, for example, the word *bank* has at least two senses:

1. bank[1]: financial institution, like *City Bank*;
2. bank[2]: sloping land, like *river bank*;

Since **synsets** are group of words sharing the same concept, bank[1]and bank[2]are members of two different synsets, although they have the same word form.

On the other hand, different word forms may also convey the same concept, such as *cab* and *taxi*, these word forms with the same concept are grouped together into one synset.

Figure: relations between words, senses and synsets

3.7.2 Synset Relations

In wordnet, synsets are linked with each other to form various kinds of relations. For example, if the concept expressed by a synset is more general than a given synset, then it is in a *hypernym* relation with the given synset. As shown in the figure below, the synset with *car*, *auto* and *automobile* as its member is the *hypernym* of the other synset with *cab*, *taxi* and *hack*. Such relation which is built on the synset level is categorized as synset relations.

Figure: example of synset relations

3.7.3 Sense Relations

Some relations in wordnet are also built on sense level, which can be further divided into two types, relations that link sense with another sense, and relations that link sense with another synset.

Note: In wordnet, synset relation and sense relation can both employ a particular relation type, such as domain topic.

Sense-Sense

Sense to sense relations emphasize the connections between different senses, especially when dealing with morphologically related words. For example, *behavioral* is the adjective to the noun *behavior*, which is known as in the *pertainym* relation with *behavior*, however, such relation doesn't exist between *behavioral* and *conduct*, which is a synonym of *behavior* and is in the same synset. Here *pertainym* is a sense-sense relation.

Figure: example of sense-sense relations

Sense-Synset

Sense-synset relations connect a particular sense with a synset. For example, *cursor* is a term in the *computer science* discipline, in wordnet, it is in the *has domain topic* relation with the *computer science* synset, but *pointer*, which is in the same synset with *cursor*, is not a term, thus has no such relation with *computer science* synset.

Figure: example of sense-synset relations

3.7.4 Other Information

A wordnet should be built in an appropriate form, two schemas are accepted:

- XML schema based on the Lexical Markup Framework (LMF)
- JSON-LD using the Lexicon Model for Ontologies

The structure of a wordnet should contain below info:

Definition

Definition is used to define senses and synsets in a wordnet, it is given in the language of the wordnet it came from.

Example

Example is used to clarify the senses and synsets in a wordnet, users can understand the definition more clearly with a given example.

Metadata

A wordnet has its own metadata, based on the Dublin Core, to state the basic info of it, below table lists all the items in the metadata of a wordnet:

contributor	Optional	str
coverage	Optional	str
creator	Optional	str
date	Optional	str
description	Optional	str
format	Optional	str
identifier	Optional	str
publisher	Optional	str
relation	Optional	str
rights	Optional	str
source	Optional	str
subject	Optional	str
title	Optional	str
type	Optional	str
status	Optional	str
note	Optional	str
confidence	Optional	float

3.8 Lemmatization and Normalization

Wn provides two methods for expanding queries: *lemmatization* and *normalization*. Wn also has a setting that allows *alternative forms* stored in the database to be included in queries.

See also:

The *wn.morphy* module is a basic English lemmatizer included with Wn.

3.8.1 Lemmatization

When querying a wordnet with wordforms from natural language text, it is important to be able to find entries for inflected forms as the database generally contains only lemmatic forms, or *lemmas* (or *lemmata*, if you prefer irregular plurals).

```
>>> import wn
>>> en = wn.Wordnet('oewn:2021')
>>> en.words('plurals')
[]
>>> en.words('plural')
[Word('oewn-plural-a'), Word('oewn-plural-n')]
```

Lemmas are sometimes called *citation forms* or *dictionary forms* as they are often used as the head words in dictionary entries. In Natural Language Processing (NLP), *lemmatization* is a technique where a possibly inflected word form is transformed to yield a lemma. In Wn, this concept is generalized somewhat to mean a transformation that yields a form matching wordforms stored in the database. For example, the English word *sparrows* is the plural inflection of *sparrow*, while the word *leaves* is ambiguous between the plural inflection of the nouns *leaf* and *leave* and the 3rd-person singular inflection of the verb *leave*.

For tasks where high-accuracy is needed, wrapping the wordnet queries with external tools that handle tokenization, lemmatization, and part-of-speech tagging will likely yield the best results as this method can make use of word context. That is, something like this:

```
for lemma, pos in fancy_shmancy_analysis(corpus):
    synsets = w.synsets(lemma, pos=pos)
```

For modest needs, however, Wn provides a way to integrate basic lemmatization directly into the queries.

Lemmatization in Wn works as follows: if a *wn.Wordnet* object is instantiated with a *lemmatizer* argument, then queries involving wordforms (e.g., *wn.Wordnet.words()*, *wn.Wordnet.senses()*, *wn.Wordnet.synsets()*) will first lemmatize the wordform and then check all resulting wordforms and parts of speech against the database as successive queries.

Lemmatization Functions

The *lemmatizer* argument of *wn.Wordnet* is a callable that takes two string arguments: (1) the original wordform, and (2) a part-of-speech or `None`. It returns a dictionary mapping parts-of-speech to sets of lemmatized wordforms. The signature is as follows:

```
lemmatizer(s: str, pos: Optional[str]) -> Dict[Optional[str], Set[str]]
```

The part-of-speech may be used by the function to determine which morphological rules to apply. If the given part-of-speech is `None`, then it is not specified and any rule may apply. A lemmatizer that only deinflects should not change any specified part-of-speech, but this is not a requirement, and a function could be provided that undoes derivational morphology (e.g., *democratic* → *democracy*).

Querying With Lemmatization

As the needs of lemmatization differs from one language to another, Wn does not provide a lemmatizer by default, and therefore it is unavailable to the convenience functions *wn.words()*, *wn.senses()*, and *wn.synsets()*. A lemmatizer can be added to a *wn.Wordnet* object. For example, using *wn.morphy*:

```
>>> import wn
>>> from wn.morphy import Morphy
>>> en = wn.Wordnet('oewn:2021', lemmatizer=Morphy())
>>> en.words('sparrows')
[Word('oewn-sparrow-n')]
>>> en.words('leaves')
[Word('oewn-leave-v'), Word('oewn-leaf-n'), Word('oewn-leave-n')]
```

Querying Without Lemmatization

When lemmatization is not used, inflected terms may not return any results:

```
>>> en = wn.Wordnet('oewn:2021')
>>> en.words('sparrows')
[]
```

Depending on the lexicon, there may be situations where results are returned for inflected lemmas, such as when the inflected form is lexicalized as its own entry:

```
>>> en.words('glasses')
[Word('oewn-glasses-n')]
```

Or if the lexicon lists the inflected form as an alternative form. For example, the English Wordnet lists irregular inflections as alternative forms:

```
>>> en.words('lemmata')
[Word('oewn-lemma-n')]
```

See below for excluding alternative forms from such queries.

3.8.2 Alternative Forms in the Database

A lexicon may include alternative forms in addition to lemmas for each word, and by default these are included in queries. What exactly is included as an alternative form depends on the lexicon. The English Wordnet, for example, adds irregular inflections (or "exceptional forms"), while the Japanese Wordnet includes the same word in multiple orthographies (original, hiragana, katakana, and two romanizations). For the English Wordnet, this means that you might get basic lemmatization for irregular forms only:

```
>>> en = wn.Wordnet('oewn:2021')
>>> en.words('learnt', pos='v')
[Word('oewn-learn-v')]
>>> en.words('learned', pos='v')
[]
```

If this is undesirable, the alternative forms can be excluded from queries with the *search_all_forms* parameter:

```
>>> en = wn.Wordnet('oewn:2021', search_all_forms=False)
>>> en.words('learnt', pos='v')
[]
>>> en.words('learned', pos='v')
[]
```

3.8.3 Normalization

While lemmatization deals with morphological variants of words, normalization handles minor orthographic variants. Normalized forms, however, may be invalid as wordforms in the target language, and as such they are only used behind the scenes for query expansion and not presented to users. For instance, a user might attempt to look up *résumé* in the English wordnet, but the wordnet only contains the form without diacritics: *resume*. With strict string matching, the entry would not be found using the wordform in the query. By normalizing the query word, the entry can be found. Similarly in the Spanish wordnet, *soñar* (to dream) and *sonar* (to ring) are two different words. A user who types *soñar* likely does not want to get results for *sonar*, but one who types *sonar* may be a non-Spanish speaker who is unaware of the missing diacritic or does not have an input method that allows them to type the diacritic, so this query would return both entries by matching against the normalized forms in the database. Wn handles all of these use cases.

When a lexicon is added to the database, potentially two wordforms are inserted for every one in the lexicon: the original wordform and a normalized form. When querying against the database, the original query string is first compared with the original wordforms and, if normalization is enabled, with the normalized forms in the database as well. If this first attempt yields no results and if normalization is enabled, the query string is normalized and tried again.

Normalization Functions

The normalized form is obtained from a *normalizer* function, passed as an argument to *wn.Wordnet*, that takes a single string argument and returns a string. That is, a function with the following signature:

```
normalizer(s: str) -> str
```

While custom *normalizer* functions could be used, in practice the choice is either the default normalizer or `None`. The default normalizer works by downcasing the string and performing NFKD normalization to remove diacritics. If the normalized form is the same as the original, only the original is inserted into the database.

Table 1: Examples of normalization

Original Form	Normalized Form
résumé	resume
soñar	sonar
San José	san jose

Querying With Normalization

By default, normalization is enabled when a *wn.Wordnet* is created. Enabling normalization does two things: it allows queries to check the original wordform in the query against the normalized forms in the database and, if no results are returned in the first step, it allows the queried wordform to be normalized as a back-off technique.

```
>>> en = wn.Wordnet('oewn:2021')
>>> en.words('résumé')
[Word('oewn-resume-n'), Word('oewn-resume-v')]
>>> es = wn.Wordnet('omw-es:1.4')
>>> es.words('soñar')
[Word('omw-es-soñar-v')]
>>> es.words('sonar')
[Word('omw-es-sonar-v'), Word('omw-es-soñar-v')]
```

Note: Users may supply a custom *normalizer* function to the *wn.Wordnet* object, but currently this is discouraged as the result is unlikely to match normalized forms in the database and there is not yet a way to customize the normalization of forms added to the database.

Querying Without Normalization

Normalization can be disabled by passing `None` as the argument of the *normalizer* parameter of *wn.Wordnet*. The queried wordform will not be checked against normalized forms in the database and neither will it be normalized as a back-off technique.

```
>>> en = wn.Wordnet('oewn:2021', normalizer=None)
>>> en.words('résumé')
[]
>>> es = wn.Wordnet('omw-es:1.4', normalizer=None)
>>> es.words('soñar')
[Word('omw-es-soñar-v')]
>>> es.words('sonar')
[Word('omw-es-sonar-v')]
```

Note: It is not possible to disable normalization for the convenience functions *wn.words()*, *wn.senses()*, and *wn.synsets()*.

3.9 Migrating from the NLTK

This guide is for users of the NLTK's `nltk.corpus.wordnet` module who are migrating to Wn. It is not guaranteed that Wn will produce the same results as the NLTK's module, but with some care its behavior can be very similar.

3.9.1 Overview

One important thing to note is that Wn will search all wordnets in the database by default where the NLTK would only search the English.

```
>>> from nltk.corpus import wordnet as nltk_wn
>>> nltk_wn.synsets('chat')                 # only English
>>> nltk_wn.synsets('chat', lang='fra')     # only French
>>> import wn
>>> wn.synsets('chat')                      # all wordnets
>>> wn.synsets('chat', lang='fr')           # only French
```

With Wn it helps to create a *wn.Wordnet* object to pre-filter the results by language or lexicon.

```
>>> en = wn.Wordnet('omw-en:1.4')
>>> en.synsets('chat')                    # only the OMW English Wordnet
```

3.9.2 Equivalent Operations

The following table lists equivalent API calls for the NLTK's wordnet module and Wn assuming the respective modules have been instantiated (in separate Python sessions) as follows:

NLTK:

```
>>> from nltk.corpus import wordnet as wn
>>> ss = wn.synsets("chat", pos="v")[0]
```

Wn:

```
>>> import wn
>>> en = wn.Wordnet('omw-en:1.4')
>>> ss = en.synsets("chat", pos="v")[0]
```

Primary Queries

NLTK	Wn
`wn.langs()`	`[lex.language for lex in wn.lexicons()]`
`wn.lemmas("chat")`	–
–	`en.words("chat")`
–	`en.senses("chat")`
`wn.synsets("chat")`	`en.synsets("chat")`
`wn.synsets("chat", pos="v")`	`en.synsets("chat", pos="v")`
`wn.all_synsets()`	`en.synsets()`
`wn.all_synsets(pos="v")`	`en.synsets(pos="v")`

Synsets – Basic

NLTK	Wn
`ss.lemmas()`	–
–	`ss.senses()`
–	`ss.words()`
`ss.lemmas_names()`	`ss.lemmas()`
`ss.definition()`	`ss.definition()`
`ss.examples()`	`ss.examples()`
`ss.pos()`	`ss.pos`

Synsets – Relations

NLTK	Wn
`ss.hypernyms()`	`ss.get_related("hypernym")`
`ss.instance_hypernyms()`	`ss.get_related("instance_hypernym")`
`ss.hypernyms() + ss.instance_hypernyms()`	`ss.hypernyms()`
`ss.hyponyms()`	`ss.get_related("hyponym")`
`ss.member_holonyms()`	`ss.get_related("holo_member")`
`ss.member_meronyms()`	`ss.get_related("mero_member")`
`ss.closure(lambda x:  x.hypernyms())`	`ss.closure("hypernym")`

Synsets – Taxonomic Structure

NLTK	Wn
`ss.min_depth()`	`ss.min_depth()`
`ss.max_depth()`	`ss.max_depth()`
`ss.hypernym_paths()`	`[list(reversed([ss] + p)) for p in ss.hypernym_paths()]`
`ss.common_hypernyms(ss)`	`ss.common_hypernyms(ss)`
`ss.lowest_common_hypernyms(ss)`	`ss.lowest_common_hypernyms(ss)`
`ss.shortest_path_distance(ss)`	`len(ss.shortest_path(ss))`

(these tables are incomplete)

3.10 wn

Wordnet Interface.

3.10.1 Project Management Functions

wn.**download**(*project_or_url*, *add=True*, *progress_handler=<class 'wn.util.ProgressBar'>*)

Download the resource specified by *project_or_url*.

First the URL of the resource is determined and then, depending on the parameters, the resource is downloaded and added to the database. The function then returns the path of the cached file.

If *project_or_url* starts with *'http://'* or *'https://'*, then it is taken to be the URL for the resource. Otherwise, *project_or_url* is taken as a *project specifier* and the URL is taken from a matching entry in Wn's project index. If no project matches the specifier, *wn.Error* is raised.

If the URL has been downloaded and cached before, the cached file is used. Otherwise the URL is retrieved and stored in the cache.

If the *add* paramter is `True` (default), the downloaded resource is added to the database.

```
>>> wn.download('ewn:2020')
Added ewn:2020 (English WordNet)
```

The *progress_handler* parameter takes a subclass of *wn.util.ProgressHandler*. An instance of the class will be created, used, and closed by this function.

Parameters

- **project_or_url** (*str*) –
- **add** (*bool*) –
- **progress_handler** (*Optional[Type[wn.util.ProgressHandler]]*) –

Return type pathlib.Path

wn.**add**(*source*, *progress_handler=<class 'wn.util.ProgressBar'>*)

Add the LMF file at *source* to the database.

The file at *source* may be gzip-compressed or plain text XML.

```
>>> wn.add('english-wordnet-2020.xml')
Added ewn:2020 (English WordNet)
```

The *progress_handler* parameter takes a subclass of *wn.util.ProgressHandler*. An instance of the class will be created, used, and closed by this function.

Parameters

- **source** (*Union[str, pathlib.Path]*) –
- **progress_handler** (*Optional[Type[wn.util.ProgressHandler]]*) –

Return type None

wn.**remove**(*lexicon*, *progress_handler=<class 'wn.util.ProgressBar'>*)

Remove lexicon(s) from the database.

The *lexicon* argument is a *lexicon specifier*. Note that this removes a lexicon and not a project, so the lexicons of projects containing multiple lexicons will need to be removed individually or, if applicable, a star specifier.

The *progress_handler* parameter takes a subclass of *wn.util.ProgressHandler*. An instance of the class will be created, used, and closed by this function.

```
>>> wn.remove('ewn:2019')  # removes a single lexicon
>>> wn.remove('*:1.3+omw')  # removes all lexicons with version 1.3+omw
```

Parameters

- **lexicon** (*str*) –
- **progress_handler** (*Optional[Type[wn.util.ProgressHandler]]*) –

Return type None

wn.**export**(*lexicons*, *destination*, *version='1.0'*)

Export lexicons from the database to a WN-LMF file.

More than one lexicon may be exported in the same file, subject to these conditions:

- identifiers on wordnet entities must be unique in all lexicons
- lexicons extensions may not be exported with their dependents

```
>>> w = wn.Wordnet(lexicon='cmnwn zsmwn')
>>> wn.export(w.lexicons(), 'cmn-zsm.xml')
```

Parameters

- **lexicons** (*Sequence[wn.Lexicon]*) – sequence of *wn.Lexicon* objects
- **destination** (*Union[str, pathlib.Path]*) – path to the destination file
- **version** (*str*) – LMF version string

Return type None

wn.**projects**()

Return the list of indexed projects.

This returns the same dictionaries of information as *wn.config.get_project_info*, but for all indexed projects.

Example

```
>>> infos = wn.projects()
>>> len(infos)
36
>>> infos[0]['label']
'Open English WordNet'
```

Return type *List*[*Dict*]

3.10.2 Wordnet Query Functions

wn.**word**(*id*, *, *lexicon=None*, *lang=None*)

Return the word with *id* in *lexicon*.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The *id* argument is then passed to the *Wordnet.word()* method.

```
>>> wn.word('ewn-cell-n')
Word('ewn-cell-n')
```

Parameters

- **id** (*str*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *wn.Word*

wn.**words**(*form=None*, *pos=None*, *, *lexicon=None*, *lang=None*)

Return the list of matching words.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The remaining arguments are passed to the *Wordnet.words()* method.

```
>>> len(wn.words())
282902
>>> len(wn.words(pos='v'))
34592
>>> wn.words(form="scurry")
[Word('ewn-scurry-n'), Word('ewn-scurry-v')]
```

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *List*[*wn.Word*]

wn.**sense**(*id*, *, *lexicon=None*, *lang=None*)

Return the sense with *id* in *lexicon*.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The *id* argument is then passed to the *Wordnet.sense()* method.

```
>>> wn.sense('ewn-flutter-v-01903884-02')
Sense('ewn-flutter-v-01903884-02')
```

Parameters

- **id** (*str*) –

- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *wn.Sense*

wn.**senses**(*form=None*, *pos=None*, ***, *lexicon=None*, *lang=None*)

Return the list of matching senses.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The remaining arguments are passed to the *Wordnet.senses()* method.

```
>>> len(wn.senses('twig'))
3
>>> wn.senses('twig', pos='n')
[Sense('ewn-twig-n-13184889-02')]
```

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *List*[*wn.Sense*]

wn.**synset**(*id*, ***, *lexicon=None*, *lang=None*)

Return the synset with *id* in *lexicon*.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The *id* argument is then passed to the *Wordnet.synset()* method.

```
>>> wn.synset('ewn-03311152-n')
Synset('ewn-03311152-n')
```

Parameters

- **id** (*str*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *wn.Synset*

wn.**synsets**(*form=None*, *pos=None*, *ili=None*, ***, *lexicon=None*, *lang=None*)

Return the list of matching synsets.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The remaining arguments are passed to the *Wordnet.synsets()* method.

```
>>> len(wn.synsets('couch'))
4
>>> wn.synsets('couch', pos='v')
[Synset('ewn-00983308-v')]
```

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –
- **ili** (*Optional[str]*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *List*[*wn.Synset*]

wn.**ili**(*id*, ***, *lexicon=None*, *lang=None*)

Return the interlingual index with *id*.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The *id* argument is then passed to the *Wordnet.ili()* method.

```
>>> wn.ili(id='i1234')
ILI('i1234')
>>> wn.ili(id='i1234').status
'presupposed'
```

Parameters

- **id** (*str*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *wn.ILI*

wn.**ilis**(*status=None*, ***, *lexicon=None*, *lang=None*)

Return the list of matching interlingual indices.

This will create a *Wordnet* object using the *lang* and *lexicon* arguments. The remaining arguments are passed to the *Wordnet.ilis()* method.

```
>>> len(wn.ilis())
120071
>>> len(wn.ilis(status='proposed'))
2573
>>> wn.ilis(status='proposed')[-1].definition()
'the neutrino associated with the tau lepton.'
>>> len(wn.ilis(lang='de'))
13818
```

Parameters

- **status** (*Optional[str]*) –
- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *List*[*wn.ILI*]

wn.**lexicons**(*, *lexicon=None*, *lang=None*)

Return the lexicons matching a language or lexicon specifier.

Example

```
>>> wn.lexicons(lang='en')
[<Lexicon ewn:2020 [en]>, <Lexicon pwn:3.0 [en]>]
```

Parameters

- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –

Return type *List*[*wn.Lexicon*]

3.10.3 The Wordnet Class

class wn.**Wordnet**(*lexicon=None*, *, *lang=None*, *expand=None*, *normalizer=<function normalize_form>*, *lemmatizer=None*, *search_all_forms=True*)

Class for interacting with wordnet data.

A wordnet object acts essentially as a filter by first selecting matching lexicons and then searching only within those lexicons for later queries. On instantiation, a *lang* argument is a BCP 47 language code that restricts the selected lexicons to those whose language matches the given code. A *lexicon* argument is a space-separated list of lexicon specifiers that more directly selects lexicons by their ID and version; this is preferable when there are multiple lexicons in the same language or multiple version with the same ID.

Some wordnets were created by translating the words from a larger wordnet, namely the Princeton WordNet, and then relying on the larger wordnet for structural relations. An *expand* argument is a second space-separated list of lexicon specifiers which are used for traversing relations, but not as the results of queries. Setting *expand* to an empty string (`expand=''`) disables expand lexicons.

The *normalizer* argument takes a callable that normalizes word forms in order to expand the search. The default function downcases the word and removes diacritics via NFKD normalization so that, for example, searching for *san josé* in the English WordNet will find the entry for *San Jose*. Setting *normalizer* to `None` disables normalization and forces exact-match searching.

The *lemmatizer* argument may be `None`, which is the default and disables lemmatizer-based query expansion, or a callable that takes a word form and optional part of speech and returns base forms of the original word. To support lemmatizers that use the wordnet for instantiation, such as *wn.morphy*, the lemmatizer may be assigned to the *lemmatizer* attribute after creation.

If the *search_all_forms* argument is `True` (the default), searches of word forms consider all forms in the lexicon; if `False`, only lemmas are searched. Non-lemma forms may include, depending on the lexicon, morphological exceptions, alternate scripts or spellings, etc.

Parameters

- **lexicon** (*Optional[str]*) –
- **lang** (*Optional[str]*) –
- **expand** (*Optional[str]*) –
- **normalizer** (*Optional[Callable[[str], str]]*) –

- **lemmatizer** (*Optional[Callable[[str, Optional[str]], Dict[Optional[str], Set[str]]]]*) –
- **search_all_forms** (*bool*) –

lemmatizer

A lemmatization function or None.

word(*id*)

Return the first word in this wordnet with identifier *id*.

Parameters **id** (*str*) –

Return type *wn.Word*

words(*form=None*, *pos=None*)

Return the list of matching words in this wordnet.

Without any arguments, this function returns all words in the wordnet's selected lexicons. A *form* argument restricts the words to those matching the given word form, and *pos* restricts words by their part of speech.

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –

Return type *List*[*wn.Word*]

sense(*id*)

Return the first sense in this wordnet with identifier *id*.

Parameters **id** (*str*) –

Return type *wn.Sense*

senses(*form=None*, *pos=None*)

Return the list of matching senses in this wordnet.

Without any arguments, this function returns all senses in the wordnet's selected lexicons. A *form* argument restricts the senses to those whose word matches the given word form, and *pos* restricts senses by their word's part of speech.

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –

Return type *List*[*wn.Sense*]

synset(*id*)

Return the first synset in this wordnet with identifier *id*.

Parameters **id** (*str*) –

Return type *wn.Synset*

synsets(*form=None*, *pos=None*, *ili=None*)

Return the list of matching synsets in this wordnet.

Without any arguments, this function returns all synsets in the wordnet's selected lexicons. A *form* argument restricts synsets to those whose member words match the given word form. A *pos* argument restricts synsets to those with the given part of speech. An *ili* argument restricts synsets to those with the given interlingual index; generally this should select a unique synset within a single lexicon.

Parameters

- **form** (*Optional[str]*) –
- **pos** (*Optional[str]*) –
- **ili** (*Optional[str]*) –

Return type *List*[*wn.Synset*]

ili(*id*)

Return the first ILI in this wordnet with identifer *id*.

Parameters **id** (*str*) –

Return type *wn.ILI*

ilis(*status=None*)

Return the list of ILIs in this wordnet.

If *status* is given, only return ILIs with a matching status.

Parameters **status** (*Optional[str]*) –

Return type *List*[*wn.ILI*]

lexicons()

Return the list of lexicons covered by this wordnet.

Return type *List*[*wn.Lexicon*]

expanded_lexicons()

Return the list of expand lexicons for this wordnet.

Return type *List*[*wn.Lexicon*]

describe()

Return a formatted string describing the lexicons in this wordnet.

Example

```
>>> oewn = wn.Wordnet('oewn:2021')
>>> print(oewn.describe())
Primary lexicons:
  oewn:2021
    Label  : Open English WordNet
    URL    : https://github.com/globalwordnet/english-wordnet
    License: https://creativecommons.org/licenses/by/4.0/
    Words  : 163161 (a: 8386, n: 123456, r: 4481, s: 15231, v: 11607)
    Senses : 211865
    Synsets: 120039 (a: 7494, n: 84349, r: 3623, s: 10727, v: 13846)
    ILIs   : 120039
```

Return type str

3.10.4 The Word Class

class wn.**Word**(*id*, *pos*, *forms*, *_lexid=0*, *_id=0*, *_wordnet=None*)

A class for words (also called lexical entries) in a wordnet.

Parameters

- **id** (*str*) –
- **pos** (*str*) –
- **forms** (*List[Tuple[str, Optional[str], Optional[str], int]]*) –
- **_lexid** (*int*) –
- **_id** (*int*) –
- **_wordnet** (*Optional[Wordnet]*) –

id

The identifier used within a lexicon.

pos

The part of speech of the Word.

lemma()

Return the canonical form of the word.

Example

```
>>> wn.words('wolves')[0].lemma()
'wolf'
```

Return type *wn.Form*

forms()

Return the list of all encoded forms of the word.

Example

```
>>> wn.words('wolf')[0].forms()
['wolf', 'wolves']
```

Return type *List*[*wn.Form*]

senses()

Return the list of senses of the word.

Example

```
>>> wn.words('zygoma')[0].senses()
[Sense('ewn-zygoma-n-05292350-01')]
```

Return type *List*[*wn.Sense*]

synsets()

Return the list of synsets of the word.

Example

```
>>> wn.words('addendum')[0].synsets()
[Synset('ewn-06411274-n')]
```

Return type *List*[*wn.Synset*]

metadata()

Return the word's metadata.

Return type *Dict*[str, *Any*]

derived_words()

Return the list of words linked through derivations on the senses.

Example

```
>>> wn.words('magical')[0].derived_words()
[Word('ewn-magic-n'), Word('ewn-magic-n')]
```

Return type *List*[*wn.Word*]

translate(*lexicon=None*, ***, *lang=None*)

Return a mapping of word senses to lists of translated words.

Parameters

- **lexicon** (*Optional[str]*) – if specified, translate to words in the target lexicon(s)
- **lang** (*Optional[str]*) – if specified, translate to words with the language code

Return type *Dict*[*wn.Sense*, *List*[*wn.Word*]]

Example

```
>>> w = wn.words('water bottle', pos='n')[0]
>>> for sense, words in w.translate(lang='ja').items():
...     print(sense, [jw.lemma() for jw in words])
...
Sense('ewn-water_bottle-n-04564934-01') ['']
```

The Form Class

class wn.**Form**

The return value of *Word.lemma()* and the members of the list returned by *Word.forms()* are *Form* objects. These are a basic subclass of Python's `str` class with an additional attribute, *script*, and methods *pronunciations()* and *tags()*. Form objects without any specified script behave exactly as a regular string (they are equal and hash to the same value), but if two Form objects are compared and they have different script values, then they are unequal and hash differently, even if the string itself is identical. When comparing a Form object to a regular string, the script value is ignored.

```
>>> inu = wn.words('', lexicon='wnja')[0]
>>> inu.forms()[3]
''
>>> inu.forms()[3].script
'hira'
```

The *script* is often unspecified (i.e., `None`) and this carries the implicit meaning that the form uses the canonical script for the word's language or wordnet, whatever it may be.

script

The script of the word form. This should be an ISO 15924 code, or `None`.

pronunciations()

Return the list of *Pronunciation* objects.

tags()

Return the list of *Tag* objects.

The Pronunciation Class

class wn.**Pronunciation**(*value*, *variety=None*, *notation=None*, *phonemic=True*, *audio=None*)

A class for word form pronunciations.

Parameters

- **value** (*str*) –
- **variety** (*Optional[str]*) –
- **notation** (*Optional[str]*) –
- **phonemic** (*bool*) –
- **audio** (*Optional[str]*) –

value

The encoded pronunciation.

variety

The language variety this pronunciation belongs to.

notation

The notation used to encode the pronunciation. For example: the International Phonetic Alphabet (IPA).

phonemic

`True` when the encoded pronunciation is a generalized phonemic description, or `False` for more precise phonetic transcriptions.

audio

A URI to an associated audio file.

The Tag Class

class wn.**Tag**(*tag*, *category*)

A general-purpose tag class for word forms.

Parameters

- **tag** (*str*) –
- **category** (*str*) –

tag

The text value of the tag.

category

The category, or kind, of the tag.

3.10.5 The Sense Class

class wn.**Sense**(*id*, *entry_id*, *synset_id*, *_lexid=0*, *_id=0*, *_wordnet=None*)

Class for modeling wordnet senses.

Parameters

- **id** (*str*) –
- **entry_id** (*str*) –
- **synset_id** (*str*) –
- **_lexid** (*int*) –
- **_id** (*int*) –
- **_wordnet** (*Optional[Wordnet]*) –

id

The identifier used within a lexicon.

word()

Return the word of the sense.

Example

```
>>> wn.senses('spigot')[0].word()
Word('pwn-spigot-n')
```

Return type *wn.Word*

synset()

Return the synset of the sense.

Example

```
>>> wn.senses('spigot')[0].synset()
Synset('pwn-03325088-n')
```

Return type *wn.Synset*

examples()

Return the list of examples for the sense.

Return type *List*[str]

lexicalized()

Return True if the sense is lexicalized.

Return type bool

adjposition()

Return the adjective position of the sense.

Values include `"a"` (attributive), `"p"` (predicative), and `"ip"` (immediate postnominal). Note that this is only relevant for adjectival senses. Senses for other parts of speech, or for adjectives that are not annotated with this feature, will return `None`.

Return type *Optional*[str]

frames()

Return the list of subcategorization frames for the sense.

Return type *List*[str]

counts()

Return the corpus counts stored for this sense.

Return type *List*[*wn.Count*]

metadata()

Return the sense's metadata.

Return type *Dict*[str, *Any*]

relations(**args*)

Return a mapping of relation names to lists of senses.

One or more relation names may be given as positional arguments to restrict the relations returned. If no such arguments are given, all relations starting from the sense are returned.

See *get_related()* for getting a flat list of related senses.

Parameters **args** (*str*) –

Return type *Dict*[str, *List*[*wn.Sense*]]

get_related(**args*)

Return a list of related senses.

One or more relation types should be passed as arguments which determine the kind of relations returned.

Example

```
>>> physics = wn.senses('physics', lexicon='ewn')[0]
>>> for sense in physics.get_related('has_domain_topic'):
...     print(sense.word().lemma())
...
coherent
chaotic
incoherent
```

Parameters **args** (*str*) –

Return type *List*[*wn.Sense*]

get_related_synsets(**args*)

Return a list of related synsets.

Parameters **args** (*str*) –

Return type *List*[*wn.Synset*]

closure(**args*)

Parameters

- **self** (*wn._core.T*) –
- **args** (*str*) –

Return type *Iterator*[wn._core.T]

relation_paths(**args*, *end=None*)

Parameters

- **self** (*wn._core.T*) –
- **args** (*str*) –
- **end** (*Optional[wn._core.T]*) –

Return type *Iterator*[*List*[wn._core.T]]

translate(*lexicon=None*, ***, *lang=None*)

Return a list of translated senses.

Parameters

- **lexicon** (*Optional[str]*) – if specified, translate to senses in the target lexicon(s)
- **lang** (*Optional[str]*) – if specified, translate to senses with the language code

Return type *List*[*wn.Sense*]

Example

```
>>> en = wn.senses('petiole', lang='en')[0]
>>> pt = en.translate(lang='pt')[0]
>>> pt.word().lemma()
'pecíolo'
```

The Count Class

class wn.**Count**(*value*, *_id=0*)

A count of sense occurrences in some corpus.

Some wordnets store computed counts of senses across some corpus or corpora. This class models those counts. It is a subtype of `int` with one additional method, *metadata()*, which may be used to give information about the source of the count (if provided by the wordnet).

Parameters **_id** (*int*) –

metadata()

Return the count's metadata.

Return type *Dict*[str, *Any*]

3.10.6 The Synset Class

class wn.**Synset**(*id*, *pos*, *ili=None*, *_lexid=0*, *_id=0*, *_wordnet=None*)

Class for modeling wordnet synsets.

Parameters

- **id** (*str*) –
- **pos** (*str*) –
- **ili** (*Optional[str]*) –
- **_lexid** (*int*) –
- **_id** (*int*) –
- **_wordnet** (*Optional[Wordnet]*) –

id

The identifier used within a lexicon.

pos

The part of speech of the Synset.

ili

The interlingual index of the Synset.

definition()

Return the first definition found for the synset.

Example

```
>>> wn.synsets('cartwheel', pos='n')[0].definition()
'a wheel that has wooden spokes and a metal rim'
```

Return type *Optional*[str]

examples()

Return the list of examples for the synset.

Example

```
>>> wn.synsets('orbital', pos='a')[0].examples()
['"orbital revolution"', '"orbital velocity"']
```

Return type *List*[str]

senses()

Return the list of sense members of the synset.

Example

```
>>> wn.synsets('umbrella', pos='n')[0].senses()
[Sense('ewn-umbrella-n-04514450-01')]
```

Return type *List*[*wn.Sense*]

lexicalized()

Return True if the synset is lexicalized.

Return type bool

lexfile()

Return the lexicographer file name for this synset, if any.

Return type *Optional*[str]

metadata()

Return the synset's metadata.

Return type *Dict*[str, *Any*]

words()

Return the list of words linked by the synset's senses.

Example

```
>>> wn.synsets('exclusive', pos='n')[0].words()
[Word('ewn-scoop-n'), Word('ewn-exclusive-n')]
```

Return type *List*[*wn.Word*]

lemmas()

Return the list of lemmas of words for the synset.

Example

```
>>> wn.synsets('exclusive', pos='n')[0].words()
['scoop', 'exclusive']
```

Return type *List*[*wn.Form*]

hypernyms()

Return the list of synsets related by any hypernym relation.

Both the `hypernym` and `instance_hypernym` relations are traversed.

Return type *List*[*wn.Synset*]

hyponyms()

Return the list of synsets related by any hyponym relation.

Both the `hyponym` and `instance_hyponym` relations are traversed.

Return type *List*[*wn.Synset*]

holonyms()

Return the list of synsets related by any holonym relation.

Any of the following relations are traversed: `holonym`, `holo_location`, `holo_member`, `holo_part`, `holo_portion`, `holo_substance`.

Return type *List*[*wn.Synset*]

meronyms()

Return the list of synsets related by any meronym relation.

Any of the following relations are traversed: `meronym`, `mero_location`, `mero_member`, `mero_part`, `mero_portion`, `mero_substance`.

Return type *List*[*wn.Synset*]

relations(**args*)

Return a mapping of relation names to lists of synsets.

One or more relation names may be given as positional arguments to restrict the relations returned. If no such arguments are given, all relations starting from the synset are returned.

See *get_related()* for getting a flat list of related synsets.

Example

```
>>> button_rels = wn.synsets('button')[0].relations()
>>> for relname, sslist in button_rels.items():
...     print(relname, [ss.lemmas() for ss in sslist])
...
hypernym [['fixing', 'holdfast', 'fastener', 'fastening']]
hyponym [['coat button'], ['shirt button']]
```

Parameters **args** (*str*) –

Return type *Dict*[str, *List*[*wn.Synset*]]

get_related(**args*)

Return the list of related synsets.

One or more relation names may be given as positional arguments to restrict the relations returned. If no such arguments are given, all relations starting from the synset are returned.

This method does not preserve the relation names that lead to the related synsets. For a mapping of relation names to related synsets, see *relations()*.

Example

```
>>> fulcrum = wn.synsets('fulcrum')[0]
>>> [ss.lemmas() for ss in fulcrum.get_related()]
[['pin', 'pivot'], ['lever']]
```

Parameters **args** (*str*) –

Return type *List*[*wn.Synset*]

closure(**args*)

Parameters

- **self** (*wn._core.T*) –
- **args** (*str*) –

Return type *Iterator*[wn._core.T]

relation_paths(**args*, *end=None*)

Parameters

- **self** (*wn._core.T*) –
- **args** (*str*) –
- **end** (*Optional[wn._core.T]*) –

Return type *Iterator*[*List*[wn._core.T]]

translate(*lexicon=None*, ***, *lang=None*)

Return a list of translated synsets.

Parameters

- **lexicon** (*Optional[str]*) – if specified, translate to synsets in the target lexicon(s)
- **lang** (*Optional[str]*) – if specified, translate to synsets with the language code

Return type *List*[*wn.Synset*]

Example

```
>>> es = wn.synsets('araña', lang='es')[0]
>>> en = es.translate(lexicon='ewn')[0]
>>> en.lemmas()
['spider']
```

hypernym_paths(*simulate_root=False*)

Shortcut for *wn.taxonomy.hypernym_paths()*.

min_depth(*simulate_root=False*)

Shortcut for *wn.taxonomy.min_depth()*.

max_depth(*simulate_root=False*)

Shortcut for *wn.taxonomy.max_depth()*.

shortest_path(*other*, *simulate_root=False*)

Shortcut for *wn.taxonomy.shortest_path()*.

common_hypernyms(*other*, *simulate_root=False*)

Shortcut for *wn.taxonomy.common_hypernyms()*.

lowest_common_hypernyms(*other*, *simulate_root=False*)

Shortcut for *wn.taxonomy.lowest_common_hypernyms()*.

3.10.7 The ILI Class

class wn.**ILI**(*id*, *status*, *definition=None*, *_id=0*)

A class for interlingual indices.

Parameters

- **id** (*Optional[str]*) –
- **status** (*str*) –
- **definition** (*Optional[str]*) –
- **_id** (*int*) –

id

The interlingual index identifier. Unlike `id` attributes for *Word*, *Sense*, and *Synset*, ILI identifers may be `None` (see the *proposed* `status`).

status

The known status of the interlingual index. Loading an interlingual index into the database provides the following explicit, authoritative status values:

- `active` – the ILI is in use
- `provisional` – the ILI is being staged for permanent inclusion

- `deprecated` – the ILI is, or should be, no longer in use

Without an interlingual index loaded, ILIs present in loaded lexicons get an implicit, temporary status from the following:

- `presupposed` – a synset uses the ILI, assuming it exists in an ILI file
- `proposed` – a synset introduces a concept not yet in an ILI and is suggesting that one should be added for it in the future

`definition()`

Return type *Optional*[str]

`metadata()`

Return the ILI's metadata.

Return type *Dict*[str, *Any*]

3.10.8 The Lexicon Class

`class wn.Lexicon(`*id*, *label*, *language*, *email*, *license*, *version*, *url=None*, *citation=None*, *logo=None*, *_id=0*`)`

A class representing a wordnet lexicon.

Parameters

- **id** (*str*) –
- **label** (*str*) –
- **language** (*str*) –
- **email** (*str*) –
- **license** (*str*) –
- **version** (*str*) –
- **url** (*Optional[str]*) –
- **citation** (*Optional[str]*) –
- **logo** (*Optional[str]*) –
- **_id** (*int*) –

`id`

The lexicon's identifier.

`label`

The full name of lexicon.

`language`

The BCP 47 language code of lexicon.

`email`

The email address of the wordnet maintainer.

`license`

The URL or name of the wordnet's license.

version

The version string of the resource.

url

The project URL of the wordnet.

citation

The canonical citation for the project.

logo

A URL or path to a project logo.

metadata()

Return the lexicon's metadata.

Return type *Dict*[str, *Any*]

specifier()

Return the *id:version* lexicon specifier.

Return type str

modified()

Return True if the lexicon has local modifications.

Return type bool

requires()

Return the lexicon dependencies.

Return type *Dict*[str, *Optional*[*wn.Lexicon*]]

extends()

Return the lexicon this lexicon extends, if any.

If this lexicon is not an extension, return None.

Return type *Optional*[*wn.Lexicon*]

extensions(*depth=1*)

Return the list of lexicons extending this one.

By default, only direct extensions are included. This is controlled by the *depth* parameter, which if you view extensions as children in a tree where the current lexicon is the root, *depth=1* are the immediate extensions. Increasing this number gets extensions of extensions, or setting it to a negative number gets all "descendant" extensions.

Parameters **depth** (*int*) –

Return type *List*[*wn.Lexicon*]

describe(*full=True*)

Return a formatted string describing the lexicon.

The *full* argument (default: `True`) may be set to `False` to omit word and sense counts.

Also see: *Wordnet.describe()*

Parameters **full** (*bool*) –

Return type str

3.10.9 The wn.config Object

Wn's data storage and retrieval can be configured through the *wn.config* object.

See also:

Installation and Configuration describes how to configure Wn using the *wn.config* instance.

wn.**config = <wn._config.WNConfig object>**

It is an instance of the *WNConfig* class, which is defined in a non-public module and is not meant to be instantiated directly. Configuration should occur through the single *wn.config* instance.

class wn._config.**WNConfig**

data_directory

The file system directory where Wn's data is stored.

database_path

The path to the database file.

allow_multithreading

If set to `True`, the database connection may be shared across threads. In this case, it is the user's responsibility to ensure that multiple threads don't try to write to the database at the same time. The default is `False`.

downloads_directory

The file system directory where downloads are cached.

add_project(*id*, *type='wordnet'*, *label=None*, *language=None*, *license=None*, *error=None*)

Add a new wordnet project to the index.

Parameters

- **id** (*str*) – short identifier of the project
- **type** (*str*) – project type (default 'wordnet')
- **label** (*Optional[str]*) – full name of the project
- **language** (*Optional[str]*) – BCP 47 language code of the resource
- **license** (*Optional[str]*) – link or name of the project's default license
- **error** (*Optional[str]*) – if set, the error message to use when the project is accessed

Return type None

add_project_version(*id*, *version*, *url=None*, *error=None*, *license=None*)

Add a new resource version for a project.

Exactly one of *url* or *error* must be specified.

Parameters

- **id** (*str*) – short identifier of the project
- **version** (*str*) – version string of the resource
- **url** (*Optional[str]*) – space-separated list of web addresses for the resource
- **license** (*Optional[str]*) – link or name of the resource's license; if not given, the project's default license will be used.
- **error** (*Optional[str]*) – if set, the error message to use when the project is accessed

Return type None

`get_project_info`(*arg*)

Return information about an indexed project version.

If the project has been downloaded and cached, the `"cache"` key will point to the path of the cached file, otherwise its value is `None`.

Parameters **arg** (*str*) – a project specifier

Return type *Dict*

Example

```
>>> info = wn.config.get_project_info('oewn:2021')
>>> info['label']
'Open English WordNet'
```

`get_cache_path`(*url*)

Return the path for caching *url*.

Note that in general this is just a path operation and does not signify that the file exists in the file system.

Parameters **url** (*str*) –

Return type pathlib.Path

`update`(*data*)

Update the configuration with items in *data*.

Items are only inserted or replaced, not deleted. If a project index is provided in the `"index"` key, then either the project must not already be indexed or any project fields (label, language, or license) that are specified must be equal to the indexed project.

Parameters **data** (*dict*) –

Return type None

`load_index`(*path*)

Load and update with the project index at *path*.

The project index is a TOML file containing project and version information. For example:

```
[ewn]
  label = "Open English WordNet"
  language = "en"
  license = "https://creativecommons.org/licenses/by/4.0/"
  [ewn.versions.2019]
    url = "https://en-word.net/static/english-wordnet-2019.xml.gz"
  [ewn.versions.2020]
    url = "https://en-word.net/static/english-wordnet-2020.xml.gz"
```

Parameters **path** (*Union[str, pathlib.Path]*) –

Return type None

3.10.10 Exceptions

exception wn.**Error**

Generic error class for invalid wordnet operations.

exception wn.**DatabaseError**

Error class for issues with the database.

exception wn.**WnWarning**

Generic warning class for dubious worndet operations.

3.11 wn.constants

Constants and literals used in wordnets.

3.11.1 Synset Relations

wn.constants.**SYNSET_RELATIONS**

- agent
- also
- attribute
- be_in_state
- causes
- classified_by
- classifies
- co_agent_instrument
- co_agent_patient
- co_agent_result
- co_instrument_agent
- co_instrument_patient
- co_instrument_result
- co_patient_agent
- co_patient_instrument
- co_result_agent
- co_result_instrument
- co_role
- direction
- domain_region
- domain_topic
- exemplifies

- entails
- eq_synonym
- has_domain_region
- has_domain_topic
- is_exemplified_by
- holo_location
- holo_member
- holo_part
- holo_portion
- holo_substance
- holonym
- hypernym
- hyponym
- in_manner
- instance_hypernym
- instance_hyponym
- instrument
- involved
- involved_agent
- involved_direction
- involved_instrument
- involved_location
- involved_patient
- involved_result
- involved_source_direction
- involved_target_direction
- is_caused_by
- is_entailed_by
- location
- manner_of
- mero_location
- mero_member
- mero_part
- mero_portion
- mero_substance
- meronym

- similar
- other
- patient
- restricted_by
- restricts
- result
- role
- source_direction
- state_of
- target_direction
- subevent
- is_subevent_of
- antonym
- feminine
- has_feminine
- masculine
- has_masculine
- young
- has_young
- diminutive
- has_diminutive
- augmentative
- has_augmentative
- anto_gradable
- anto_simple
- anto_converse
- ir_synonym

3.11.2 Sense Relations

wn.constants.**SENSE_RELATIONS**

- antonym
- also
- participle
- pertainym
- derivation
- domain_topic

- has_domain_topic
- domain_region
- has_domain_region
- exemplifies
- is_exemplified_by
- similar
- other
- feminine
- has_feminine
- masculine
- has_masculine
- young
- has_young
- diminutive
- has_diminutive
- augmentative
- has_augmentative
- anto_gradable
- anto_simple
- anto_converse
- simple_aspect_ip
- secondary_aspect_ip
- simple_aspect_pi
- secondary_aspect_pi

wn.constants.**SENSE_SYNSET_RELATIONS**

- domain_topic
- domain_region
- exemplifies
- other

wn.constants.**REVERSE_RELATIONS**

```
{
    'hypernym': 'hyponym',
    'hyponym': 'hypernym',
    'instance_hypernym': 'instance_hyponym',
    'instance_hyponym': 'instance_hypernym',
    'antonym': 'antonym',
    'eq_synonym': 'eq_synonym',
```

(continues on next page)

(continued from previous page)

```
    'similar': 'similar',
    'meronym': 'holonym',
    'holonym': 'meronym',
    'mero_location': 'holo_location',
    'holo_location': 'mero_location',
    'mero_member': 'holo_member',
    'holo_member': 'mero_member',
    'mero_part': 'holo_part',
    'holo_part': 'mero_part',
    'mero_portion': 'holo_portion',
    'holo_portion': 'mero_portion',
    'mero_substance': 'holo_substance',
    'holo_substance': 'mero_substance',
    'also': 'also',
    'state_of': 'be_in_state',
    'be_in_state': 'state_of',
    'causes': 'is_caused_by',
    'is_caused_by': 'causes',
    'subevent': 'is_subevent_of',
    'is_subevent_of': 'subevent',
    'manner_of': 'in_manner',
    'in_manner': 'manner_of',
    'attribute': 'attribute',
    'restricts': 'restricted_by',
    'restricted_by': 'restricts',
    'classifies': 'classified_by',
    'classified_by': 'classifies',
    'entails': 'is_entailed_by',
    'is_entailed_by': 'entails',
    'domain_topic': 'has_domain_topic',
    'has_domain_topic': 'domain_topic',
    'domain_region': 'has_domain_region',
    'has_domain_region': 'domain_region',
    'exemplifies': 'is_exemplified_by',
    'is_exemplified_by': 'exemplifies',
    'role': 'involved',
    'involved': 'role',
    'agent': 'involved_agent',
    'involved_agent': 'agent',
    'patient': 'involved_patient',
    'involved_patient': 'patient',
    'result': 'involved_result',
    'involved_result': 'result',
    'instrument': 'involved_instrument',
    'involved_instrument': 'instrument',
    'location': 'involved_location',
    'involved_location': 'location',
    'direction': 'involved_direction',
    'involved_direction': 'direction',
    'target_direction': 'involved_target_direction',
    'involved_target_direction': 'target_direction',
    'source_direction': 'involved_source_direction',
```

(continues on next page)

(continued from previous page)

```
    'involved_source_direction': 'source_direction',
    'co_role': 'co_role',
    'co_agent_patient': 'co_patient_agent',
    'co_patient_agent': 'co_agent_patient',
    'co_agent_instrument': 'co_instrument_agent',
    'co_instrument_agent': 'co_agent_instrument',
    'co_agent_result': 'co_result_agent',
    'co_result_agent': 'co_agent_result',
    'co_patient_instrument': 'co_instrument_patient',
    'co_instrument_patient': 'co_patient_instrument',
    'co_result_instrument': 'co_instrument_result',
    'co_instrument_result': 'co_result_instrument',
    'pertainym': 'pertainym',
    'derivation': 'derivation',
    'simple_aspect_ip': 'simple_aspect_pi',
    'simple_aspect_pi': 'simple_aspect_ip',
    'secondary_aspect_ip': 'secondary_aspect_pi',
    'secondary_aspect_pi': 'secondary_aspect_ip',
    'feminine': 'has_feminine',
    'has_feminine': 'feminine',
    'masculine': 'has_masculine',
    'has_masculine': 'masculine',
    'young': 'has_young',
    'has_young': 'young',
    'diminutive': 'has_diminutive',
    'has_diminutive': 'diminutive',
    'augmentative': 'has_augmentative',
    'has_augmentative': 'augmentative',
    'anto_gradable': 'anto_gradable',
    'anto_simple': 'anto_simple',
    'anto_converse': 'anto_converse',
    'ir_synonym': 'ir_synonym',
}
```

3.11.3 Parts of Speech

wn.constants.**PARTS_OF_SPEECH**

- `n` – Noun
- `v` – Verb
- `a` – Adjective
- `r` – Adverb
- `s` – Adjective Satellite
- `t` – Phrase
- `c` – Conjunction
- `p` – Adposition
- `x` – Other

- u – Unknown

wn.constants.**NOUN = 'n'**

wn.constants.**VERB = 'v'**

wn.constants.**ADJECTIVE = 'a'**

wn.constants.**ADJ**
Alias of *ADJECTIVE*

wn.constants.**ADJECTIVE_SATELLITE = 's'**

wn.constants.**ADJ_SAT**
Alias of *ADJECTIVE_SATELLITE*

wn.constants.**PHRASE = 't'**

wn.constants.**CONJUNCTION = 'c'**

wn.constants.**CONJ**
Alias of *CONJUNCTION*

wn.constants.**ADPOSITION = 'p'**

wn.constants.**ADP = 'p'**
Alias of *ADPOSITION*

wn.constants.**OTHER = 'x'**

wn.constants.**UNKNOWN = 'u'**

3.11.4 Adjective Positions

wn.constants.**ADJPOSITIONS**

- a – Attributive
- ip – Immediate Postnominal
- p – Predicative

3.11.5 Lexicographer Files

wn.constants.**LEXICOGRAPHER_FILES**

```
{
    'adj.all': 0,
    'adj.pert': 1,
    'adv.all': 2,
    'noun.Tops': 3,
    'noun.act': 4,
    'noun.animal': 5,
    'noun.artifact': 6,
    'noun.attribute': 7,
    'noun.body': 8,
```

(continues on next page)

(continued from previous page)

```
    'noun.cognition': 9,
    'noun.communication': 10,
    'noun.event': 11,
    'noun.feeling': 12,
    'noun.food': 13,
    'noun.group': 14,
    'noun.location': 15,
    'noun.motive': 16,
    'noun.object': 17,
    'noun.person': 18,
    'noun.phenomenon': 19,
    'noun.plant': 20,
    'noun.possession': 21,
    'noun.process': 22,
    'noun.quantity': 23,
    'noun.relation': 24,
    'noun.shape': 25,
    'noun.state': 26,
    'noun.substance': 27,
    'noun.time': 28,
    'verb.body': 29,
    'verb.change': 30,
    'verb.cognition': 31,
    'verb.communication': 32,
    'verb.competition': 33,
    'verb.consumption': 34,
    'verb.contact': 35,
    'verb.creation': 36,
    'verb.emotion': 37,
    'verb.motion': 38,
    'verb.perception': 39,
    'verb.possession': 40,
    'verb.social': 41,
    'verb.stative': 42,
    'verb.weather': 43,
    'adj.ppl': 44,
}
```

3.12 wn.ic

Information Content is a corpus-based metrics of synset or sense specificity.

The mathematical formulae for information content are defined in *Formal Description*, and the corresponding Python API function are described in *Calculating Information Content*. These functions require information content weights obtained either by *computing them from a corpus*, or by *loading pre-computed weights from a file*.

Note: The term *information content* can be ambiguous. It often, and most accurately, refers to the result of the `information_content()` function (IC(c) in the mathematical notation), but is also sometimes used to refer to the corpus frequencies/weights (freq(c) in the mathematical notation) returned by `load()` or `compute()`, as these weights are the basis of the value computed by `information_content()`. The Wn documentation tries to consistently refer

to former as the *information content value*, or just *information content*, and the latter as *information content weights*, or *weights*.

3.12.1 Formal Description

The Information Content (IC) of a concept (synset) is a measure of its specificity computed from the wordnet's taxonomy structure and corpus frequencies. It is defined by Resnik 1995 ([RES95]), following information theory, as the negative log-probability of a concept:

$$\mathrm{IC}(c) = -\log p(c)$$

A concept's probability is the empirical probability over a corpus:

$$p(c) = \frac{\mathrm{freq}(c)}{N}$$

Here, N is the total count of words of the same category as concept c ([RES95] only considered nouns) where each word has some representation in the wordnet, and freq is defined as the sum of corpus counts of words in $\mathrm{words}(c)$, which is the set of words subsumed by concept c:

$$\mathrm{freq}(c) = \sum_{w \in \mathrm{words}(c)} \mathrm{count}(w)$$

It is common for freq to not contain actual frequencies but instead weights distributed evenly among the synsets for a word. These weights are calculated as the word frequency divided by the number of synsets for the word:

$$\mathrm{freq}_{\mathrm{distributed}}(c) = \sum_{w \in \mathrm{words}(c)} \frac{\mathrm{count}(w)}{|\mathrm{synsets}(w)|}$$

3.12.2 Example

In the Princeton WordNet 3.0 (hereafter *WordNet*, but note that the equivalent lexicon in Wn is the *OMW English Wordnet based on WordNet 3.0* with specifier `omw-en:1.4`), the frequency of a concept like **stone fruit** is not just the number of occurrences of *stone fruit*, but also includes the counts of the words for its hyponyms (*almond*, *olive*, etc.) and other taxonomic descendants (*Jordan almond*, *green olive*, etc.). The word *almond* has two synsets: one for the fruit or nut, another for the plant. Thus, if the word *almond* is encountered n times in a corpus, then the weight (either the frequency n or distributed weight $\frac{n}{2}$) is added to the total weights for both synsets and to those of their ancestors, but not for descendant synsets, such as for **Jordan almond**. The fruit/nut synset of almond has two hypernym paths which converge on **fruit**:

1. **almond stone fruit fruit**
2. **almond nut seed fruit**

The weight is added to each ancestor (**stone fruit**, **nut**, **seed**, **fruit**, ...) once. That is, the weight is not added to the convergent ancestor for **fruit** twice, but only once.

3.12.3 Calculating Information Content

wn.ic.**information_content**(*synset*, *freq*)

Calculate the Information Content value for a synset.

The information content of a synset is the negative log of the synset probability (see *synset_probability()*).

Parameters

- **synset** (wn.Synset) –
- **freq** (*Dict[str, Dict[Optional[str], float]]*) –

Return type float

wn.ic.**synset_probability**(*synset*, *freq*)

Calculate the synset probability.

The synset probability is defined as freq(ss)/N where freq(ss) is the IC weight for the synset and N is the total IC weight for all synsets with the same part of speech.

Note: this function is not generally used directly, but indirectly through *information_content()*.

Parameters

- **synset** (wn.Synset) –
- **freq** (*Dict[str, Dict[Optional[str], float]]*) –

Return type float

3.12.4 Computing Corpus Weights

If pre-computed weights are not available for a wordnet or for some domain, they can be computed given a corpus and a wordnet.

The corpus is an iterable of words. For large corpora it may help to use a generator for this iterable, but the entire vocabulary (i.e., unique words and counts) will be held at once in memory. Multi-word expressions are also possible if they exist in the wordnet. For instance, WordNet has *stone fruit*, with a single space delimiting the words, as an entry.

The *wn.Wordnet* object must be instantiated with a single lexicon, although it may have expand-lexicons for relation traversal. For best results, the wordnet should use a lemmatizer to help it deal with inflected wordforms from running text.

wn.ic.**compute**(*corpus*, *wordnet*, *distribute_weight=True*, *smoothing=1.0*)

Compute Information Content weights from a corpus.

Parameters

- **corpus** (*Iterable[str]*) – An iterable of string tokens. This is a flat list of words and the order does not matter. Tokens may be single words or multiple words separated by a space.
- **wordnet** (wn.Wordnet) – An instantiated *wn.Wordnet* object, used to look up synsets from words.
- **distribute_weight** (*bool*) – If True, the counts for a word are divided evenly among all synsets for the word.
- **smoothing** (*float*) – The initial value given to each synset.

Return type *Dict*[str, *Dict*[*Optional*[str], float]]

Example

```
>>> import wn, wn.ic, wn.morphy
>>> ewn = wn.Wordnet('ewn:2020', lemmatizer=wn.morphy.morphy)
>>> freq = wn.ic.compute(["Dogs", "run", ".", "Cats", "sleep", "."], ewn)
>>> dog = ewn.synsets('dog', pos='n')[0]
>>> cat = ewn.synsets('cat', pos='n')[0]
>>> frog = ewn.synsets('frog', pos='n')[0]
>>> freq['n'][dog.id]
1.125
>>> freq['n'][cat.id]
1.1
>>> freq['n'][frog.id]  # no occurrence; smoothing value only
1.0
>>> carnivore = dog.lowest_common_hypernyms(cat)[0]
>>> freq['n'][carnivore.id]
1.3250000000000002
```

3.12.5 Reading Pre-computed Information Content Files

The *load()* function reads pre-computed information content weights files as used by the WordNet::Similarity Perl module or the NLTK Python package. These files are computed for a specific version of a wordnet using the synset offsets from the WNDB format, which Wn does not use. These offsets therefore must be converted into an identifier that matches those used by the wordnet. By default, *load()* uses the lexicon identifier from its *wordnet* argument with synset offsets (padded with 0s to make 8 digits) and parts-of-speech from the weights file to format an identifier, such as `omw-en-00001174-n`. For wordnets that use a different identifier scheme, the *get_synset_id* parameter of *load()* can be given a callable created with *wn.util.synset_id_formatter()*. It can also be given another callable with the same signature as shown below:

```
get_synset_id(*, offset: int, pos: str) -> str
```

When loading pre-computed information content files, it is recommended to use the ones with smoothing (i.e., `*-add1.dat` or `*-resnik-add1.dat`) to avoid math domain errors when computing the information content value.

Warning: The weights files are only valid for the version of wordnet for which they were created. Files created for WordNet 3.0 do not work for WordNet 3.1 because the offsets used in its identifiers are different, although the *get_synset_id* parameter of *load()* could be given a function that performs a suitable mapping. Some Open Multilingual Wordnet wordnets use the WordNet 3.0 offsets in their identifiers and can therefore technically use the weights, but this usage is discouraged because the distributional properties of text in another language and the structure of the other wordnet will not be compatible with that of the English WordNet. For these cases, it is recommended to compute new weights using *compute()*.

`wn.ic.load`(*source*, *wordnet*, *get_synset_id=None*)

Load an Information Content mapping from a file.

Parameters

- **source** (*Union[str, pathlib.Path]*) – A path to an information content weights file.
- **wordnet** (`wn.Wordnet`) – A *wn.Wordnet* instance with synset identifiers matching the offsets in the weights file.

- **get_synset_id** (*Optional[Callable]*) – A callable that takes a synset offset and part of speech and returns a synset ID valid in *wordnet*.

Raises ***wn.Error*** – If *wordnet* does not have exactly one lexicon.

Return type *Dict*[str, *Dict*[*Optional*[str], float]]

Example

```
>>> import wn, wn.ic
>>> pwn = wn.Wordnet('pwn:3.0')
>>> path = '~/nltk_data/corpora/wordnet_ic/ic-brown-resnik-add1.dat'
>>> freq = wn.ic.load(path, pwn)
```

3.13 wn.lmf

Reader for the Lexical Markup Framework (LMF) format.

wn.lmf.**load**(*source*, *progress_handler=<class 'wn.util.ProgressBar'>*)

Load wordnets encoded in the WN-LMF format.

Parameters

- **source** (*Union[str, pathlib.Path]*) – path to a WN-LMF file
- **progress_handler** (*Optional[Type[wn.util.ProgressHandler]]*) –

Return type wn.lmf.LexicalResource

wn.lmf.**scan_lexicons**(*source*)

Scan *source* and return only the top-level lexicon info.

Parameters **source** (*Union[str, pathlib.Path]*) –

Return type *List*[*Dict*]

wn.lmf.**is_lmf**(*source*)

Return True if *source* is a WN-LMF file.

Parameters **source** (*Union[str, pathlib.Path]*) –

Return type bool

3.14 wn.morphy

A simple English lemmatizer that finds and removes known suffixes.

See also:

The Princeton WordNet documentation describes the original implementation of Morphy.

The *Lemmatization and Normalization* guide describes how Wn handles lemmatization in general.

3.14.1 Initialized and Uninitialized Morphy

There are two ways of using Morphy in Wn: initialized and uninitialized.

Unintialized Morphy is a simple callable that returns lemma *candidates* for some given wordform. That is, the results might not be valid lemmas, but this is not a problem in practice because subsequent queries against the database will filter out the invalid ones. This callable is obtained by creating a *Morphy* object with no arguments:

```
>>> from wn import morphy
>>> m = morphy.Morphy()
```

As an uninitialized Morphy cannot predict which lemmas in the result are valid, it always returns the original form and any transformations it can find for each part of speech:

```
>>> m('lemmata', pos='n')  # exceptional form
{'n': {'lemmata'}}
>>> m('lemmas', pos='n')   # regular morphology with part-of-speech
{'n': {'lemma', 'lemmas'}}
>>> m('lemmas')            # regular morphology for any part-of-speech
{None: {'lemmas'}, 'n': {'lemma'}, 'v': {'lemma'}}
>>> m('wolves')            # invalid forms may be returned
{None: {'wolves'}, 'n': {'wolf', 'wolve'}, 'v': {'wolve', 'wolv'}}
```

This lemmatizer can also be used with a *wn.Wordnet* object to expand queries:

```
>>> import wn
>>> ewn = wn.Wordnet('ewn:2020')
>>> ewn.words('lemmas')
[]
>>> ewn = wn.Wordnet('ewn:2020', lemmatizer=morphy.Morphy())
>>> ewn.words('lemmas')
[Word('ewn-lemma-n')]
```

An initialized Morphy is created with a *wn.Wordnet* object as its argument. It then uses the wordnet to build lists of valid lemmas and exceptional forms (this takes a few seconds). Once this is done, it will only return lemmas it knows about:

```
>>> ewn = wn.Wordnet('ewn:2020')
>>> m = morphy.Morphy(ewn)
>>> m('lemmata', pos='n')  # exceptional form
{'n': {'lemma'}}
>>> m('lemmas', pos='n')   # regular morphology with part-of-speech
{'n': {'lemma'}}
>>> m('lemmas')            # regular morphology for any part-of-speech
{'n': {'lemma'}}
>>> m('wolves')            # invalid forms are pre-filtered
{'n': {'wolf'}}
```

In order to use an initialized Morphy lemmatizer with a *wn.Wordnet* object, it must be assigned to the object after creation:

```
>>> ewn = wn.Wordnet('ewn:2020')  # default: lemmatizer=None
>>> ewn.words('lemmas')
[]
>>> ewn.lemmatizer = morphy.Morphy(ewn)
```

(continues on next page)

(continued from previous page)

```
>>> ewn.words('lemmas')
[Word('ewn-lemma-n')]
```

There is little to no difference in the results obtained from a *wn.Wordnet* object using an initialized or uninitialized *Morphy* object, but there may be slightly different performance profiles for future queries.

3.14.2 Default Morphy Lemmatizer

As a convenience, an uninitialized Morphy lemmatizer is provided in this module via the *morphy* member.

wn.morphy.**morphy**

A *Morphy* object created without a *wn.Wordnet* object.

3.14.3 The Morphy Class

class wn.morphy.**Morphy**(*wordnet=None*)

The Morphy lemmatizer class.

Objects of this class are callables that take a wordform and an optional part of speech and return a dictionary mapping parts of speech to lemmas. If objects of this class are not created with a *wn.Wordnet* object, the returned lemmas may be invalid.

Parameters **wordnet** (*Optional*[wn.Wordnet]) – optional *wn.Wordnet* instance

Example

```
>>> import wn
>>> from wn.morphy import Morphy
>>> ewn = wn.Wordnet('ewn:2020')
>>> m = Morphy(ewn)
>>> m('axes', pos='n')
{'n': {'axe', 'ax', 'axis'}}
>>> m('geese', pos='n')
{'n': {'goose'}}
>>> m('gooses')
{'n': {'goose'}, 'v': {'goose'}}
>>> m('goosing')
{'v': {'goose'}}
```

3.15 wn.project

Wordnet and ILI Packages and Collections

wn.project.**iterpackages**(*path*)

Yield any wordnet or ILI packages found at *path*.

The *path* argument can point to one of the following:

- a lexical resource file or ILI file
- a wordnet package directory

- a wordnet collection directory
- a tar archive containing one of the above
- a compressed (gzip or lzma) resource file or tar archive

Parameters **path** (*Union[str, pathlib.Path]*) –

Return type *Iterator*[*wn.project.Package*]

wn.project.**is_package_directory**(*path*)

Return `True` if *path* appears to be a wordnet or ILI package.

Parameters **path** (*Union[str, pathlib.Path]*) –

Return type bool

wn.project.**is_collection_directory**(*path*)

Return `True` if *path* appears to be a wordnet collection.

Parameters **path** (*Union[str, pathlib.Path]*) –

Return type bool

class wn.project.**Package**(*path*)

This class represents a wordnet or ILI package – a directory with a resource file and optional metadata.

Parameters **path** (*Union[str, pathlib.Path]*) –

resource_file()

Return the path of the package's resource file.

Return type pathlib.Path

readme()

Return the path of the README file, or `None` if none exists.

Return type *Optional*[pathlib.Path]

license()

Return the path of the license, or `None` if none exists.

Return type *Optional*[pathlib.Path]

citation()

Return the path of the citation, or `None` if none exists.

Return type *Optional*[pathlib.Path]

class wn.project.**Collection**(*path*)

This class represents a wordnet or ILI collection – a directory with one or more wordnet/ILI packages and optional metadata.

Parameters **path** (*Union[str, pathlib.Path]*) –

packages()

Return the list of packages in the collection.

Return type *List*[*wn.project.Package*]

readme()

Return the path of the README file, or `None` if none exists.

Return type *Optional*[pathlib.Path]

license()

Return the path of the license, or `None` if none exists.

Return type *Optional*[pathlib.Path]

citation()

Return the path of the citation, or `None` if none exists.

Return type *Optional*[pathlib.Path]

3.16 wn.similarity

Synset similarity metrics.

3.16.1 Taxonomy-based Metrics

The *Path*, *Leacock-Chodorow*, and *Wu-Palmer* similarity metrics work by finding path distances in the hypernym/hyponym taxonomy. As such, they are most useful when the synsets are, in fact, arranged in a taxonomy. For the Princeton WordNet and derivative wordnets, such as the Open English Wordnet and OMW English Wordnet based on WordNet 3.0 available to Wn, synsets for nouns and verbs are arranged taxonomically: the nouns mostly form a single structure with a single root while verbs form many smaller structures with many roots. Synsets for the other parts of speech do not use hypernym/hyponym relations at all. This situation may be different for other wordnet projects or future versions of the English wordnets.

The similarity metrics tend to fail when the synsets are not connected by some path. When the synsets are in different parts of speech, or even in separate lexicons, this failure is acceptable and expected. But for cases like the verbs in the Princeton WordNet, it might be more useful to pretend that there is some unique root for all verbs so as to create a path connecting any two of them. For this purpose, the *simulate_root* parameter is available on the *path()*, *lch()*, and *wup()* functions, where it is passed on to calls to *wn.Synset.shortest_path()* and *wn.Synset.lowest_common_hypernyms()*. Setting *simulate_root* to `True` can, however, give surprising results if the words are from a different lexicon. Currently, computing similarity for synsets from a different part of speech raises an error.

Path Similarity

When p is the length of the shortest path between two synsets, the path similarity is:

$$\frac{1}{p+1}$$

The similarity score ranges between 0.0 and 1.0, where the higher the score is, the more similar the synsets are. The score is 1.0 when a synset is compared to itself, and 0.0 when there is no path between the two synsets (i.e., the path distance is infinite).

wn.similarity.**path**(*synset1*, *synset2*, *simulate_root=False*)

Return the Path similarity of *synset1* and *synset2*.

Parameters

- **synset1** (wn.Synset) – The first synset to compare.
- **synset2** (wn.Synset) – The second synset to compare.

- **simulate_root** (*bool*) – When `True`, a fake root node connects all other roots; default: `False`.

Return type float

Example

```
>>> import wn
>>> from wn.similarity import path
>>> ewn = wn.Wordnet('ewn:2020')
>>> spatula = ewn.synsets('spatula')[0]
>>> path(spatula, ewn.synsets('pancake')[0])
0.058823529411764705
>>> path(spatula, ewn.synsets('utensil')[0])
0.2
>>> path(spatula, spatula)
1.0
>>> flip = ewn.synsets('flip', pos='v')[0]
>>> turn_over = ewn.synsets('turn over', pos='v')[0]
>>> path(flip, turn_over)
0.0
>>> path(flip, turn_over, simulate_root=True)
0.16666666666666666
```

Leacock-Chodorow Similarity

When p is the length of the shortest path between two synsets and d is the maximum taxonomy depth, the Leacock-Chodorow similarity is:

$$-\log\left(\frac{p+1}{2d}\right)$$

`wn.similarity.`**lch**(*synset1*, *synset2*, *max_depth*, *simulate_root=False*)

Return the Leacock-Chodorow similarity between *synset1* and *synset2*.

Parameters

- **synset1** (*wn.Synset*) – The first synset to compare.
- **synset2** (*wn.Synset*) – The second synset to compare.
- **max_depth** (*int*) – The taxonomy depth (see *wn.taxonomy.taxonomy_depth()*)
- **simulate_root** (*bool*) – When `True`, a fake root node connects all other roots; default: `False`.

Return type float

Example

```
>>> import wn, wn.taxonomy
>>> from wn.similarity import lch
>>> ewn = wn.Wordnet('ewn:2020')
>>> n_depth = wn.taxonomy.taxonomy_depth(ewn, 'n')
>>> spatula = ewn.synsets('spatula')[0]
>>> lch(spatula, ewn.synsets('pancake')[0], n_depth)
0.8043728156701697
>>> lch(spatula, ewn.synsets('utensil')[0], n_depth)
2.0281482472922856
>>> lch(spatula, spatula, n_depth)
3.6375861597263857
>>> v_depth = taxonomy.taxonomy_depth(ewn, 'v')
>>> flip = ewn.synsets('flip', pos='v')[0]
>>> turn_over = ewn.synsets('turn over', pos='v')[0]
>>> lch(flip, turn_over, v_depth, simulate_root=True)
1.3862943611198906
```

Wu-Palmer Similarity

When *LCS* is the lowest common hypernym (also called "least common subsumer") between two synsets, i is the shortest path distance from the first synset to *LCS*, j is the shortest path distance from the second synset to *LCS*, and k is the number of nodes (distance + 1) from *LCS* to the root node, then the Wu-Palmer similarity is:

$$\frac{2k}{i + j + 2k}$$

wn.similarity.**wup**(*synset1*, *synset2*, *simulate_root=False*)

Return the Wu-Palmer similarity of *synset1* and *synset2*.

Parameters

- **synset1** (wn.Synset) – The first synset to compare.
- **synset2** (wn.Synset) – The second synset to compare.
- **simulate_root** – When True, a fake root node connects all other roots; default: False.

Raises *wn.Error* – When no path connects the *synset1* and *synset2*.

Return type float

Example

```
>>> import wn
>>> from wn.similarity import wup
>>> ewn = wn.Wordnet('ewn:2020')
>>> spatula = ewn.synsets('spatula')[0]
>>> wup(spatula, ewn.synsets('pancake')[0])
0.2
>>> wup(spatula, ewn.synsets('utensil')[0])
0.8
```

(continues on next page)

(continued from previous page)

```
>>> wup(spatula, spatula)
1.0
>>> flip = ewn.synsets('flip', pos='v')[0]
>>> turn_over = ewn.synsets('turn over', pos='v')[0]
>>> wup(flip, turn_over, simulate_root=True)
0.2857142857142857
```

3.16.2 Information Content-based Metrics

The *Resnik*, *Jiang-Conrath*, and *Lin* similarity metrics work by computing the information content of the synsets and/or that of their lowest common hypernyms. They therefore require information content weights (see `wn.ic`), and the values returned necessarily depend on the weights used.

Resnik Similarity

The Resnik similarity (Resnik 1995) is the maximum information content value of the common subsumers (hypernym ancestors) of the two synsets. Formally it is defined as follows, where c_1 and c_2 are the two synsets being compared.

$$\max_{c \in S(c_1,c_2)} \mathrm{IC}(c)$$

Since a synset's information content is always equal or greater than the information content of its hypernyms, $S(c_1, c_2)$ above is more efficiently computed using the lowest common hypernyms instead of all common hypernyms.

`wn.similarity.res`(*synset1*, *synset2*, *ic*)

Return the Resnik similarity between *synset1* and *synset2*.

Parameters

- **synset1** (`wn.Synset`) – The first synset to compare.
- **synset2** (`wn.Synset`) – The second synset to compare.
- **ic** (*Dict[str, Dict[Optional[str], float]]*) – Information Content weights.

Return type float

Example

```
>>> import wn, wn.ic, wn.taxonomy
>>> from wn.similarity import res
>>> pwn = wn.Wordnet('pwn:3.0')
>>> ic = wn.ic.load('~/nltk_data/corpora/wordnet_ic/ic-brown.dat', pwn)
>>> spatula = pwn.synsets('spatula')[0]
>>> res(spatula, pwn.synsets('pancake')[0], ic)
0.8017591149538994
>>> res(spatula, pwn.synsets('utensil')[0], ic)
5.87738923441087
```

Jiang-Conrath Similarity

The Jiang-Conrath similarity metric (Jiang and Conrath, 1997) combines the ideas of the taxonomy-based and information content-based metrics. It is defined as follows, where c_1 and c_2 are the two synsets being compared and c_0 is the lowest common hypernym of the two with the highest information content weight:

$$\frac{1}{\mathrm{IC}(c_1) + \mathrm{IC}(c_2) - 2(\mathrm{IC}(c_0))}$$

This equation is the simplified form given in the paper were several parameterized terms are cancelled out because the full form is not often used in practice.

There are two special cases:

1. If the information content of c_0, c_1, and c_2 are all zero, the metric returns zero. This occurs when both c_1 and c_2 are the root node, but it can also occur if the synsets did not occur in the corpus and the smoothing value was set to zero.
2. Otherwise if $c_1 + c_2 = 2c_0$, the metric returns infinity. This occurs when the two synsets are the same, one is a descendant of the other, etc., such that they have the same frequency as each other and as their lowest common hypernym.

`wn.similarity.`**`jcn`**(*synset1*, *synset2*, *ic*)

Return the Jiang-Conrath similarity of two synsets.

Parameters

- **synset1** (`wn.Synset`) – The first synset to compare.
- **synset2** (`wn.Synset`) – The second synset to compare.
- **ic** (*`Dict[str, Dict[Optional[str], float]]`*) – Information Content weights.

Return type float

Example

```
>>> import wn, wn.ic, wn.taxonomy
>>> from wn.similarity import jcn
>>> pwn = wn.Wordnet('pwn:3.0')
>>> ic = wn.ic.load('~/nltk_data/corpora/wordnet_ic/ic-brown.dat', pwn)
>>> spatula = pwn.synsets('spatula')[0]
>>> jcn(spatula, pwn.synsets('pancake')[0], ic)
0.04061799236354239
>>> jcn(spatula, pwn.synsets('utensil')[0], ic)
0.10794048564613007
```

Lin Similarity

Another formulation of information content-based similarity is the Lin metric (Lin 1997), which is defined as follows, where c_1 and c_2 are the two synsets being compared and c_0 is the lowest common hypernym with the highest information content weight:

$$\frac{2(\mathrm{IC}(c_0))}{\mathrm{IC}(c_1) + \mathrm{IC}(c_0)}$$

One special case is if either synset has an information content value of zero, in which case the metric returns zero.

wn.similarity.**lin**(*synset1*, *synset2*, *ic*)

Return the Lin similarity of two synsets.

Parameters

- **synset1** (wn.Synset) – The first synset to compare.
- **synset2** (wn.Synset) – The second synset to compare.
- **ic** (*Dict[str, Dict[Optional[str], float]]*) – Information Content weights.

Return type float

Example

```
>>> import wn, wn.ic, wn.taxonomy
>>> from wn.similarity import lin
>>> pwn = wn.Wordnet('pwn:3.0')
>>> ic = wn.ic.load('~/nltk_data/corpora/wordnet_ic/ic-brown.dat', pwn)
>>> spatula = pwn.synsets('spatula')[0]
>>> lin(spatula, pwn.synsets('pancake')[0], ic)
0.061148956278604116
>>> lin(spatula, pwn.synsets('utensil')[0], ic)
0.5592415686750427
```

3.17 wn.taxonomy

Functions for working with hypernym/hyponym taxonomies.

3.17.1 Overview

Among the valid synset relations for wordnets (see *wn.constants.SYNSET_RELATIONS*), those used for describing *is-a* taxonomies are given special treatment and they are generally the most well-developed relations in any wordnet. Typically these are the `hypernym` and `hyponym` relations, which encode *is-a-type-of* relationships (e.g., a *hermit crab* is a type of *decapod*, which is a type of *crustacean*, etc.). They also include `instance_hypernym` and `instance_hyponym`, which encode *is-an-instance-of* relationships (e.g., *Oregon* is an instance of *American state*).

The taxonomy forms a multiply-inheriting hierarchy with the synsets as nodes. In the English wordnets, such as the Princeton WordNet and its derivatives, nearly all nominal synsets form such a hierarchy with single root node, while verbal synsets form many smaller hierarchies without a common root. Other wordnets may have different properties, but as many are based off of the Princeton WordNet, they tend to follow this structure.

Functions to find paths within the taxonomies form the basis of all *wordnet similarity measures*. For instance, the *Leacock-Chodorow Similarity* measure uses both *shortest_path()* and (indirectly) *taxonomy_depth()*.

3.17.2 Wordnet-level Functions

Root and leaf synsets in the taxonomy are those with no ancestors (`hypernym`, `instance_hypernym`, etc.) or hyponyms (`hyponym`, `instance_hyponym`, etc.), respectively.

Finding root and leaf synsets

wn.taxonomy.**roots**(*wordnet*, *pos=None*)

Return the list of root synsets in *wordnet*.

Parameters

- **wordnet** (Wordnet) – The wordnet from which root synsets are found.
- **pos** (*Optional[str]*) – If given, only return synsets with the specified part of speech.

Return type *List*[*Synset*]

Example

```
>>> import wn, wn.taxonomy
>>> ewn = wn.Wordnet('ewn:2020')
>>> len(wn.taxonomy.roots(ewn, pos='v'))
573
```

wn.taxonomy.**leaves**(*wordnet*, *pos=None*)

Return the list of leaf synsets in *wordnet*.

Parameters

- **wordnet** (Wordnet) – The wordnet from which leaf synsets are found.
- **pos** (*Optional[str]*) – If given, only return synsets with the specified part of speech.

Return type *List*[*Synset*]

Example

```
>>> import wn, wn.taxonomy
>>> ewn = wn.Wordnet('ewn:2020')
>>> len(wn.taxonomy.leaves(ewn, pos='v'))
10525
```

Computing the taxonomy depth

The taxonomy depth is the maximum depth from a root node to a leaf node within synsets for a particular part of speech.

wn.taxonomy.**taxonomy_depth**(*wordnet*, *pos*)

Return the list of leaf synsets in *wordnet*.

Parameters

- **wordnet** (Wordnet) – The wordnet for which the taxonomy depth will be calculated.

- **pos** (*str*) – The part of speech for which the taxonomy depth will be calculated.

Return type int

Example

```
>>> import wn, wn.taxonomy
>>> ewn = wn.Wordnet('ewn:2020')
>>> wn.taxonomy.taxonomy_depth(ewn, 'n')
19
```

3.17.3 Synset-level Functions

wn.taxonomy.**hypernym_paths**(*synset*, *simulate_root=False*)

Return the list of hypernym paths to a root synset.

Parameters

- **synset** (Synset) – The starting synset for paths to a root.
- **simulate_root** (*bool*) – If `True`, find the path to a simulated root node.

Return type *List*[*List*[*Synset*]]

Example

```
>>> import wn, wn.taxonomy
>>> dog = wn.synsets('dog', pos='n')[0]
>>> for path in wn.taxonomy.hypernym_paths(dog):
...     for i, ss in enumerate(path):
...         print(' ' * i, ss, ss.lemmas()[0])
...
 Synset('pwn-02083346-n') canine
  Synset('pwn-02075296-n') carnivore
   Synset('pwn-01886756-n') eutherian mammal
    Synset('pwn-01861778-n') mammalian
     Synset('pwn-01471682-n') craniate
      Synset('pwn-01466257-n') chordate
       Synset('pwn-00015388-n') animal
        Synset('pwn-00004475-n') organism
         Synset('pwn-00004258-n') animate thing
          Synset('pwn-00003553-n') unit
           Synset('pwn-00002684-n') object
            Synset('pwn-00001930-n') physical entity
             Synset('pwn-00001740-n') entity
 Synset('pwn-01317541-n') domesticated animal
  Synset('pwn-00015388-n') animal
   Synset('pwn-00004475-n') organism
    Synset('pwn-00004258-n') animate thing
     Synset('pwn-00003553-n') unit
      Synset('pwn-00002684-n') object
```

(continues on next page)

(continued from previous page)

```
        Synset('pwn-00001930-n') physical entity
         Synset('pwn-00001740-n') entity
```

wn.taxonomy.**min_depth**(*synset*, *simulate_root=False*)

Return the minimum taxonomy depth of the synset.

Parameters

- **synset** (Synset) – The starting synset for paths to a root.
- **simulate_root** (*bool*) – If True, find the depth to a simulated root node.

Return type int

Example

```
>>> import wn, wn.taxonomy
>>> dog = wn.synsets('dog', pos='n')[0]
>>> wn.taxonomy.min_depth(dog)
8
```

wn.taxonomy.**max_depth**(*synset*, *simulate_root=False*)

Return the maximum taxonomy depth of the synset.

Parameters

- **synset** (Synset) – The starting synset for paths to a root.
- **simulate_root** (*bool*) – If True, find the depth to a simulated root node.

Return type int

Example

```
>>> import wn, wn.taxonomy
>>> dog = wn.synsets('dog', pos='n')[0]
>>> wn.taxonomy.max_depth(dog)
13
```

wn.taxonomy.**shortest_path**(*synset*, *other*, *simulate_root=False*)

Return the shortest path from *synset* to the *other* synset.

Parameters

- **other** (Synset) – endpoint synset of the path
- **simulate_root** (*bool*) – if True, ensure any two synsets are always connected by positing a fake root node
- **synset** (Synset) –

Return type *List*[*Synset*]

Example

```
>>> import wn, wn.taxonomy
>>> dog = ewn.synsets('dog', pos='n')[0]
>>> squirrel = ewn.synsets('squirrel', pos='n')[0]
>>> for ss in wn.taxonomy.shortest_path(dog, squirrel):
...     print(ss.lemmas())
...
['canine', 'canid']
['carnivore']
['eutherian mammal', 'placental', 'placental mammal', 'eutherian']
['rodent', 'gnawer']
['squirrel']
```

wn.taxonomy.**common_hypernyms**(*synset*, *other*, *simulate_root=False*)

Return the common hypernyms for the current and *other* synsets.

Parameters

- **other** (Synset) – synset that is a hyponym of any shared hypernyms
- **simulate_root** (*bool*) – if True, ensure any two synsets always share a hypernym by positing a fake root node
- **synset** (Synset) –

Return type *List*[*Synset*]

Example

```
>>> import wn, wn.taxonomy
>>> dog = ewn.synsets('dog', pos='n')[0]
>>> squirrel = ewn.synsets('squirrel', pos='n')[0]
>>> for ss in wn.taxonomy.common_hypernyms(dog, squirrel):
...     print(ss.lemmas())
...
['entity']
['physical entity']
['object', 'physical object']
['unit', 'whole']
['animate thing', 'living thing']
['organism', 'being']
['fauna', 'beast', 'animate being', 'brute', 'creature', 'animal']
['chordate']
['craniate', 'vertebrate']
['mammalian', 'mammal']
['eutherian mammal', 'placental', 'placental mammal', 'eutherian']
```

wn.taxonomy.**lowest_common_hypernyms**(*synset*, *other*, *simulate_root=False*)

Return the common hypernyms furthest from the root.

Parameters

- **other** (Synset) – synset that is a hyponym of any shared hypernyms

- **simulate_root** (*bool*) – if `True`, ensure any two synsets always share a hypernym by positing a fake root node
- **synset** (`Synset`) –

Return type *List*[*Synset*]

Example

```
>>> import wn, wn.taxonomy
>>> dog = ewn.synsets('dog', pos='n')[0]
>>> squirrel = ewn.synsets('squirrel', pos='n')[0]
>>> len(wn.taxonomy.lowest_common_hypernyms(dog, squirrel))
1
>>> wn.taxonomy.lowest_common_hypernyms(dog, squirrel)[0].lemmas()
['eutherian mammal', 'placental', 'placental mammal', 'eutherian']
```

3.18 wn.util

Wn utility classes.

wn.util.**synset_id_formatter**(*fmt='{prefix}-{offset:08}-{pos}'*, ***kwargs*)

Return a function for formatting synset ids.

The *fmt* argument can be customized. It will be formatted using any other keyword arguments given to this function and any given to the resulting function. By default, the format string expects a `prefix` string argument for the namespace (such as a lexicon id), an `offset` integer argument (such as a WNDB offset), and a `pos` string argument.

Parameters

- **fmt** (*str*) – A Python format string
- ****kwargs** – Keyword arguments for the format string.

Return type *Callable*

Example

```
>>> pwn_synset_id = synset_id_formatter(prefix='pwn')
>>> pwn_synset_id(offset=1174, pos='n')
'pwn-00001174-n'
```

class wn.util.**ProgressHandler**(***, *message=''*, *count=0*, *total=0*, *refresh_interval=0*, *unit=''*, *status=''*, *file=<_io.TextIOWrapper name='<stderr>' mode='w' encoding='UTF-8'>*)

An interface for updating progress in long-running processes.

Long-running processes in Wn, such as *wn.download()* and *wn.add()*, call to a progress handler object as they go. The default progress handler used by Wn is *ProgressBar*, which updates progress by formatting and printing a textual bar to stderr. The *ProgressHandler* class may be used directly, which does nothing, or users may create their own subclasses for, e.g., updating a GUI or some other handler.

The initialization parameters, except for `file`, are stored in a *kwargs* member and may be updated after the handler is created through the *set()* method. The *update()* method is the primary way a counter is updated.

The *flash()* method is sometimes called for simple messages. When the process is complete, the *close()* method is called, optionally with a message.

Parameters

- **message** (*str*) –
- **count** (*int*) –
- **total** (*int*) –
- **refresh_interval** (*int*) –
- **unit** (*str*) –
- **status** (*str*) –
- **file** (*TextIO*) –

kwargs

A dictionary storing the updateable parameters for the progress handler. The keys are:

- `message` (`str`) – a generic message or name
- `count` (`int`) – the current progress counter
- `total` (`int`) – the expected final value of the counter
- `unit` (`str`) – the unit of measurement
- `status` (`str`) – the current status of the process

close()

Close the progress handler.

This might be useful for closing file handles or cleaning up resources.

Return type None

flash(*message*)

Issue a message unrelated to the current counter.

This may be useful for multi-stage processes to indicate the move to a new stage, or to log unexpected situations.

Parameters **message** (*str*) –

Return type None

set(***kwargs*)

Update progress handler parameters.

Calling this method also runs *update()* with an increment of 0, which causes a refresh of any indicator without changing the counter.

Return type None

update(*n=1*, *force=False*)

Update the counter with the increment value *n*.

This method should update the `count` key of *kwargs* with the increment value *n*. After this, it is expected to update some user-facing progress indicator.

If *force* is `True`, any indicator will be refreshed regardless of the value of the refresh interval.

Parameters

- **n** (*int*) –
- **force** (*bool*) –

Return type None

class wn.util.**ProgressBar**(*, *message=''*, *count=0*, *total=0*, *refresh_interval=0*, *unit=''*, *status=''*, *file=<_io.TextIOWrapper name='<stderr>' mode='w' encoding='UTF-8'>*)

A *ProgressHandler* subclass for printing a progress bar.

Example

```
>>> p = ProgressBar(message='Progress: ', total=10, unit=' units')
>>> p.update(3)
Progress: [#########                         ] (3/10 units)
```

See *format()* for a description of how the progress bar is formatted.

Parameters

- **message** (*str*) –
- **count** (*int*) –
- **total** (*int*) –
- **refresh_interval** (*int*) –
- **unit** (*str*) –
- **status** (*str*) –
- **file** (*TextIO*) –

FMT = '\r{message}{bar}{counter}{status}'

The default formatting template.

close()

Print a newline so the last printed bar remains on screen.

Return type None

flash(*message*)

Overwrite the progress bar with *message*.

Parameters **message** (*str*) –

Return type None

format()

Format and return the progress bar.

The bar is is formatted according to *FMT*, using variables from `kwargs` and two computed variables:

- `bar`: visualization of the progress bar, empty when `total` is 0
- `counter`: display of `count`, `total`, and `units`

```
>>> p = ProgressBar(message='Progress', count=2, total=10, unit='K')
>>> p.format()
'\rProgress [######                          ] (2/10K) '
>>> p = ProgressBar(count=2, status='Counting...')
```

(continues on next page)

(continued from previous page)

```
>>> p.format()
'\r (2) Counting...'
```

Return type str

update(*n=1*, *force=False*)

Increment the count by *n* and print the reformatted bar.

Parameters

- **n** (*int*) –
- **force** (*bool*) –

Return type None

3.19 wn.validate

Wordnet lexicon validation.

This module is for checking whether the the contents of a lexicon are valid according to a series of checks. Those checks are:

Code	Message
E101	ID is not unique within the lexicon.
W201	Lexical entry has no senses.
W202	Redundant sense between lexical entry and synset.
W203	Redundant lexical entry with the same lemma and synset.
E204	Synset of sense is missing.
W301	Synset is empty (not associated with any lexical entries).
W302	ILI is repeated across synsets.
W303	Proposed ILI is missing a definition.
W304	Existing ILI has a spurious definition.
E401	Relation target is missing or invalid.
W402	Relation type is invalid for the source and target.
W403	Redundant relation between source and target.
W404	Reverse relation is missing.
W501	Synset's part-of-speech is different from its hypernym's.
W502	Relation is a self-loop.

wn.validate.**validate**(*lex*, *select=('E', 'W')*, *progress_handler=<class 'wn.util.ProgressBar'>*)

Check *lex* for validity and return a report of the results.

The *select* argument is a sequence of check codes (e.g., `E101`) or categories (`E` or `W`).

The *progress_handler* parameter takes a subclass of *wn.util.ProgressHandler*. An instance of the class will be created, used, and closed by this function.

Parameters

- **lex** (*Union[wn.lmf.Lexicon, wn.lmf.LexiconExtension]*) –
- **select** (*Sequence[str]*) –
- **progress_handler** (*Optional[Type[wn.util.ProgressHandler]]*) –

Return type *Dict*[str, *Dict*[str, *Union*[str, *Dict*[str, *Dict*]]]]

3.20 wn.web

This module provides a RESTful API with JSON:API responses to queries against a Wn database. This API implements the primary queries of the Python API (see *Primary Queries*). For instance, to search all words in the `ewn:2020` lexicon with the form *jet* and part-of-speech *v*, we can perform the following query:

```
/lexicons/ewn:2020/words?form=jet&pos=v
```

This query would return the following response:

```
{
  "data": [
    {
      "id": "ewn-jet-v",
      "type": "word",
      "attributes": {
        "pos": "v",
        "lemma": "jet",
        "forms": ["jet", "jetted", "jetting"]
      },
      "links": {
        "self": "http://example.com/lexicons/ewn:2020/words/ewn-jet-v"
      },
      "relationships": {
        "senses": {
          "links": {"related": "http://example.com/lexicons/ewn:2020/words/ewn-jet-v/
↪senses"}
        },
        "synsets": {
          "data": [
            {"type": "synset", "id": "ewn-01518922-v"},
            {"type": "synset", "id": "ewn-01946093-v"}
          ]
        },
        "lexicon": {
          "links": {"related": "http://example.com/lexicons/ewn:2020"}
        }
      },
      "included": [
        {
          "id": "ewn-01518922-v",
          "type": "synset",
          "attributes": {"pos": "v", "ili": "i29306"},
          "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-01518922-v
↪"}
        },
        {
          "id": "ewn-01946093-v",
          "type": "synset",
          "attributes": {"pos": "v", "ili": "i31432"},
```

(continues on next page)

(continued from previous page)

```
          "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-01946093-v
↪"}
        }
      ]
    }
  ],
  "meta": {"total": 1}
}
```

Currently, only `GET` requests are handled.

3.20.1 Installing Dependencies

By default, Wn does not install the requirements needed for this module. Install them with the `[web]` extra:

```
$ pip install wn[web]
```

3.20.2 Running and Deploying the Server

This module does not provide an ASGI server, so one will need to be installed and ran separately. Any ASGI-compliant server should work.

For example, the Uvicorn server may be used directly for local development, optionally with the `--reload` option for hot reloading:

```
$ uvicorn --reload wn.web:app
```

For production, see Uvicorn's documentation about deployment.

3.20.3 Requests: API Endpoints

The module provides the following endpoints:

Endpoint	Description
`/words`	List words in all available lexicons
`/senses`	List senses in all available lexicons
`/synsets`	List synsets in all available lexicons
`/lexicons`	List available lexicons
`/lexicons/:lex`	Get lexicon with specifier `:lex`
`/lexicons/:lex/words`	List words for lexicon with specifier `:lex`
`/lexicons/:lex/words/:id/senses`	List senses for word `:id` in lexicon `:lex`
`/lexicons/:lex/words/:id`	Get word with ID `:id` in lexicon `:lex`
`/lexicons/:lex/senses`	List senses for lexicon with specifier `:lex`
`/lexicons/:lex/senses/:id`	Get sense with ID `:id` in lexicon `:lex`
`/lexicons/:lex/synsets`	List synsets for lexicon with specifier `:lex`
`/lexicons/:lex/synsets/:id`	Get synset with ID `:id` in lexicon `:lex`
`/lexicons/:lex/synsets/:id/members`	Get member senses for synset `:id` in lexicon `:lex`

3.20.4 Requests: Query Parameters

lang

Specifies the language in BCP 47 of the lexicon(s) from which results are returned.

Example:

```
/words?lang=fr
```

Valid for:

```
/lexicons
/words
/senses
/synsets
```

form

Specifies the word form of the objects that are returned.

Example:

```
/words?form=chat
```

Valid for:

```
/words
/senses
/synsets
/lexicon/:lex/words
/lexicon/:lex/senses
/lexicon/:lex/synsets
```

pos

Specifies the part-of-speech of the objects that are returned. Valid values are given in *Parts of Speech*.

Example:

```
/words?pos=v
```

Valid for:

```
/words
/senses
/synsets
/lexicon/:lex/words
/lexicon/:lex/senses
/lexicon/:lex/synsets
```

ili

Specifies the interlingual index of a synset.

Example:

```
/synsets?ili=i57031
```

Valid for:

```
/synsets
/lexicon/:lex/synsets
```

page[offset] and page[limit]

Used for pagination: `page[offset]` specifies the starting index of a set of results, and `page[limit]` specifies how many results from the offset will be returned.

Example:

```
/words?page[offset]=150
```

Valid for:

```
/words
/senses
/synsets
/lexicon/:lex/words
/lexicon/:lex/senses
/lexicon/:lex/synsets
```

3.20.5 Responses

Responses are JSON data following the JSON:API specification. A full description of JSON:API is left to the linked specification, but a brief walkthrough is provided here. First, the top-level structure of "to-one" responses (e.g., getting a single synset) is:

```
{
  "data": { ... },  // primary response data as a JSON object
  "meta": { ... }   // metadata for the response
}
```

For "to-many" responses (e.g., getting a list of matching synsets), it is the same as above except the `data` key maps to an array and it includes pagination links:

```
{
  "data": [{ ... }, ...],  // primary response data as an array of objects
  "links": { ... },        // pagination links
  "meta": { ... }          // metadata; e.g., total number of results
}
```

Each JSON:API *resource object* (the primary data given by the `data` key) has the following structure:

```
{
  "id": "...",                // Lexicon specifier or entity ID
  "type": "...",              // "lexicon", "word", "sense", or "synset"
  "attributes": { ... },      // Basic resource information
  "links": { "self": ... },   // URL for this specific resource
  "relationships": { ... },   // Word senses, synset members, other relations
  "included": [ ... ],        // Data for related resources
}
```

Lexicons

```
{
  "id": "ewn:2020",
  "type": "lexicon",
  "attributes": {
    "version": "2020",
    "label": "English WordNet",
    "language": "en",
    "license": "https://creativecommons.org/licenses/by/4.0/"
  },
  "links": {"self": "http://example.com/lexicons/ewn:2020"},
  "relationships": {
    "words": {"links": {"related": "http://example.com/lexicons/ewn:2020/words"}},
    "synsets": {"links": {"related": "http://example.com/lexicons/ewn:2020/synsets"}},
    "senses": {"links": {"related": "http://example.com/lexicons/ewn:2020/senses"}}
  }
}
```

Words

```
{
  "id": "ewn-brick-v",
  "type": "word",
  "attributes": {"pos": "v", "lemma": "brick", "forms": ["brick"]},
  "links": {"self": "http://example.com/lexicons/ewn:2020/words/ewn-brick-v"},
  "relationships": {
    "senses": {"links": {"related": "http://example.com/lexicons/ewn:2020/words/ewn-
↪brick-v/senses"}},
    "synsets": {"data": [{"type": "synset", "id": "ewn-90011761-v"}]},
    "lexicon": {"links": {"related": "http://example.com/lexicons/ewn:2020"}}
  },
  "included": [
    {
      "id": "ewn-90011761-v",
      "type": "synset",
      "attributes": {"pos": "v", "ili": null},
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-90011761-v"}
    }
  ]
}
```

Senses


```
{
  "id": "ewn-explain-v-00941308-01",
  "type": "sense",
  "links": {"self": "http://example.com/lexicons/ewn:2020/senses/ewn-explain-v-00941308-
↪01"},
  "relationships": {
    "word": {"links": {"related": "http://example.com/lexicons/ewn:2020/words/ewn-
↪explain-v"}},
    "synset": {"links": {"related": "http://example.com/lexicons/ewn:2020/synsets/ewn-
↪00941308-v"}},
    "lexicon": {"links": {"related": "http://example.com/lexicons/ewn:2020"}},
    "derivation": {
      "data": [
        {"type": "sense", "id": "ewn-explanatory-s-01327635-01"},
        {"type": "sense", "id": "ewn-explanation-n-07247081-01"}
      ]
    }
  },
  "included": [
    {
      "id": "ewn-explanatory-s-01327635-01",
      "type": "sense",
      "links": {"self": "http://example.com/lexicons/ewn:2020/senses/ewn-explanatory-s-
↪01327635-01"}
    },
    {
      "id": "ewn-explanation-n-07247081-01",
      "type": "sense",
      "links": {"self": "http://example.com/lexicons/ewn:2020/senses/ewn-explanation-n-
↪07247081-01"}
    }
  ]
}
```


Synsets


```
{
  "id": "ewn-03204585-n",
  "type": "synset",
  "attributes": {"pos": "n", "ili": "i52917"},
  "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-03204585-n"},
  "relationships": {
    "members": {"links": {"related": "http://example.com/lexicons/ewn:2020/synsets/ewn-
↪03204585-n/members"}},
    "words": {
      "data": [
        {"type": "word", "id": "ewn-dory-n"},
        {"type": "word", "id": "ewn-rowboat-n"},
        {"type": "word", "id": "ewn-dinghy-n"}
```


(continues on next page)

(continued from previous page)


```
      ]
    },
    "lexicon": {"links": {"related": "http://example.com/lexicons/ewn:2020"}},
    "hypernym": {"data": [{"type": "synset", "id": "ewn-04252125-n"}]},
    "mero_part": {
      "data": [
        {"type": "synset", "id": "ewn-03911849-n"},
        {"type": "synset", "id": "ewn-04439177-n"}
      ]
    },
    "hyponym": {
      "data": [
        {"type": "synset", "id": "ewn-04122550-n"},
        {"type": "synset", "id": "ewn-04584425-n"}
      ]
    }
  },
  "included": [
    {
      "id": "ewn-dory-n",
      "type": "word",
      "attributes": {"pos": "n", "lemma": "dory", "forms": ["dory"]},
      "links": {"self": "http://example.com/lexicons/ewn:2020/words/ewn-dory-n"}
    },
    {
      "id": "ewn-rowboat-n",
      "type": "word",
      "attributes": {"pos": "n", "lemma": "rowboat", "forms": ["rowboat"]},
      "links": {"self": "http://example.com/lexicons/ewn:2020/words/ewn-rowboat-n"}
    },
    {
      "id": "ewn-dinghy-n",
      "type": "word",
      "attributes": {"pos": "n", "lemma": "dinghy", "forms": ["dinghy"]},
      "links": {"self": "http://example.com/lexicons/ewn:2020/words/ewn-dinghy-n"}
    },
    {
      "id": "ewn-04252125-n",
      "type": "synset",
      "attributes": {"pos": "n", "ili": "i59107"},
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-04252125-n"}
    },
    {
      "id": "ewn-03911849-n",
      "type": "synset",
      "attributes": {"pos": "n", "ili": "i57094"},
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-03911849-n"}
    },
    {
      "id": "ewn-04439177-n",
      "type": "synset",
      "attributes": {"pos": "n", "ili": "i60240"},
```


(continues on next page)

(continued from previous page)

```
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-04439177-n"}
    },
    {
      "id": "ewn-04122550-n",
      "type": "synset",
      "attributes": {"pos": "n", "ili": "i58319"},
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-04122550-n"}
    },
    {
      "id": "ewn-04584425-n",
      "type": "synset",
      "attributes": {"pos": "n", "ili": "i61103"},
      "links": {"self": "http://example.com/lexicons/ewn:2020/synsets/ewn-04584425-n"}
    }
  ]
}
```

BIBLIOGRAPHY

[VOSSEN1998] Piek Vossen. 1998. *Introduction to EuroWordNet.* Computers and the Humanities, 32(2): 73–89.

[Vossen99] Vossen, Piek, Wim Peters, and Julio Gonzalo. "Towards a universal index of meaning." In Proceedings of ACL-99 workshop, Siglex-99, standardizing lexical resources, pp. 81-90. University of Maryland, 1999.

[Bond16] Bond, Francis, Piek Vossen, John Philip McCrae, and Christiane Fellbaum. "CILI: the Collaborative Interlingual Index." In Proceedings of the 8th Global WordNet Conference (GWC), pp. 50-57. 2016.

[RES95] Resnik, Philip. "Using information content to evaluate semantic similarity." In Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95), Montreal, Canada, pp. 448-453. 1995.

PYTHON MODULE INDEX

W

INDEX

Symbols

A

C

D

E

N

O

P

R

S

T

U

V

W