Package ‘taxalight’

May 8, 2026

Version 0.1.5

Title A Lightweight and Lightning-Fast Taxonomic Naming Interface

Description Creates a local Lightning Memory-Mapped Database ('LMDB') of many commonly used taxonomic authorities and provides functions that can quickly query this data. Supported taxonomic authorities include the Integrated Taxonomic Information System ('ITIS'), National Center for Biotechnology Information ('NCBI'), Global Biodiversity Information Facility ('GBIF'), Catalogue of Life ('COL'), and Open Tree Taxonomy ('OTT'). Name and identifier resolution using 'LMDB' can be hundreds of times faster than either relational databases or internet-based queries. Precise data provenance information for data derived from naming providers is also included.

License MIT + file LICENSE

Encoding UTF-8

ByteCompile true

Depends R (>= 4.0)

Imports thor, contentid

Suggests jsonlite, spelling, testthat, covr, knitr, rmarkdown, progress, utf8, crayon

URL https://github.com/cboettig/taxalight

BugReports https://github.com/cboettig/taxalight

Language en-US

RoxygenNote 7.1.1

NeedsCompilation no

Author Carl Boettiger [aut, cre] (ORCID: <https://orcid.org/0000-0002-1642-628X>), Kari Norman [aut] (ORCID: <https://orcid.org/0000-0002-2029-2325>)

Maintainer Carl Boettiger <cboettig@gmail.com>

Repository CRAN

Date/Publication 2021-09-14 16:20:06 UTC

Contents

`get_ids` *Return the accepted taxonomic identifier,* `acceptedNameUsageID` *given a scientific name*

Description

Return the accepted taxonomic identifier, `acceptedNameUsageID` given a scientific name

Usage

```
get_ids(
  name,
  provider = getOption("tl_default_provider", "itis"),
  version = tl_latest_version(),
  dir = tl_dir()
)
```

Arguments

`name`	character vector of scientific names
`provider`	Abbreviation for a known naming provider. Provider data should first be imported with `[tl_create]`. Note: setting provider to `"itis_test"` is for testing purposes only, use `"itis"` for the full ITIS data. See details
`version`	version of the authority to use (e.g. four-digit year)
`dir`	storage location for the LMDB databases

Value

a vector of matching accepted identifiers. Note that if the name provided is considered to be a synonym by the provider, then the ID corresponds to the accepted name and not the synonym. (i.e. `get_names(get_ids(synonym))`) will return the accepted name and not the synonym name.

Examples

```
 # slow initial import
sp <- c("Dendrocygna autumnalis", "Dendrocygna bicolor")
get_ids(sp, "itis_test") # use "itis_test" test data for example only
```

`get_names` *Return* `scientificName` *names given taxonomic identifiers*

Description

Return `scientificName` names given taxonomic identifiers

Usage

```
get_names(
  id,
  provider = getOption("tl_default_provider", "itis"),
  version = tl_latest_version(),
  dir = tl_dir()
)
```

Arguments

`id`	a character vector of taxonomic identifiers, including provider prefix
`provider`	Abbreviation for a known naming provider. Provider data should first be imported with `[tl_create]`. Note: setting provider to `"itis_test"` is for testing purposes only, use `"itis"` for the full ITIS data. See details
`version`	version of the authority to use (e.g. four-digit year)
`dir`	storage location for the LMDB databases

Value

a vector of matching scientific names

Examples

```
 # slow initial import
get_names(c("ITIS:180092", "ITIS:179913"), "itis_test") # uses test version
```

`tl` *taxalight query: rapidly look up scientific names from a local database*

Description

taxalight query: rapidly look up scientific names from a local database

Usage

```
tl(
  x,
  provider = getOption("tl_default_provider", "itis"),
  version = tl_latest_version(),
  dir = tl_dir()
)
```

Arguments

`x`	character vector of either scientific names or taxonomic identifiers (with prefix). Can mix and match too.
`provider`	Abbreviation for a known naming provider. Provider data should first be imported with `[tl_create]`. Note: setting provider to `"itis_test"` is for testing purposes only, use `"itis"` for the full ITIS data. See details
`version`	version of the authority to use (e.g. four-digit year)
`dir`	storage location for the LMDB databases

Details

Naming providers currently recognized by `taxalight` are:

- `itis`: Integrated Taxonomic Information System, `https://www.itis.gov/`
- `ncbi`: National Center for Biotechnology Information, `https://www.ncbi.nlm.nih.gov/taxonomy`
- `col`: Catalogue of Life, `http://www.catalogueoflife.org/`
- `gbif`: Global Biodiversity Information Facility, `https://www.gbif.org/`
- `ott`: OpenTree Taxonomy: `https://github.com/OpenTreeOfLife/reference-taxonomy`
- `itis_test`: a small subset of ITIS, cached locally for testing purposes only.

The default provider is `itis`, which can be reconfigured by setting `tl_default_provider` in `[options]`.

Value

a data.frame in Darwin Core format with rows matching the acceptedNameUsageID or scientificName requested.

See Also

tl_create

Examples

```
 # slow initial import
sp <- c("Dendrocygna autumnalis", "Dendrocygna bicolor")
id <- c("ITIS:180092", "ITIS:179913")

## example uses "itis_test" provider for illustration only:
tl(sp, "itis_test")
tl(id, "itis_test")
```

`tl_create` *Create a Lightning Memory-Mapped Database (LMDB) for a given provider*

Description

Download raw data and store in a local LMDB database. Importing data is a time-consuming step that needs be run only once per machine and will persist through sessions.

Usage

```
tl_create(
  provider = getOption("tl_default_provider", "itis"),
  version = tl_latest_version(),
  dir = tl_dir(),
  lines = 100000L
)
```

Arguments

`provider`	Abbreviation for a known naming provider. Provider data should first be imported with `[tl_create]`. Note: setting provider to `"itis_test"` is for testing purposes only, use `"itis"` for the full ITIS data. See details
`version`	version of the authority to use (e.g. four-digit year)
`dir`	storage location for the LMDB databases
`lines`	number of lines to read in each chunk.

Details

Naming providers currently recognized by `taxalight` are:

- `itis`: Integrated Taxonomic Information System, https://www.itis.gov/
- `ncbi`: National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/taxonomy
- `col`: Catalogue of Life, http://www.catalogueoflife.org/
- `gbif`: Global Biodiversity Information Facility, https://www.gbif.org/
- `ott`: OpenTree Taxonomy: https://github.com/OpenTreeOfLife/reference-taxonomy
- `itis_test`: a small subset of ITIS, cached locally for testing purposes only.

The default provider is `itis`, which can be reconfigured by setting `tl_default_provider` in `[options]`.

Examples

```
## example uses "itis_test" for illustration only:



 # test may take > 5s
tl_create("itis_test")
```

`tl_dir` *taxalight data directory*

Description

taxalight stores data for persistent access in the directory given by tl_dir() by default. All functions can override this choice by passing an alternative path to the `dir` argument, or configure the location system-wide by setting the environmental variable `TAXALIGHT_HOME`, eg. in their `.Renviron` file, see `Sys.setenv()`. If unset, the default location is the default for the operating system, as provided by the core R function `tools::R_user_dir()`. Users can manually purge the data storage at any time by deleting this directory.

Usage

```
tl_dir()
```

Examples

```
tl_dir()
```

Index