

idpp.db.builder

Subpackage for populating the IdPP database with data from literature sources.

Overview

The build scripts in this subpackage each provide a function for adding a literature data source into the database.

```
from idpp.db.builder.hmdb import add_hmdb_chunks_to_idppdb
from idpp.db.builder.mona import add_mona_chunks_to_idppdb
from idpp.db.builder.nist20 import add_nist20_msms_to_idppdb
from idpp.db.builder.report import add_report_datasets_to_idppdb
from idpp.db.builder.ccs_compendium import add_ccs_compendium_to_idppdb
from idpp.db.builder.ccsbase import add_ccsbase_to_idppdb
from idpp.db.builder.metlin_ccs import add_metlin_ccs_to_idppdb
```

There are also utility functions for taking some of the larger datasets and chunking them into more manageable sizes.

! Note

The literature datasets are not housed in this repository and must be downloaded from their respective sources.

Build functions

idpp.db.builder.hmdb

idpp.db.builder.hmdb.hmdb_chunks_exist(*chunk_dir: str*) → bool

Check that chunked HMDB download files exist

idpp.db.builder.hmdb.chunk_hmdb_xml(*hmdb_metabolites_xml: str, chunk_dir: str*) → None

process a download of “hmdb_metabolites.xml” from HMDB into chunks of a manageable size (~5 MB)

Parameters: | **hmdb_metabolites** : **str**

path to “hmdb_metabolites.xml” (download from HMDB)

| **chunk_dir** : **str**

directory to save the chunked files into

idpp.db.builder.hmdb.add_hmdb_chunks_to_idppdb(*db: IdPPdb, chunk_dir: str*) → None

Add chunked HMDB metabolite data to IdPPdb

Parameters: `db : IdPPdb`
IdPP database interface

`chunk_dir : str`
directory containing chunked HMDB download XML files

`idpp.db.builder.mona`

`idpp.db.builder.mona.mona_chunks_exist(chunk_dir: str)→ bool`

Check that chunked MoNA download files exist

`idpp.db.builder.mona.chunk_mona_json(mona_export_json: str, chunk_dir: str)→ None`

process a download of experimental MS/MS spectra from MoNA into chunks of a manageable size (~5 MB)

Parameters: `mona_export_json : str`
path to MoNA-export-Experimental_Spectra.json downloaded file

`chunk_dir : str`
directory to save the chunked files into

`idpp.db.builder.mona.add_mona_chunks_to_idppdb(db: IdPPdb, chunk_dir: str)→ None`

Add downloaded MoNA experimental MS/MS database (a JSON file) to IdPPdb

Parameters: `db : IdPPdb`
IdPP database interface

`chunk_dir : str`
path to directory with chunks from MoNA-export-Experimental_Spectra.json downloaded file

`idpp.db.builder.nist20`

`idpp.db.builder.nist20.add_nist20_msms_to_idppdb(db: IdPPdb, nist20_dbf: str)→ None`

Add NIST20 MS/MS database (a SQLite database file) to IdPPdb

Parameters: `db : IdPPdb`
IdPP database interface

`nist20_dbf : str`
path to NIST20 database file

`idpp.db.builder.report`

`idpp.db.builder.report.add_report_datasets_to_idppdb(db: IdPPdb, report_dir: str)→ None`

Add RT datasets from RepoRT (<https://github.com/michaelwitting/RepoRT>) to IdPPdb

Parameters: `db` : `IdPPdb`
IdPP database interface
`report_dir` : `str`
path to the RepoRT repository directory

`idpp.db.builder.ccs_compendium`

`idpp.db.builder.ccs_compendium.add_ccs_compendium_to_idppdb`(*db*: `IdPPdb`, *compendium_file*: `str`)→ `None`

Add downloaded CCS compendium (a .csv) to IdPPdb

Parameters: `db` : `IdPPdb`
IdPP database interface
`compendium_file` : `str`
path to CCS compendium download file (csv)

`idpp.db.builder.ccsbase`

`idpp.db.builder.ccsbase.add_ccsbase_to_idppdb`(*db*: `IdPPdb`, *c3sdb_file*: `str`)→ `None`

Add downloaded MoNA experimental MS/MS database (a JSON file) to IdPPdb

Parameters: `db` : `IdPPdb`
IdPP database interface
`c3sdb_file` : `str`
path to local C3S.db database

`idpp.db.builder.metlin_ccs`

`idpp.db.builder.metlin_ccs.add_metlin_ccs_to_idppdb`(*db*: `IdPPdb`, *metlin_ccs_file*: `str`)→ `None`

Add downloaded METLIN-CCS (.xlsx) to IdPPdb

Parameters: `db` : `IdPPdb`
IdPP database interface
`metlin_ccs_file` : `str`
path to METLIN-CCS download file (.xlsx)