

How to import Mesh disease classification records into database

Description:

Mesh terms are the medical classifications and categories of descriptors in a hierarchy structure. For more information please refer to official National Library of Medicine website:

(<https://www.nlm.nih.gov/>)

Import script description:

The provided script inventories all available mesh IDs in ChemBL database and once all mesh IDs are found, the script finds the corresponding disease classification in the downloaded MeshData file and persists disease classification back to database

How and where to Download MeSH Data:

1. Please navigate to following URL

<https://www.nlm.nih.gov/databases/download/mesh.html>


2. Please click on current Production year under ASCII Format.

Download MeSH Data

[Terms and Conditions](#)

Get the Data via Bulk Download

Medical Subject Headings (MeSH) is a hierarchically-organized terminology for indexing and cataloging of biomedical information. It is used for the indexing of PubMed and other NLM databases. Please see the [Terms and Conditions](#) for more information regarding the use and re-use of MeSH.



XML Format

[Current Production Year MeSH](#) [Archive of Prior Production Year MeSH](#)

ASCII Format

[Current Production Year MeSH](#) [Archive of Prior Production Year MeSH](#)

MARC 21 Format

[Current Production Year MeSH](#) [Archive of Prior Production Year MeSH](#)

RDF Format

[Current Production Year MeSH](#) [Archive of Prior Production Year MeSH](#)

3. Please download the file “d2020.bin” and place it in your preferred folder

Index of /online/mesh/MESH_FILES/asciimesh/

Not secure | ftp://nlmpubs.nlm.nih.gov/online/mesh/MESH_FILES/asciimesh/

Index of /online/mesh/MESH_FILES/asciimesh/

[parent directory]

Name	Size	Date Modified
c2019.bin	95.3 MB	12/1/19, 10:41:00 PM
c2020.bin	99.1 MB	4/21/20, 9:41:00 PM
d2020.bin	28.4 MB	4/21/20, 9:31:00 PM
q2020.bin	81.2 kB	4/21/20, 9:30:00 PM

How to run the code:

Pre-installation requirement:

The following libraries are required to run the import script “mesh_term_import_script.py”

Please run the following commands one by one to ensure the libraries are installed, if planning to run this from parent project folder, requirement.txt includes these libraries.

```
1. pip install numbers
2. pip install pandas
3. pip install SQLAlchemy
4. pip install pylint-flask-sqlalchemy
5. pip install re2
```

Database credentials:

open mesh_term_import_script.py file and provide database credential as following:

```
10 DATABASE_URL = "postgresql://YOUR_USERNAME:YOUR_PASSWORD@DB_URL:5432/drugdata"
11
```

Run the script:

- Navigate the folder that mesh_term_import_script.py exist and ensure the source file “d2020.bin” exist.
- Run the script by running: python mesh_term_import_script.py

```
C:\Development\School\CSCI-E 599\M3\code_4_22_2020\import_script_doc\import_scripts\mesh>python mesh_term_import_script.py
script finished inserting records into database
```