Setup for accessing UMLS database for different purposes stored in mysql server and accessed using python scripting on Linux based systems.

- You need to first register here and then download the UMLS database from here.

Task 1: Installing and populating **mysql server** with UMLS database:-

1. Open terminal (alt+ctrl+t).
2. Run command **sudo apt update** and **sudo apt grade,** enter 'y' on prompt.
3. Run command **sudo apt install mysql-server** (click here for details).
4. Enter into the mysql command line using: **sudo mysql.**
5. Query to set password for user: **ALTER USER 'root'@'localhost' IDENTIFIED WITH mysql_native_password BY 'your_password';**
6. Query: **show databases;**
   to verify installation. Should return:

```
+--------------------+
| Database           |
+--------------------+
| information_schema |
| mysql              |
| performance_schema |
| sys                |
+--------------------+
```

   4 rows in set (0.01 sec)
7. Query: **SHOW VARIABLES LIKE 'socket';** to find the socket location.
8. Query: **create database umls;**
9. **exit;**

   Here we installed mysql server with details:
   - 'username' = 'root'
   - 'password' = 'your_password'
   - 'host' = 'localhost'
   - 'socket' = '/var/run/mysqld/mysqld.sock' //can be found in Task 1 - step 7
10. Go to location **/etc/mysql/** and find file **my.cnf,** open it in sudo mode and below all text, add lines (change permissions for **my.cnf** as needed):

```
[client]
loose-local-infile=1

[mysqld]
local_infile=1
```

So that it will now be able to read text from local files in UMLS database.

11. Extract the UMLS database and navigate to directory **/umls-2024AA-metathesaurus-full/2024AA/META**
12. Here, file named **populate_mysql_db.sh** can be found, open it and set parameters as:

    MYSQL_HOME='/usr'
    user='root'
    password='your_password'
    db_name='umls'

    And leave other things as they are.

13. Replace file **mysql_tables.sql** with the one in the **files** directory with the same name.
14. Execute **bash populate_mysql_db.sh.** (try **sudo bash populate_mysql_db.sh** if required.)
15. From the directory **files**, download directory **NET** and it has additional tables that are not included in the umls databases somehow.
16. Login again into mysql using **mysql -u root -p** and query **use umls;**
17. Once the database changes, get the location of directory **NET** and run the following commands to create additional tables and read data from files into them:
    (replace [path_to_NET] by actual path of directory NET)

```
-- Create SRDEF table
CREATE TABLE SRDEF (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    RT VARCHAR(10) NOT NULL,
    UI VARCHAR(10) NOT NULL,
    STY_RTN VARCHAR(100),
    DEF TEXT,
    EX TEXT,
    UN TEXT,
    NH VARCHAR(2),
    ABR VARCHAR(10),
    RIN VARCHAR(10)
) CHARACTER SET utf8mb4;

-- Load data into SRDEF table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRDEF' INTO TABLE SRDEF
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create SRFIL table
CREATE TABLE SRFIL (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    FIL VARCHAR(100) NOT NULL,
    DES VARCHAR(200) NOT NULL,
    FMT VARCHAR(100) NOT NULL,
    CLS INT UNSIGNED NOT NULL,
    RWS INT UNSIGNED NOT NULL,
    BTS INT UNSIGNED NOT NULL
) CHARACTER SET utf8mb4;

-- Load data into SRFIL table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRFIL' INTO TABLE SRFIL
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create SRFLD table
CREATE TABLE SRFLD (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    COL VARCHAR(40) NOT NULL,
    DES VARCHAR(200) NOT NULL,
    REF VARCHAR(40),
    FIL VARCHAR(50) NOT NULL
) CHARACTER SET utf8mb4;

-- Load data into SRFLD table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRFLD' INTO TABLE SRFLD
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create SRSTR table
CREATE TABLE SRSTR (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    STY_RL1 VARCHAR(100) NOT NULL,
    RL VARCHAR(50) NOT NULL,
```

```
    STY_RL2 VARCHAR(100),
    LS VARCHAR(10) NOT NULL
) CHARACTER SET utf8mb4;

-- Load data into SRSTR table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRSTR' INTO TABLE SRSTR
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create SRSTRE1 table
CREATE TABLE SRSTRE1 (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    UI1 VARCHAR(100) NOT NULL,
    UI2 VARCHAR(100) NOT NULL,
    UI3 VARCHAR(100) NOT NULL
) CHARACTER SET utf8mb4;

-- Load data into SRSTRE1 table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRSTRE1' INTO TABLE SRSTRE1
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create SRSTRE2 table
CREATE TABLE SRSTRE2 (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    UI1 VARCHAR(100) NOT NULL,
    UI2 VARCHAR(100) NOT NULL,
    UI3 VARCHAR(100) NOT NULL
) CHARACTER SET utf8mb4;

-- Load data into SRSTRE2 table
LOAD DATA LOCAL INFILE '[path_to_NET]/SRSTRE2' INTO TABLE SRSTRE2
FIELDS TERMINATED BY '|'
LINES TERMINATED BY '\n';
```

```
-- Create the SU table for Semantic Types and Relations
```

```
CREATE TABLE SU (
    RECNUM INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
    UI VARCHAR(10) NOT NULL,
    STY VARCHAR(100),
    ABR VARCHAR(10),
    STN VARCHAR(20),
    DEF TEXT,
    EX TEXT,
    UN TEXT,
    NH VARCHAR(2),
    HL TEXT,
    RL VARCHAR(100),
    RTN VARCHAR(20),
    RIN VARCHAR(100),
    INH VARCHAR(2),
    STL TEXT,
    STLB VARCHAR(100)
) CHARACTER SET utf8mb4;

-- Load data into SU table
LOAD DATA LOCAL INFILE '[path_to_NET]/SU' INTO TABLE SU
FIELDS TERMINATED BY '\n'
LINES TERMINATED BY '\n\n'
(@UI, @STY, @ABR, @STN, @DEF, @EX, @UN, @NH, @HL, @RL, @RTN, @RIN, @INH,
@STL, @STLB)
SET
    UI = IF(@UI LIKE 'UI:%', TRIM(SUBSTRING(@UI, 5)), NULL),
    STY = IF(@STY LIKE 'STY:%', TRIM(SUBSTRING(@STY, 5)), NULL),
    ABR = IF(@ABR LIKE 'ABR:%', TRIM(SUBSTRING(@ABR, 5)), NULL),
    STN = IF(@STN LIKE 'STN:%', TRIM(SUBSTRING(@STN, 5)), NULL),
    DEF = IF(@DEF LIKE 'DEF:%', TRIM(SUBSTRING(@DEF, 5)), NULL),
    EX = IF(@EX LIKE 'EX:%', TRIM(SUBSTRING(@EX, 4)), NULL),
    UN = IF(@UN LIKE 'UN:%', TRIM(SUBSTRING(@UN, 4)), NULL),
    NH = IF(@NH LIKE 'NH:%', TRIM(SUBSTRING(@NH, 4)), NULL),
    HL = IF(@HL LIKE 'HL:%', TRIM(SUBSTRING(@HL, 4)), NULL),
    RL = IF(@RL LIKE 'RL:%', TRIM(SUBSTRING(@RL, 4)), NULL),
    RTN = IF(@RTN LIKE 'RTN:%', TRIM(SUBSTRING(@RTN, 5)), NULL),
    RIN = IF(@RIN LIKE 'RIN:%', TRIM(SUBSTRING(@RIN, 5)), NULL),
    INH = IF(@INH LIKE 'INH:%', TRIM(SUBSTRING(@INH, 5)), NULL),
    STL = IF(@STL LIKE 'STL:%', TRIM(SUBSTRING(@STL, 5)), NULL),
    STLB = IF(@STLB LIKE 'STLB:%', TRIM(SUBSTRING(@STLB, 6)), NULL);
```

Task 2: Installing necessary softwares in addition to Task1:-

1. Download WordNet-3.0 from here.
2. Open terminal and go to the downloaded file location and run the commands in exact given sequence:
   -- install tcl and tk
   a. apt-get install libx11-dev
   b. apt-get install tcl-dev
   c. apt-get install tcl tcltls tcllib tclx
   d. apt-get install tk-dev
   e. apt-get install libgtk-3-dev
   f. apt-get build-dep gtkwave
   g. apt install python3-tk
   h. apt-get install tk-dev
   -- install and configure wordnet with tcl and tk
   i. tar -xzf WordNet-3.0.tar.gz
   j. cd WordNet-3.0
   k. ./configure --with-tcl=/usr/lib/tcl8.6 --with-tk=/usr/lib/tk8.6
   l. make
   m. make install
   -- install perl and other add-ons
   n. apt-get install libdbd-mysql-perl (click here for details)
   o. sudo cpanm UMLS::Interface --force
   p. sudo cpanm UMLS::Similarity --force
   q. cpanm WordNet::QueryData
   r. cpanm WordNet::Similarity

Task 3: Setting up PyUMLS for Linux based system:-

1. Install pandas and PyUMLSSimilarity using **pip install pandas PyUMLS-Similarity.**
2. PyUMLSSimilarity is by default installed for Windows based systems, make these changes in **PyUMLS_Similarity.py file** for linux based systems**:**
   a. Set line 12 to executable perl file in **usr/bin/**
      for accessing: `self.perl_bin_path = r"/usr/bin/perl"`
   b. Create a temporary file to hold UMLS concept names and set line 28, 205 and 342 as: `in_file_path = r"[path_to_umls-similarity-temp.txt]"`
   c. Set perl bin path on line 157, 240 and 380 to:
      `cwd = r'[path_to_/perl5/bin]'`
   d. Execute using **similarity.py**.