

Pequeño manual para el uso del componente quickumls usando medspacy para el idioma español

Contenido

Introducción.....	1
1. Pasos para descomprimir e instalar umls:	2
2. Configurar el componente de quickumls	11
Referencias.....	12

Introducción

Para poder usar el componente quickumls se requieren varios pasos de configuración que pueden ser generalizados en dos: 1. Descomprimir e instalar umls y 2. Configurar el componente quickumls en un ambiente de desarrollo (es decir, notebooks).

La idea es poder usar medspacy quickumls como se describe en (Eyre et al., 2022)

1. Pasos para descomprimir e instalar umls:

1. Registrarse en umls (<https://uts.nlm.nih.gov/uts/signup-login>)

2. Descargar la totalidad de los archivos umls files

<https://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html>

UMLS Terminology Services

[About ▾](#)[Browse ▾](#)[Download ▾](#)[APIs ▾](#)

Welcome to UMLS Terminology Services (UTS). Your UTS account provides access to the Unified Medical Language System (UMLS), RxNorm downloads, SNOMED CT downloads and more.

Unified Medical Language System (UMLS)

A set of files and software that brings together many health and biomedical vocabularies and standards to enable interoperability between computer systems.

[Home](#) | [Browse](#) | [Download](#) | [API](#)

Value Set Authority Center (VSAC)

A repository and authoring tool for standard lists of codes and terms from biomedical vocabularies.

[Home](#) | [Browse](#) | [Download](#) | [API](#)

Please submit your comments to the [UMLS Feedback](#) using the subject line: UMLS Subject

2023AB Full UMLS Release Files

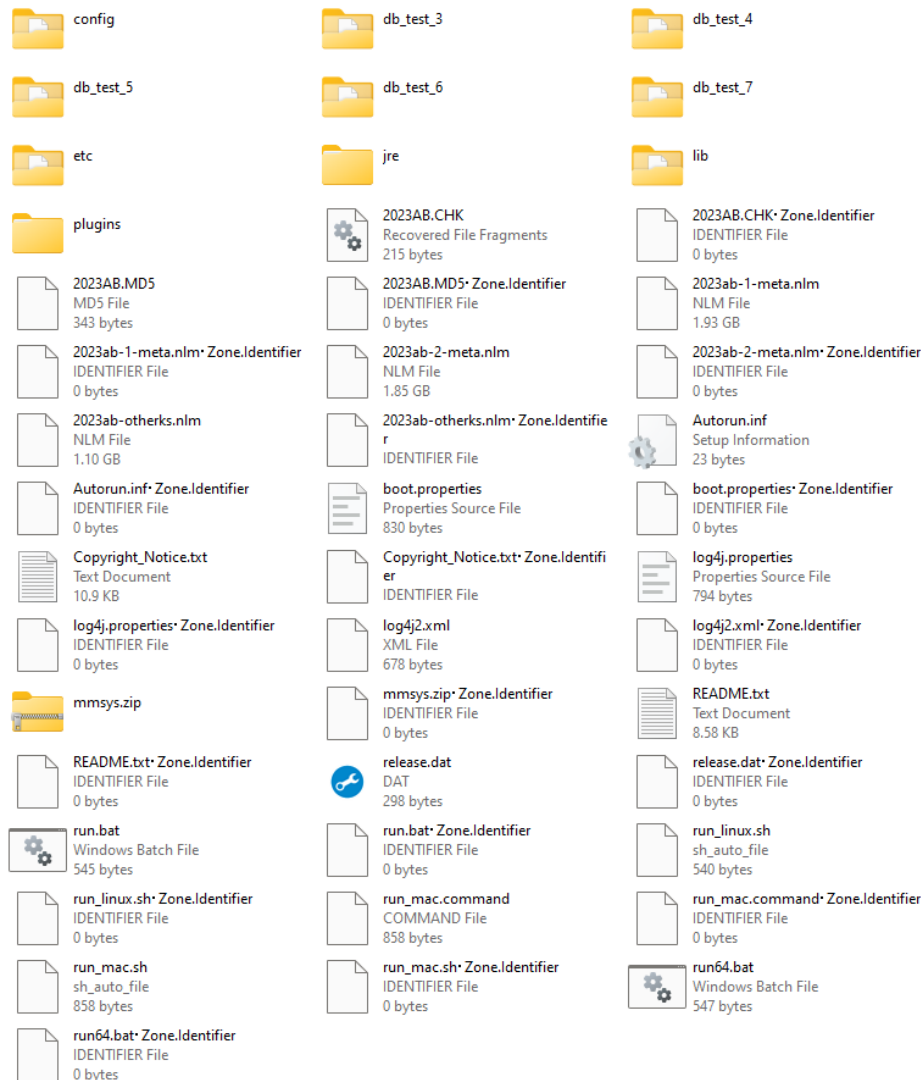
Release	Description	File Size (compressed)	File Size (uncompressed)	Date
Full Release (umls-2023AB-full.zip)	The full release includes the UMLS Metathesaurus, Semantic Network, Specialist Lexicon and Lexical Tools, database load scripts, and MetamorphoSys for customizing your UMLS subset and browsing the data.	5.1 GB	34.2 GB	November 6, 2023

Download Notes:

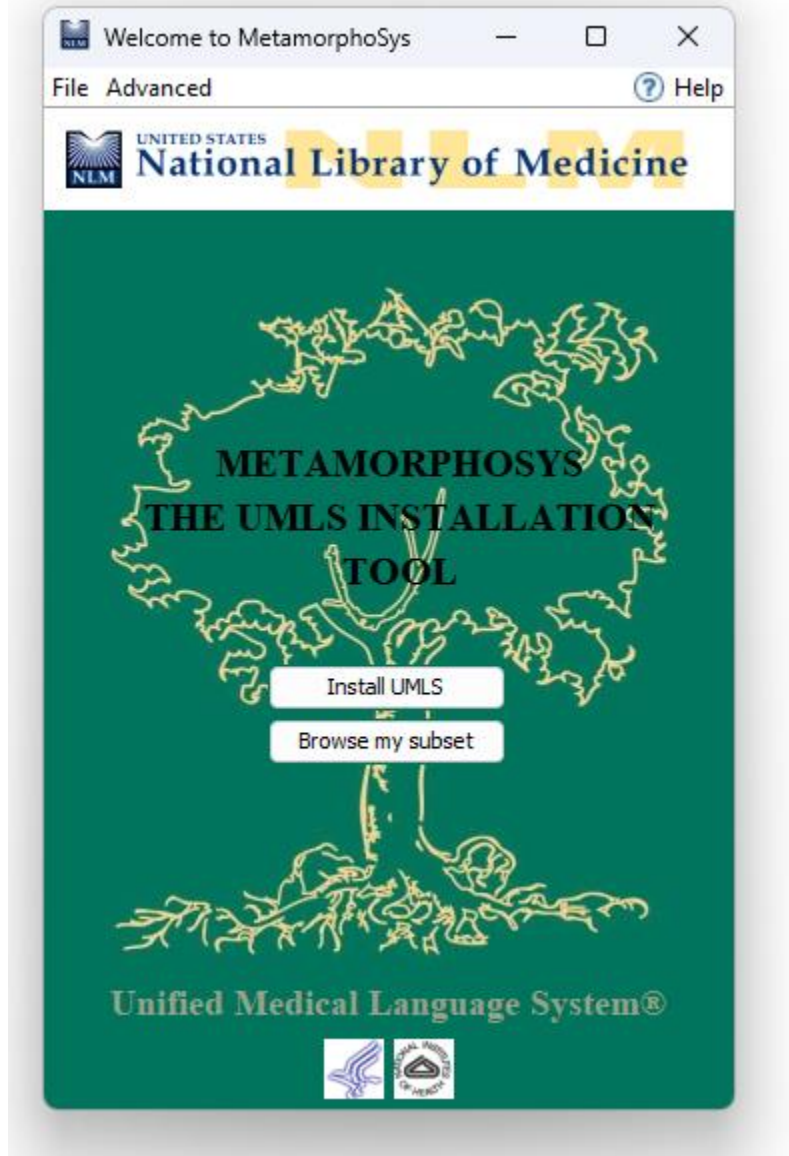
- Download the UMLS zip file and extract the contents to a single directory.
- Read the README file after downloading, which includes information on installing the UMLS Knowledge Sources and highlights changes for this release.
- Each UMLS release includes MetamorphoSys, required to install Knowledge Sources files, and to create, search and browse customized Metathesaurus subsets. MetamorphoSys requires a minimum of 40 GB of free hard disk and takes 2-10 hours to run on a range of platforms tested. The actual time will depend on your configuration, hardware and operating system platforms.

3. Instalar UMLS usando Metamorphosys

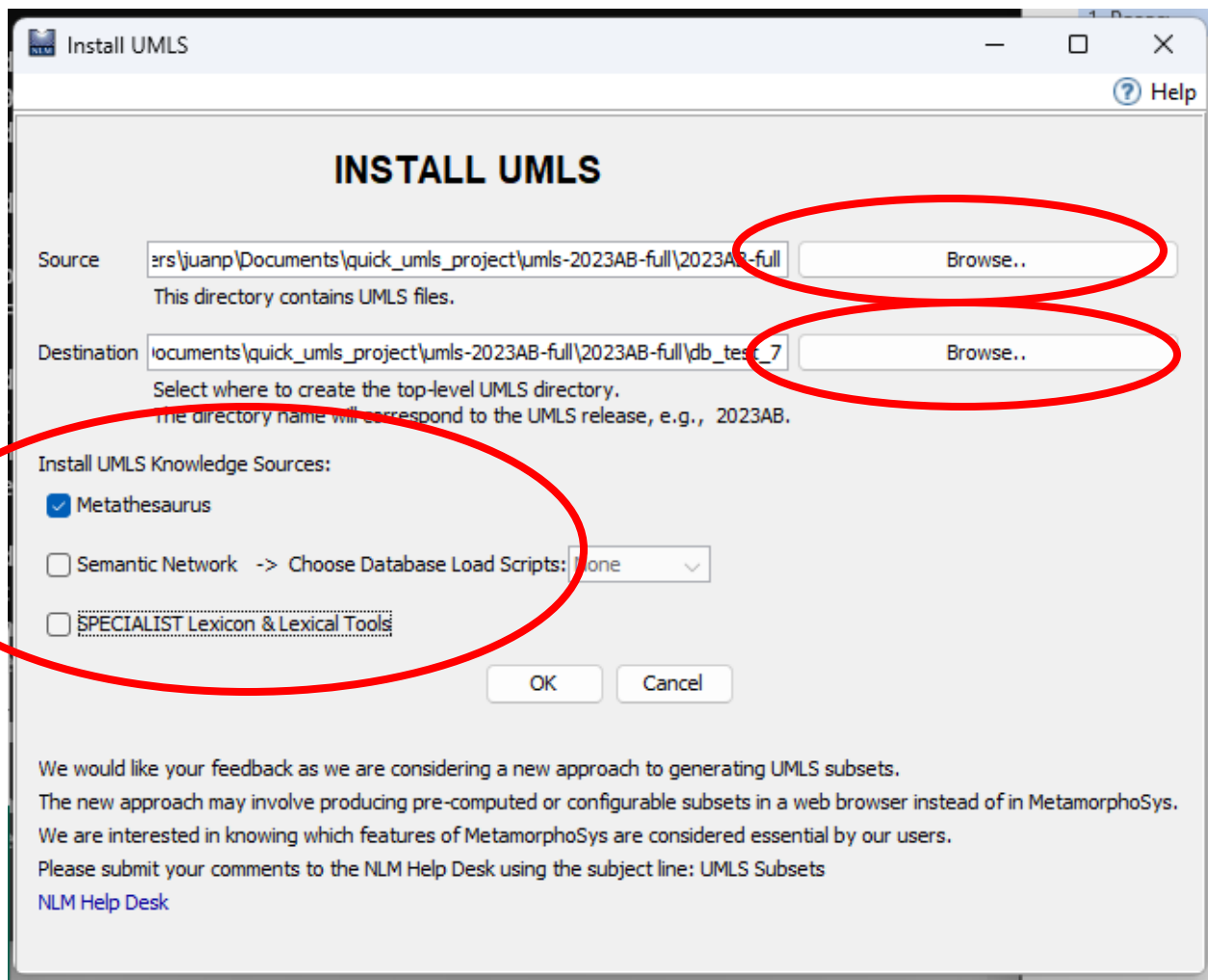
- a. Descomprimir umls-2023AB-full.zip (puede tomar un par de horas descomprimirlos)
- b. Los siguientes archivos deben estar en la carpeta seleccionada de la descompresión:
 - mmsys.zip (zipped MetamorphoSys application)
 - 2016aa-1-meta.nlm (compressed Metathesaurus data)
 - 2016aa-2-meta.nlm (compressed Metathesaurus data)
 - 2016aa-otherks.nlm (compressed Semantic Network and SPECIALIST Lexicon)
 - 2016AA.CHK
 - 2016AA.MD5
 - Copyright_Notice.txt
 - README.txt
- c. Descomprimir el archivo mmsys.zip en la misma carpeta anterior. Debe quedar muy similar a como se muestra en la imagen siguiente. Exceptuando las carpetas db_test.



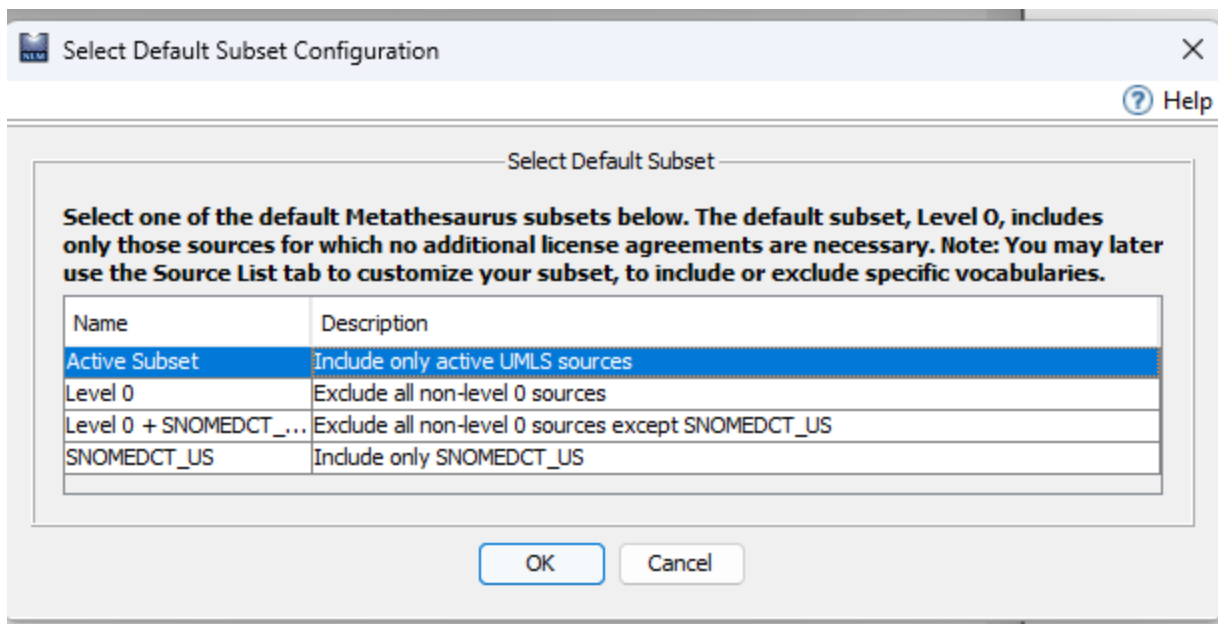
- d. Iniciar metamorphosys. Doble click en alguno de los siguientes archivos ejecutables de acuerdo con el sistema operativo de preferencia, o por medio de una terminal ubicarse en la carpeta anterior y ejecutar alguno de los siguientes comandos:
- ./run.bat or ./run64.bat (Windows)
 - ./run_mac.sh (or click on the run_mac.command file)
 - ./run_linux.sh
- b. Debe aparecer una interfaz visual como la siguiente imagen



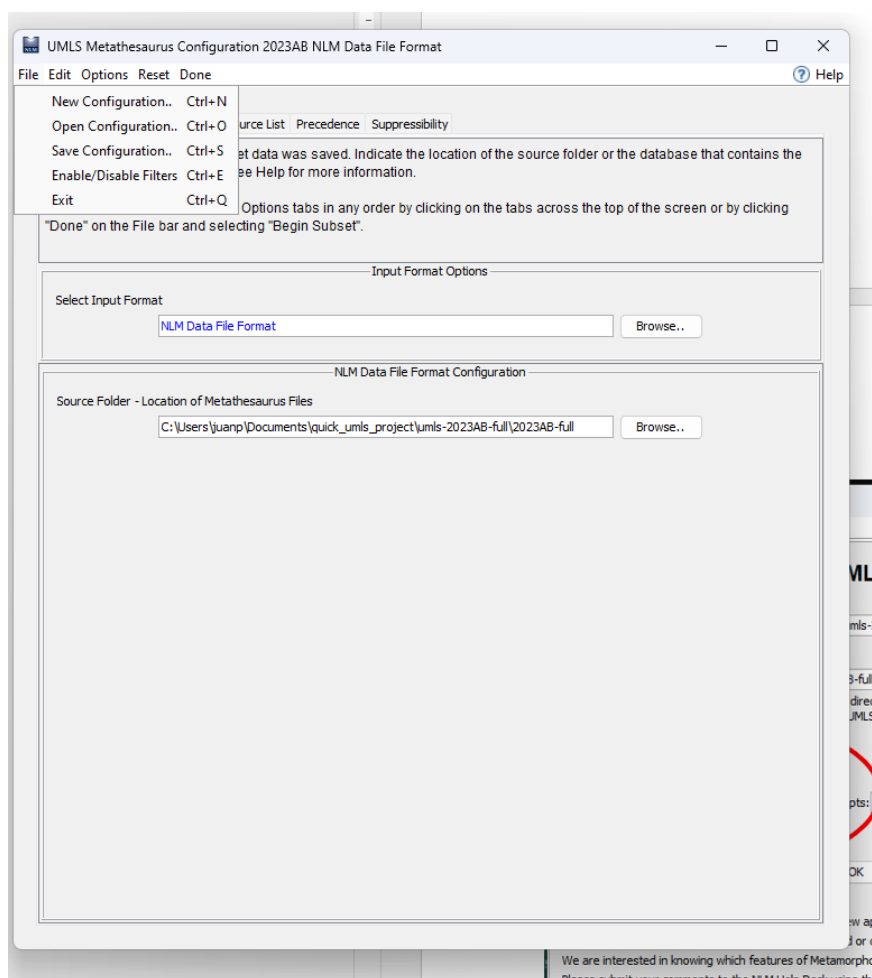
- e. Seleccionar Install UMLS.
- f. En Source seleccionar la carpeta con los archivos descomprimidos.
- g. En Destination seleccionar una carpeta de destino. Sugerencia nombre de la carpeta: db_test
- h. Deseleccionar Semantic Network y SPECIALIST Lexicon & Lexical tools y click en ok



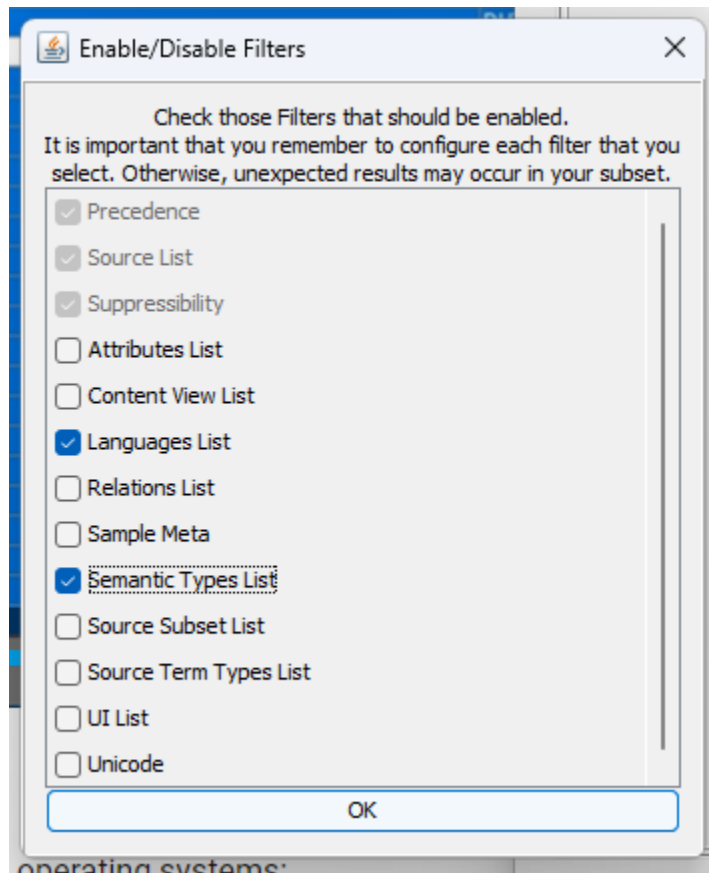
- i. Seleccionar nueva configuración
- j. Aceptar los acuerdos de uso y licencia
- k. Seleccionar solo el Active Subset y click en ok



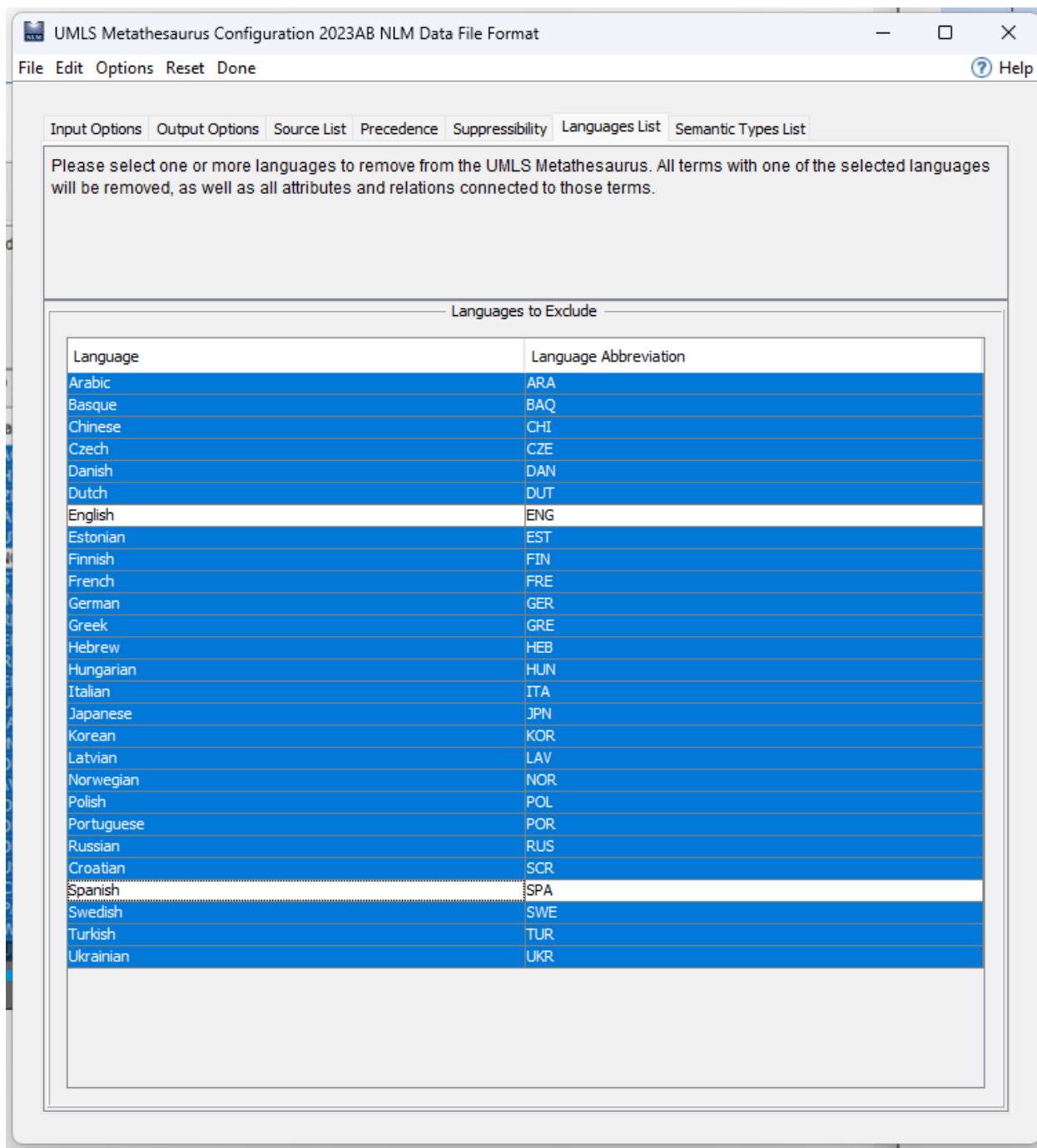
I. Seleccionar Files y Enable/Disable filters. Como se muestra en la siguiente imagen.



- m. Seleccionar Languages List y Semantic Types List. Click en ok



- n. Dejar incluidos los lenguajes ingles y español en la pestaña Languages List. Nota: el componente de quick umls no funciona sin el idioma ingles de base, para usarlo los vocabularios en español se debe tener ambos idiomas en la configuración de umls que se está creando. Ver imagen siguiente.



- o. Seleccionar T033 Finding y T047 Disease or Syndrome en Semantic Types List. En ambos casos ventanas emergentes sugieren vocabularios, dar click en ok y aceptar dichas sugerencias. Asegúrese que al seleccionar un vocabulario no deja de seleccionar otro.

UMLS Metathesaurus Configuration 2023AB NLM Data File Format

File Edit Options Reset Done

Help

Input Options

Output Options

Source List

Precedence

Suppressibility

Languages List

Semantic Types List

Select one or more semantic types below. Your customized Metathesaurus subset will exclude or include any concepts assigned to selected semantic types.

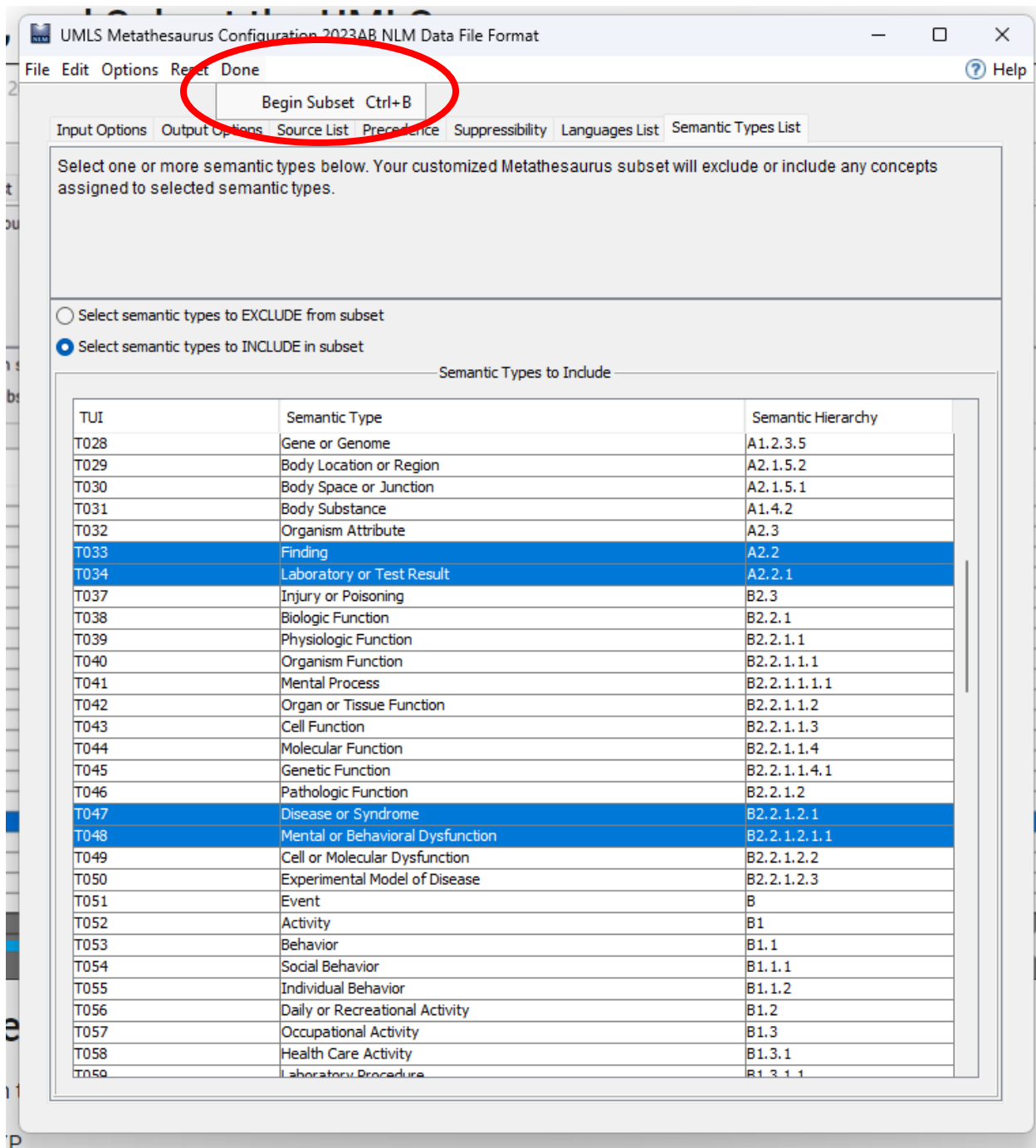
☐ Select semantic types to EXCLUDE from subset

☒ Select semantic types to INCLUDE in subset

Semantic Types to Include

TUI	Semantic Type	Semantic Hierarchy
T028	Gene or Genome	A1.2.3.5
T029	Body Location or Region	A2.1.5.2
T030	Body Space or Junction	A2.1.5.1
T031	Body Substance	A1.4.2
T032	Organism Attribute	A2.3
T033	Finding	A2.2
T034	Laboratory or Test Result	A2.2.1
T037	Injury or Poisoning	B2.3
T038	Biologic Function	B2.2.1
T039	Physiologic Function	B2.2.1.1
T040	Organism Function	B2.2.1.1.1
T041	Mental Process	B2.2.1.1.1.1
T042	Organ or Tissue Function	B2.2.1.1.2
T043	Cell Function	B2.2.1.1.3
T044	Molecular Function	B2.2.1.1.4
T045	Genetic Function	B2.2.1.1.4.1
T046	Pathologic Function	B2.2.1.2
T047	Disease or Syndrome	B2.2.1.2.1
T048	Mental or Behavioral Dysfunction	B2.2.1.2.1.1
T049	Cell or Molecular Dysfunction	B2.2.1.2.2
T050	Experimental Model of Disease	B2.2.1.2.3
T051	Event	B
T052	Activity	B1
T053	Behavior	B1.1
T054	Social Behavior	B1.1.1
T055	Individual Behavior	B1.1.2
T056	Daily or Recreational Activity	B1.2
T057	Occupational Activity	B1.3
T058	Health Care Activity	B1.3.1
T059	Laboratory Procedure	B1.3.1.1

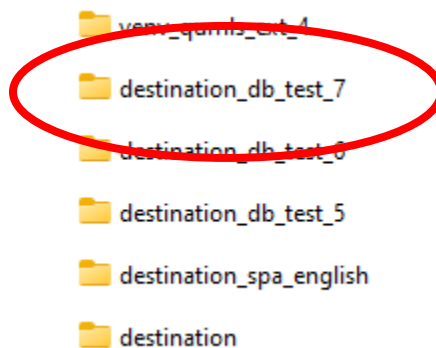
p. Click en Done y Begin Subset



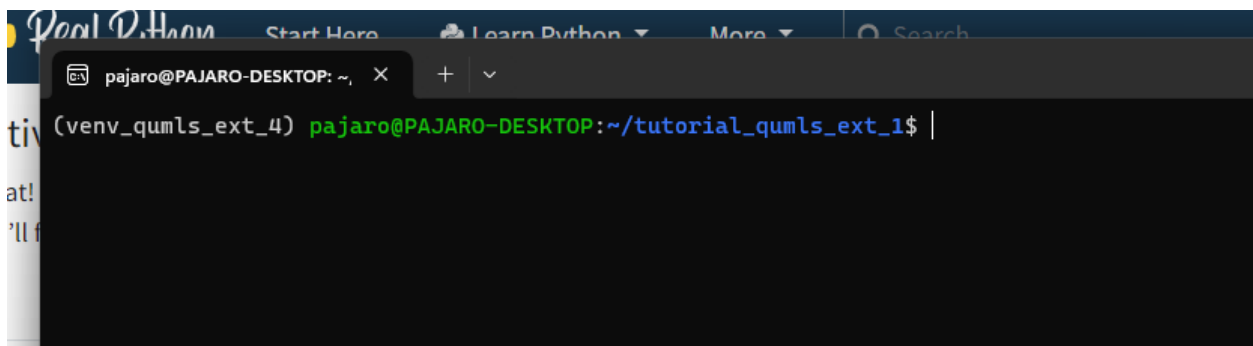
- q. Revisar que los archivos hayan sido creados en la carpeta de destino que selecciono en pasos anteriores. Importante que los archivos MRCONSO.RRF y MRSTY.RRF estén en dicha carpeta.
- r. OPCIONAL: puede seguir o complementar su información con la siguiente guía en: https://www.nlm.nih.gov/research/umls/implementation_resources/metamorphosis/help.html

2. Configurar el componente de quickumls

- a. Crear una subcarpeta de destino, preferiblemente en la dentro de la carpeta proyecto donde se crearán los notebooks (o colabs) con los modelos de aprendizaje automático. Sugerencia <destination_quickumls>
 - Como ejemplo la carpeta creada se denomina <destination_db_test_6> y se encuentra dentro de la carpeta proyecto denominada <tutorial_qumls_ext_1>

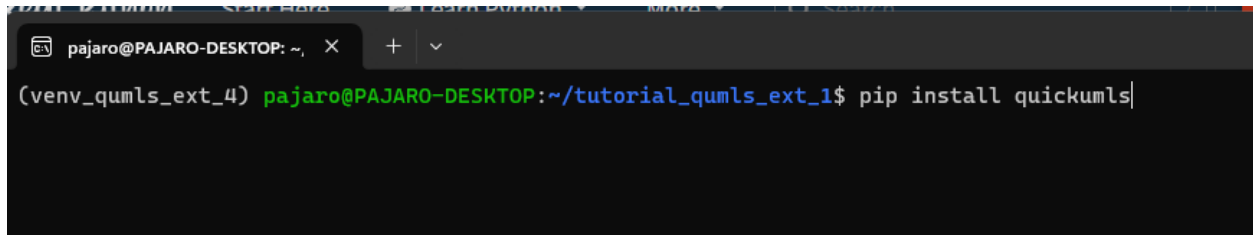


- b. Desde una terminal configurar un ambiente virtual en su carpeta proyecto:
 - Conda: <https://conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#creating-an-environment-with-commands>
 - Python: <https://realpython.com/python-virtual-environments-a-primer/>
- c. activar el ambiente desde la terminal
 - Conda: conda activate <nombre_ambiente_virtual>
 - Python (sistema operativo Linux): source <nombre_ambiente_virtual>/bin/actíivate
 - Python (sistema operativo Windows):
<nombre_ambiente_virtual>\Scripts\activate
 - En su terminal debe verse similar a la siguiente imagen:



d. Instalar quickumls.

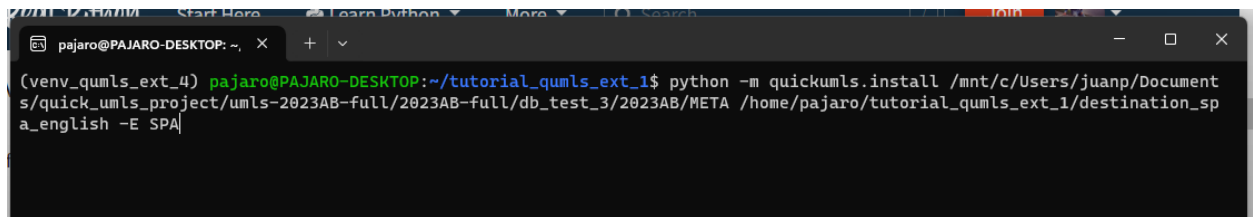
- pip install quickumls
- Ejemplo en la siguiente imagen



```
pajaro@PAJARO-DESKTOP: ~$ pip install quickumls
```

e. Crear una instalación de quickumls.

- python -m quickumls.install <umls_installation_path> <destination_path> -E SPA
- <umls_installation_path> es la carpeta donde realizo todo el paso 1.
- <destination_quickmls> es la carpeta creada al inicio de este paso 2.
- Ejemplo en la siguiente imagen:



```
pajaro@PAJARO-DESKTOP: ~$ python -m quickumls.install /mnt/c/Users/juana/Documents/quick_umls_project/umls-2023AB-full/2023AB-full/db_test_3/2023AB/META /home/pajaro/tutorial_qumls_ext_1/destination_spa_english -E SPA
```

f. Correr el colab hello_qumls_spa.ipynb del repositorio.

g. NOTA IMPORTANTE: Las librerías Medspacy y Spacy deben ser instaladas después de quickumls. Estos pasos se realizan en el colab ejemplo.

h. Revisar los notebooks de medspacy para un correcto uso de la librería

- <https://github.com/medspacy/medspacy/tree/master/notebooks>

Referencias

Eyre, H., Chapman, A. B., Peterson, K. S., Shi, J., Alba, P. R., Jones, M. M., Box, T. L., DuVall, S. L., &

Patterson, O. V. (2022). Launching into clinical space with medspaCy: A new clinical text

processing toolkit in Python. *AMIA Annual Symposium Proceedings, 2021*, 438–447.