This project uses SNOMED CT data to compare terms and their relationships with one another.

1. **SETTING UP IMPORTS:**

NOTE: This project was created using PyCharm. Therefore, needed imports were auto downloaded. The imports for this project are shown below:

ADDITIONAL NOTE: This project was done in python 3.8. It was noted during testing that some packages conflict with any version of python beyond 3.8.

```
from gensim import corpora
from gensim.models import LsiModel, Word2Vec
from nltk.tokenize import RegexpTokenizer
from nltk.stem.porter import PorterStemmer
from gensim import similarities
import json
from os.path import exists
import time
```

2. **SETTING UP FILES:**

This project uses three text documents to initialize:

1. sct2_Concept_Snapshot_INT_20210731
2. sct2_Description_Snapshot-en_INT_20210731
3. sct2_Relationship_Snapshot_INT_20210731

Please save these files locally wherever the script is being run from.

3. **RUNNING:**

Once the text files and imports are set up from the previous 2 steps, initialization will take care of the rest! The concepts will be loaded as well as the relationships

NOTE: After running the first time, the program will save the relationship values as "rel_table.json" and load from this file in future runs

Once initialized, you will be prompted to enter a term as well as it's tag as shown below:

```
Enter term to compare: myocardial infarction
Enter the tag: disorder
```

NOTE: user input is cleaned up to some extent so adding capitalisation and parenthesis will work as well like shown below:

```
Enter term to compare: Myocardial infarction
Enter the tag: (Disorder)
```

Once this input is taken from the user, let the code do the rest! Completion time varies based on number of similar terms and the tag of the chosen term

SAMPLE RUN WITH "MYOCARDIAL INFARCTION (DISORDER)":

```
Initializing data...
Active concepts...
Active concepts loaded!
Active relationships...
Found rel_table.json, Loading...
rel_table.json loaded!
Data loading complete!


Initialization completed after 0.33507680892944336 seconds!


Enter term to compare: myocardial infarction
Enter the tag: disorder
Building Model for myocardial infarction . This won't take long!
Similarity calculation complete after 31.972397565841675
seconds!
Check results in output.txt

Process finished with exit code 0
```

SAMPLE OUTPUT WITH "MYOCARDIAL INFARCTION (DISORDER)":

```
22298006 myocardial infarction
25106000 impending infarction
Similarity Score: 0.9962486

22298006 myocardial infarction
129574000 postoperative myocardial infarction
Similarity Score: 0.99540836

22298006 myocardial infarction
311796008 postoperative subendocardial myocardial infarction
Similarity Score: 0.9953943

22298006 myocardial infarction
903008 chorioretinal infarction
Similarity Score: 0.9951792

22298006 myocardial infarction
703211006 subsequent st segment elevation myocardial infarction
Similarity Score: 0.9939004

22298006 myocardial infarction
233823002 silent myocardial ischemia
Similarity Score: 0.9912474

22298006 myocardial infarction
80475003 thrombotic mesenteric infarction
Similarity Score: 0.9902761

22298006 myocardial infarction
57658009 embolic mesenteric infarction
Similarity Score: 0.98873335

22298006 myocardial infarction
827164008 delayed postmyocardial infarction pericarditis
Similarity Score: 0.9803578

22298006 myocardial infarction
827163002 early postmyocardial infarction pericarditis
```

Similarity Score: 0.9408106

22298006 myocardial infarction
401303003 acute st segment elevation myocardial infarction
Similarity Score: 0.8835001

22298006 myocardial infarction
404234000 st. louis meningitis
Similarity Score: 0.86858207

22298006 myocardial infarction
703209002 subsequent st segment elevation myocardial infarction of inferior wall
Similarity Score: 0.8203378

22298006 myocardial infarction
72150009 nonocclusive intestinal infarction
Similarity Score: 0.8057749

22298006 myocardial infarction
401314000 acute non-st segment elevation myocardial infarction
Similarity Score: 0.8049302

22298006 myocardial infarction
304914007 acute q wave myocardial infarction
Similarity Score: 0.8014688

22298006 myocardial infarction
15713161000119100 acute st segment elevation myocardial infarction of septum
Similarity Score: 0.7895063

22298006 myocardial infarction
62695002 acute anteroseptal myocardial infarction
Similarity Score: 0.7810046

22298006 myocardial infarction
233835003 acute widespread myocardial infarction
Similarity Score: 0.7794961

22298006 myocardial infarction

52035003 acute anteroapical myocardial infarction
Similarity Score: 0.7779336

22298006 myocardial infarction
57054005 acute myocardial infarction
Similarity Score: 0.7771469

22298006 myocardial infarction
703213009 acute st segment elevation myocardial infarction of inferior wall
Similarity Score: 0.76063514

22298006 myocardial infarction
12238111000119106 acute st segment elevation myocardial infarction of inferolateral wall
Similarity Score: 0.7604519

22298006 myocardial infarction
15712961000119108 acute st segment elevation myocardial infarction of anteroseptal wall
Similarity Score: 0.760412

22298006 myocardial infarction
15712881000119105 acute st segment elevation myocardial infarction of anterolateral wall
Similarity Score: 0.76023024

22298006 myocardial infarction
15712841000119100 acute st segment elevation myocardial infarction of posterolateral wall
Similarity Score: 0.7589687

22298006 myocardial infarction
12238151000119107 acute st segment elevation myocardial infarction of inferoposterior wall
Similarity Score: 0.75888824

22298006 myocardial infarction
15713201000119105 acute st segment elevation myocardial infarction of posterobasal wall
Similarity Score: 0.75854313

22298006 myocardial infarction
703210007 subsequent st segment elevation myocardial infarction of anterior wall
Similarity Score: 0.7530788

22298006 myocardial infarction
896697005 acute st segment elevation myocardial infarction of right ventricle
Similarity Score: 0.7208415

22298006 myocardial infarction
15712921000119103 acute st segment elevation myocardial infarction of lateral wall
Similarity Score: 0.7117729

22298006 myocardial infarction
15713041000119103 acute st segment elevation myocardial infarction of posterior wall
Similarity Score: 0.7041735

22298006 myocardial infarction
703164000 acute st segment elevation myocardial infarction of anterior wall
Similarity Score: 0.7009299