How to use scispaCy Entity Linkers for Biomedical Named Entities

Wuraola Oyewusi · Follow

5 min read · Aug 24, 2020

This is a sequel to a previous tutorial by me, there's been an update to scispaCy library

Link to project github
Link to Notebook with uncleared output : This is to help you know if you are getting the right output
Link to Notebook with cleared output

Now that you have done extracted Biomedical Named Entities. How can you make more of the entities? You can link them to a knowledge base(s). This is exactly what the scispaCy entity linkers do.

The choice of knowledge base to link depends on the nature of extracted named entities and the question the user is trying to answer. It is preferable to read further about each database have a good grasp of the possibilities

with them. The names of the knowledge base give an idea of the kind of information accessible with it.

Previous versions supported only one knowledge base but from the library documentation for v2.5.0, five(5) knowledge bases are now supported umls (Unified Medical Language System), mesh (Medical Subject Headings), rxnorm (RxNorm), go(Gene Ontology), hpo(Human Phenotype Ontology).

In this tutorial before linking entities to the available knowledge bases biomedical entities will be extracted from the sample text using the 4 available scispaCy NER models. To read the specificity of each NER model , click this link
All identified entities will then be parsed through different knowledge bases.

Download the required scispaCy models

```
%%capture
!pip install scispacy
!pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.2.5/en_core_sci_md-
0.2.5.tar.gz      #scispacy medium model
!pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.2.5/en_ner_bc5cdr_md-
0.2.5.tar.gz    #biomedical NER model trained on BC5CDR corpus
!pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.2.5/en_ner_bionlp13cg_md-
0.2.5.tar.gz  #biomedical NER model trained on BIONLP13CG corpus
!pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.2.5/en_ner_craft_md-
0.2.5.tar.gz    #biomedical NER model trained on CRAFT corpus
!pip install https://s3-us-west-2.amazonaws.com/ai2-s2-scispacy/releases/v0.2.5/en_ner_jnlpba_md-
0.2.5.tar.gz     #biomedical NER model trained on JNLPBA corpus
```

Download all required models

Import all libraries

Link to sample text: https://www.ncbi.nlm.nih.gov/books/NBK92477/.

Any text sample can be used, data can also be a series of text files.

```
import spacy
import scispacy
from spacy import displacy
import en_core_sci_md
import en_ner_bc5cdr_md
import en_ner_jnlpba_md
import en_ner_craft_md
import en_ner_bionlp13cg_md
from scispacy.abbreviation import AbbreviationDetector
from scispacy.linking import EntityLinker
from collections import OrderedDict,Counter
from pprint import pprint
```

This function displays a tuple of entities extracted by the scispaCy NER models and the st function is to prevent duplication of entities in our dataset. displacy will also helps to visualize the recognized entities.

```
def display_entities(model,document):
    """ A function that returns a tuple of displacy image of named or unnamed word entities and
        a set of unique entities recognized based on scispacy model in use
        Args:
            model: A pretrained model from spaCy or ScispaCy
            document: text data to be analysed"""
    nlp = model.load()
    doc = nlp(document)
    displacy_image = displacy.render(doc, jupyter=True,style='ent')
    entity_and_label = set([(X.text, X.label_) for X in doc.ents])
    return  displacy_image, entity_and_label
```

The images below show the code and output for named entity extraction using different models. as expected the type of entities recognized and extracted is dependent on type of model

```
bc5dr_entities = display_entities(en_ner_bc5cdr_md,sample_text)
bc5dr_entities_dataframe = pd.DataFrame(bc5dr_entities[1],columns=['Entity','Label'])  #save returned
values of entities and their labels in a pandas dataframe
bc5dr_entities_dataframe['Ner_model'] = 'bc5dr'  #include a column with constant value of NER model
bc5dr_entities_dataframe
```

scispaCy NER based on bc5dr corpus

```
1  bc5dr_entities = display_entities(en_ner_bc5cdr_md,sample_text)
```

CORONAVIRUS DISEASE RESEARCH: KEYS CHEMICAL TO DIAGNOSIS, TREATMENT, AND PREVENTION OF SARS DISEASE

Mark R. Denison, M.D.

For coronavirus investigators, the recognition of a new coronavirus as the cause of severe acute respiratory syndrome DISEASE (SARS DISEASE) was certainly remarkable, yet perhaps not surprising (Baric et al., 1995). The cadre of investigators who have worked with this intriguing family of viruses over the past 30 years are familiar with many of the features of coronavirus biology, pathogenesis, and disease that manifested so dramatically in the worldwide SARS DISEASE epidemic. Advances in the biology of coronaviruses have resulted in greater understanding of their capacity for adaptation to new environments, transspecies infection DISEASE , and emergence of new diseases. New tools of cell and molecular biology have led to increased understanding of intracellular replication and viral cell biology, and the advent in the past five years of reverse genetic approaches to study coronaviruses has made it possible to begin to define the determinants of viral replication, transpecies adaptation, and human disease DISEASE . This summary will discuss the basic life cycle and replication of the well-studied coronavirus, mouse hepatitis virus DISEASE (MHV), identifying the unique characteristics of coronavirus biology and highlighting critical points where research has made significant advances, and which might represent targets for antivirals or vaccines. Areas where rapid progress has been made in SCoV research will be described. Finally, areas of need for research in coronavirus replication, genetics, and pathogenesis will be summarized.

Coronavirus Life Cycle

```
bionlp13cg_entities_dataframe = pd.DataFrame(bionlp13cg_entities[1],columns=['Entity','Label']) #save returned values of entities and their labels in a pandas dataframe
bionlp13cg_entities_dataframe['Ner_model'] = 'bionlp13cg'  #include a column with constant value of NER model
bionlp13cg_entities_dataframe
```

scispaCy NER based on bioblp13cg corpus

```
1  bionlp13cg_entities = display_entities(en_ner_bionlp13cg_md,sample_text)
```

CORONAVIRUS RESEARCH: KEYS SIMPLE_CHEMICAL TO DIAGNOSIS, TREATMENT, AND PREVENTION OF SARS

Mark R. Denison, M.D.

For coronavirus investigators, the recognition of a new coronavirus ORGANISM as the cause of severe acute respiratory syndrome (SARS) was certainly remarkable, yet perhaps not surprising (Baric et al., 1995). The cadre of investigators who have worked with this intriguing family of viruses over the past 30 years are familiar with many of the features of coronavirus ORGANISM biology, pathogenesis, and disease that manifested so dramatically in the worldwide SARS epidemic. Advances in the biology of coronaviruses ORGANISM have resulted in greater understanding of their capacity for adaptation to new environments, transspecies infection, and emergence of new diseases. New tools of cell CELL and molecular biology have led to increased understanding of intracellular IMMATERIAL_ANATOMICAL_ENTITY replication and viral cell CELL biology, and the advent in the past five years of reverse genetic approaches to study coronaviruses ORGANISM has made it possible to begin to define the determinants of viral replication, transpecies adaptation, and human ORGANISM disease. This summary will discuss the basic life cycle and replication of the well-studied coronavirus, mouse hepatitis virus ORGANISM (MHV ORGANISM), identifying the unique characteristics of coronavirus ORGANISM biology and highlighting critical points where research has made significant advances, and which might represent targets for antivirals or vaccines. Areas where rapid progress has been made in SCoV research will be described. Finally, areas of need for research in coronavirus ORGANISM replication, genetics, and pathogenesis will be summarized.

Coronavirus Life Cycle

The best studied model for coronavirus ORGANISM replication and pathogenesis has been the group 2 murine coronavirus ORGANISM , mouse hepatitis virus ORGANISM , and much of what is known of the stages of the coronavirus ORGANISM life cycle has been determined in animals and in culture using this virus. Thus this discussion will focus on MHV ORGANISM with comparisons to SCoV and other coronaviruses ORGANISM . This is appropriate because bioinformatics analyses suggest that SCoV, while a distinct virus, has significant similarities in organization, putative protein functions, and replication to the group II coronaviruses, particularly within the replicase GENE_OR_GENE_PRODUCT gene (Snijder et al., ORGANISM 2003). Excellent, detailed reviews of MHV ORGANISM and coronavirus ORGANISM replication are available elsewhere (Holmes and Lai, 1996; Lai and Cavanagh, 1997).

The coronavirus ORGANISM virion is an enveloped particle containing the spike (S), membrane CELLULAR_COMPONENT (M), and envelope (E) proteins. In addition, some strains of coronaviruses ORGANISM , but not SCoV, express a

```
craft_entities_dataframe = pd.DataFrame(craft_entities[1],columns=['Entity','Label'])  #save returned values of entities and their labels in a pandas dataframe
craft_entities_dataframe['Ner_model'] = 'craft' #include a column with constant value of NER model
craft_entities_dataframe
```

scispaCy NER based on craft corpus

```
1  craft_entities = display_entities(en_ner_craft_md,sample_text)
```

CORONAVIRUS RESEARCH: KEYS TO DIAGNOSIS, TREATMENT, AND PREVENTION OF SARS

Mark R. Denison, M.D.

For coronavirus investigators, the recognition of a new coronavirus as the cause of severe acute respiratory syndrome (SARS) was certainly remarkable, yet perhaps not surprising (Baric et al., 1995). The cadre of investigators who have worked with this intriguing family of viruses TAXON over the past 30 years are familiar with many of the features of coronavirus biology, pathogenesis, and disease that manifested so dramatically in the worldwide SARS epidemic. Advances in the biology of coronaviruses have resulted in greater understanding of their capacity for adaptation to new environments, transspecies infection, and emergence of new diseases. New tools of cell CL and molecular CHEBI biology have led to increased understanding of intracellular GO replication and viral cell TAXON biology, and the advent in the past five years of reverse genetic SO approaches to study coronaviruses has made it possible to begin to define the determinants of viral TAXON replication, transpecies adaptation, and human TAXON disease. This summary will discuss the basic life cycle and replication of the well-studied coronavirus, mouse TAXON hepatitis virus TAXON (MHV), identifying the unique characteristics of coronavirus biology and highlighting critical points where research has made significant advances, and which might represent targets for antivirals CHEBI or vaccines. Areas where rapid progress has been made in SCoV research will be described. Finally, areas of need for research in coronavirus replication, genetics, and pathogenesis will be summarized.

Coronavirus Life Cycle

The best studied model for coronavirus replication and pathogenesis has been the group 2 murine TAXON coronavirus, mouse TAXON hepatitis virus TAXON, and much of what is known of the stages of the coronavirus life cycle has been determined in animals TAXON and in culture using this virus TAXON. Thus this discussion will focus on MHV with comparisons to SCoV and other coronaviruses. This is appropriate because bioinformatics analyses suggest that SCoV, while a distinct virus TAXON, has significant similarities in organization, putative protein CHEBI functions, and replication to the group II coronaviruses, particularly within the replicase gene SO (Snijder et al., 2003). Excellent, detailed reviews of MHV and coronavirus replication are available elsewhere (Holmes and Lai, 1996; Lai and Cavanagh, 1997).

```
jnlpa_entities_dataframe = pd.DataFrame(jnlpba_entities[1],columns=['Entity','Label']) #save returned values of entities and their labels in a pandas dataframe
jnlpa_entities_dataframe['Ner_model'] = 'jnlpa' # include a column with constant value of NER model
jnlpa_entities_dataframe
```

scispaCy NER based on jnlpa corpus

```
[27]  jnlpba_entities = display_entities(en_ner_jnlpba_md,sample_text)
```

CORONAVIRUS RESEARCH: KEYS TO DIAGNOSIS, TREATMENT, AND PREVENTION OF SARS

Mark R. Denison, M.D.

For coronavirus investigators, the recognition of a new coronavirus as the cause of severe acute respiratory syndrome (SARS) was certainly remarkable, yet perhaps not surprising (Baric et al., 1995). The cadre of investigators who have worked with this intriguing family of viruses over the past 30 years are familiar with many of the features of coronavirus biology, pathogenesis, and disease that manifested so dramatically in the worldwide SARS epidemic. Advances in the biology of coronaviruses have resulted in greater understanding of their capacity for adaptation to new environments, transspecies infection, and emergence of new diseases. New tools of cell and molecular biology have led to increased understanding of intracellular replication and viral cell biology, and the advent in the past five years of reverse genetic approaches to study coronaviruses has made it possible to begin to define the determinants of viral replication, transpecies adaptation, and human disease. This summary will discuss the basic life cycle and replication of the well-studied coronavirus, mouse hepatitis virus (MHV), identifying the unique characteristics of coronavirus biology and highlighting critical points where research has made significant advances, and which might represent targets for antivirals or vaccines. Areas where rapid progress has been made in SCoV PROTEIN research will be described. Finally, areas of need for research in coronavirus replication, genetics, and pathogenesis will be summarized.

Coronavirus Life Cycle

The best studied model for coronavirus replication and pathogenesis has been the group 2 murine coronavirus, mouse hepatitis virus, and much of what is known of the stages of the coronavirus life cycle has been determined in animals and in culture using this virus. Thus this discussion will focus on MHV PROTEIN with comparisons to SCoV PROTEIN and other coronaviruses. This is appropriate because bioinformatics analyses suggest that SCoV PROTEIN, while a distinct virus, has significant similarities in organization, putative protein functions, and replication to the group II coronaviruses, particularly within the replicase gene DNA (Snijder et al., 2003). Excellent, detailed reviews of MHV and coronavirus replication are available elsewhere (Holmes and Lai, 1996; Lai and Cavanagh, 1997).

The coronavirus virion is an enveloped particle containing the spike (S), membrane (M), and envelope (E) proteins PROTEIN. In addition, some strains of coronaviruses, but not SCoV PROTEIN, express a hemagglutinin protein PROTEIN

A total of 422 biomedical named entities were extracted from the sample corpus using 4 NER models from scispaCy.

Concatenate all the extracted entities and save the data for future use.

```
entities_and_label_from_4_NER_model_dataframe =
pd.concat([bc5dr_entities_dataframe,bionlp13cg_entities_dataframe,craft_entities_dataframe,jnlpa_entitie
s_dataframe])
#Concatenate all pandas dataframe into one.
entities_and_label_from_4_NER_model_dataframe.to_csv('entities_and_label_from_4_scispacy_NER_models.csv'
, index=False) #Save dataframe to csv
entities_and_label_from_4_NER_model_dataframe.info()
```

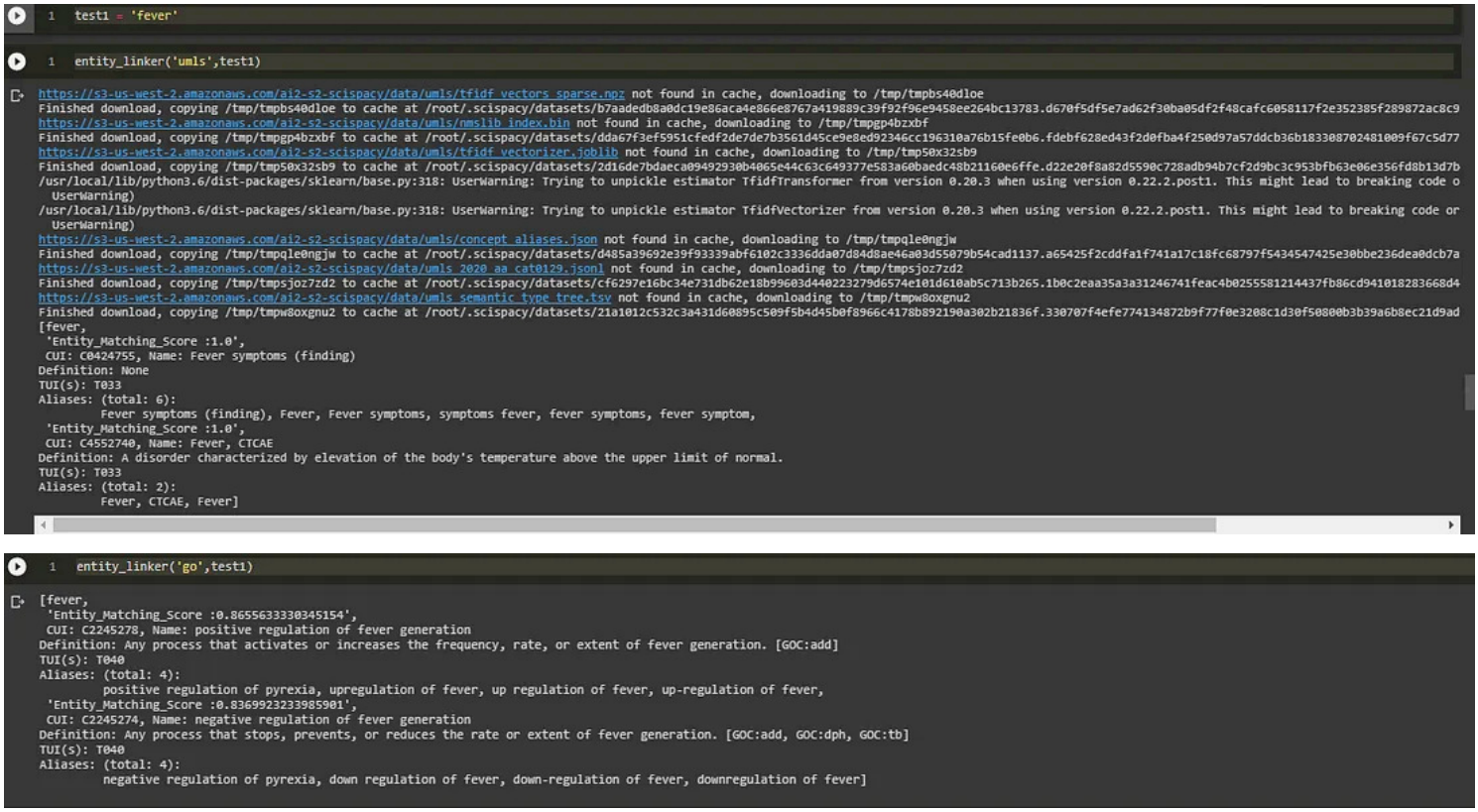
The function below is a general function to link biomedical entities to the scispaCy knowledge bases.

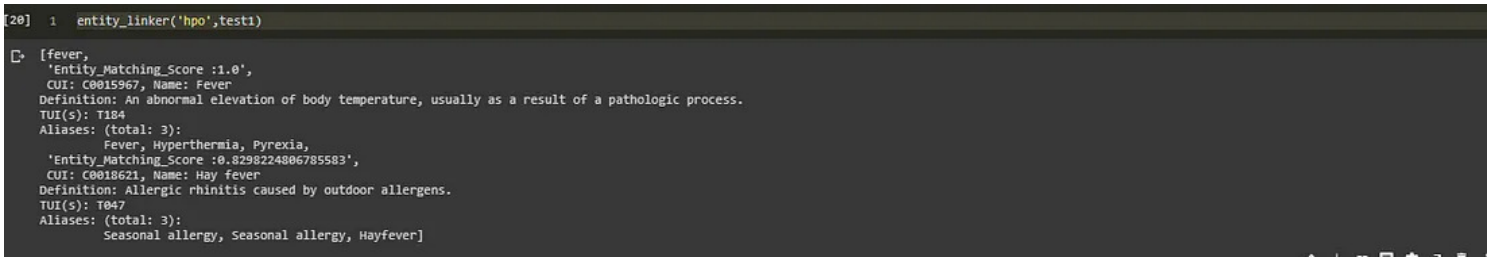
```
def entity_linker(linker_name,document):
    """ A function that accepts the name of a scispacy knowledge base and documents and returns the
entity link details"""
    linker = EntityLinker(k = 10,max_entities_per_mention = 2, name=linker_name)  #parameters are
tunable,so it can be set to return more than 2 entity matches
    nlp = en_core_sci_sm.load()
    nlp.add_pipe(linker)
    doc = nlp(document)
    try:
        entity = doc.ents[0]
    except IndexError:
        entity = 'Nan'
    entity_details = []
    entity_details.append(entity)
    try:
        for linker_ent in entity._.kb_ents:
            Concept_Id, Score = linker_ent
            entity_details.append('Entity_Matching_Score :{}'.format(Score))
            entity_details.append(linker.kb.cui_to_entity[linker_ent[0]])
    except AttributeError:
        pass
    return entity_details
```

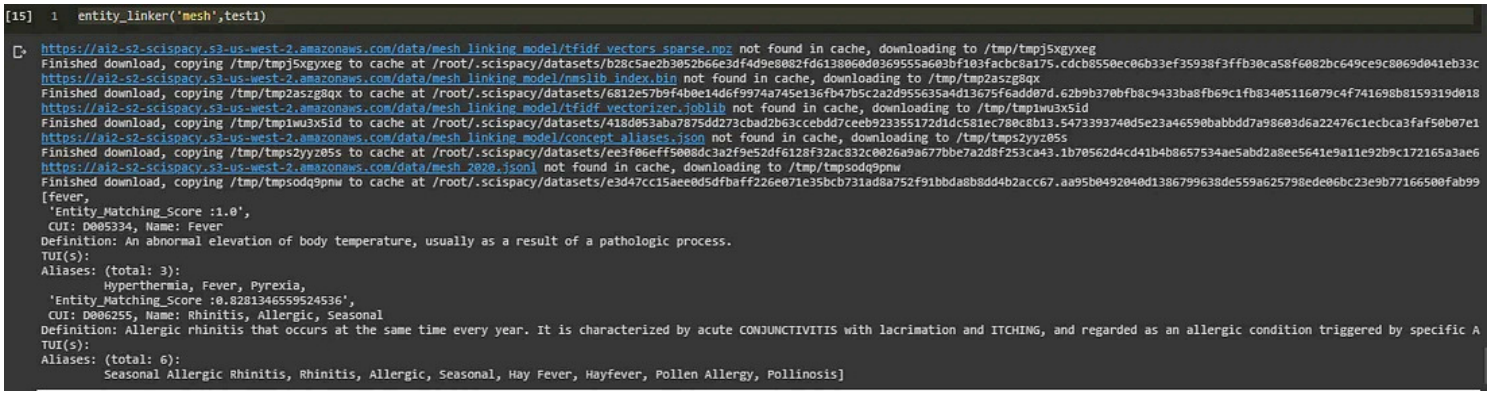
One of the goals of this tutorial is to show how different knowledge bases can return different entity linkage based on the type of data the knowledge base is designed for. Here, the same entity is parsed by four knowledge bases and they returned different concepts, matching score and definitions.

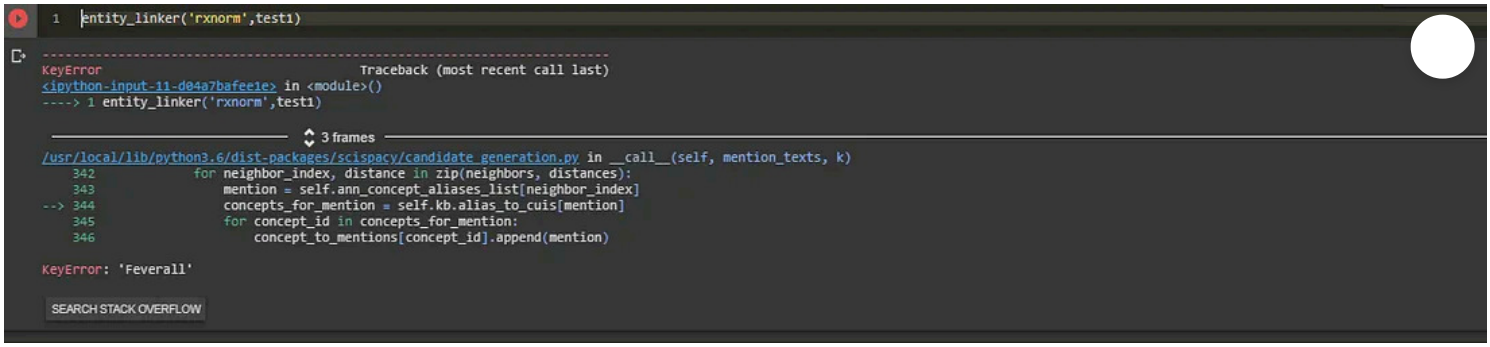
The entity_linker function was tested with the 4 sciSpacy knowledge bases “umls”,” mesh”,”go”,”hpo”. The function will return 2 entities and their scores as it relates to the Knowledge base. As at the time of this writing the “rxnorm” knowledge base returned KeyError, I will investigate further on why this is happening

The images below show the general entity linker in action.









To apply the entity linker to all entities in the pandas dataframe. Readers have probably come across the concept of optimizing code. The code below shows some of the lines of code being moved out of the entity_linker function and adjusted to be able to link to one database.

If readers can, they can compare the difference between using the code with certain lines outside the function or the general entity linker used above. (It is an interesting difference that answers the question of when functions should be a general function or tweaked.

```
[22] 1 #Load pre extracted dataset into pandas
     2 entities_and_label_from_4_NER_model_dataframe = pd.read_csv('/content/entities_and_label_from_4_scispacy_NER_models.csv')
```

```
[23] 1 #pd.options.display.max_colwidth = 1000      #increase display width of the pandas dataframe
     2 entities_and_label_from_4_NER_model_dataframe.head()
```

	Entity	Label	Ner_model
0	viroplexis	DISEASE	bc5dr
1	productive infection	DISEASE	bc5dr
2	gastroenteritis	DISEASE	bc5dr
3	hepatitis virus	DISEASE	bc5dr
4	angiotensin	CHEMICAL	bc5dr

```
linker = EntityLinker(k = 10,max_entities_per_mention = 2, name='umls')  #parameters are tunable,so it
```



```
    #nlp = en_core_sci_sm.load()
    #nlp.add_pipe(linker)
    doc = nlp(document)
    try:
        entity = doc.ents[0]
    except IndexError:
        entity = 'Nan'
    entity_details = []
    entity_details.append(entity)
    try:
        for linker_ent in entity._.kb_ents:
            Concept_Id, Score = linker_ent
            entity_details.append('Entity_Matching_Score :{}'.format(Score))
            entity_details.append(linker.kb.cui_to_entity[linker_ent[0]])
    except AttributeError:
        pass
    return entity_details
```

```
linker = EntityLinker(k = 10,max_entities_per_mention = 2, name='mesh')  #parameters are tunable,so it
can be set to return more than 2 entity matches
nlp = en_core_sci_sm.load()
nlp.add_pipe(linker)
def mesh_entity_linker(document):
    """ A function that accepts the name of a scispacy knowledge base and documents and returns the
entity link details"""
    #linker = EntityLinker(k = 10,max_entities_per_mention = 2, name=linker_name)  #parameters are
tunable,so it can be set to return more than 2 entity matches
    #nlp = en_core_sci_sm.load()
    #nlp.add_pipe(linker)
    doc = nlp(document)
    try:
        entity = doc.ents[0]
    except IndexError:
        entity = 'Nan'
    entity_details = []
    entity_details.append(entity)
    try:
        for linker_ent in entity._.kb_ents:
            Concept_Id, Score = linker_ent
            entity_details.append('Entity_Matching_Score :{}'.format(Score))
            entity_details.append(linker.kb.cui_to_entity[linker_ent[0]])
    except AttributeError:
        pass
    return entity_details
```

```
[14]  1  entities_and_label_from_4_NER_model_dataframe['umls_output'] = entities_and_label_from_4_NER_model_dataframe['Entity'].swifter.apply(lambda x : umls_entity_linker(x))

Pandas Apply: 100% 422/422 [02:24<00:00, 2.91it/s]

[16]  1  entities_and_label_from_4_NER_model_dataframe['mesh_output'] = entities_and_label_from_4_NER_model_dataframe['Entity'].swifter.apply(lambda x : mesh_entity_linker(x))

Pandas Apply: 100% 422/422 [01:11<00:00, 5.87it/s]

[10]  1  entities_and_label_from_4_NER_model_dataframe['go_output'] = entities_and_label_from_4_NER_model_dataframe['Entity'].swifter.apply(lambda x : go_entity_linker(x))

Pandas Apply: 100% 422/422 [04:03<00:00, 1.74it/s]

[19]  1  entities_and_label_from_4_NER_model_dataframe['hpo_output'] = entities_and_label_from_4_NER_model_dataframe['Entity'].swifter.apply(lambda x : hpo_entity_linker(x))

Pandas Apply: 100% 422/422 [00:05<00:00, 73.65it/s]
```

This is the view of the resulting dataframe showing what each entity links to in the available knowledge bases(s). Some entities had definitions to link to in the 4 scispaCy knowledge bases connected to.

```
[21] 1 pd.options.display.max_colwidth = 1000      #increase display width of the pandas dataframe
     2 entities_and_label_from_4_NER_model_dataframe.head()
```

	Entity	Label	Ner_model	go_output	umls_output	mesh_output	hpo_output
0	viroplexis	DISEASE	bc5dr	[Nan]	[Nan]	[Nan]	[Nan]
1	productive infection	DISEASE	bc5dr	[(productive, infection)]	[(productive, infection), Entity_Matching_Score :0.822649121284849, (C2314882, Reproductive tract infection, [Infection, Reproductive Tract, Reproductive Tract Infection, reproductive tract infection, Infections, Reproductive Tract, Reproductive Tract Infections, Genital Tract Infections, Genital Tract Infection, Infection, Genital Tract, Infections, Genital Tract], [T047], Infections of the genital tract in females or males. They can be caused by endogenous, iatrogenic, or sexually transmitted organisms.), Entity_Matching_Score :0.7882543802261353, (C1171315, reproductive system infection, [], [T047], invasion and growth of microorganisms in the male or female reproductive system; may be clinically inapparent or result in local cellular injury.)]	[(productive, infection), Entity_Matching_Score :0.7774598598480225, (D060737, Reproductive Tract Infections, [Reproductive Tract Infections, Genital Tract Infections], [], Infections of the genital tract in females or males. They can be caused by endogenous, iatrogenic, or sexually transmitted organisms.), Entity_Matching_Score :0.7372956275939941, (D004467, Economics, [Easterlin Hypothesis, Macroeconomic Factors, Capital, Cost of Living, Remittances, Microeconomic Factors, Factors, Economic, Consumption, Consumer Price Index, Policies, Economic, Utility Theory, Economic Policy, Economic Policies, Home Economics, Economics, Household Consumption, Economic Conditions, Economics, Home, Production, Policy, Economic, Economic Factors, Conditions, Economic], [], The science of utilization, distribution, and consumption of services and materials.)]	[(productive, infection)]
2	gastroenteritis	DISEASE	bc5dr	[(gastroenteritis)]	[(gastroenteritis), Entity_Matching_Score :1.0, (C0017160, Gastroenteritis, [Gastroenteritis, Gastroenteritis, Gastroenteritis, Gastroenteritis, Gastroenteritis, Gastroenteritis, Gastroenteritis, Gastroenteritis, gastroenteritis, gastroenteritis, Gastroenteritides, gastroenteritides, gastroenteritis nos, GE - Gastroenteritis, Gastrointestinal tract inflamed, Gastroenteritis (disorder), Inflammation of the mucous membrane of the stomach and bowel], [T047], INFLAMMATION of any segment of the GASTROINTESTINAL TRACT from ESOPHAGUS to RECTUM. Causes of gastroenteritis are many including genetic, infection, HYPERSENSITIVITY, drug effects, and	[(gastroenteritis), Entity_Matching_Score :1.0, (D005759, Gastroenteritis, [Gastroenteritis], [], INFLAMMATION of any segment of the GASTROINTESTINAL TRACT from ESOPHAGUS to RECTUM. Causes of gastroenteritis are many including genetic, infection, HYPERSENSITIVITY, drug effects, and CANCER.), Entity_Matching_Score :0.7732845544815063, (D006517, Murine hepatitis virus, [Murine coronavirus, Hepatitis Virus, Mouse,	[(gastroenteritis), Entity_Matching_Score :0.8153096437454224, (C3808828, Recurrent gastroenteritis, [], [T033], Increased susceptibility to gastroenteritis, an infectious inflammationof the stomach and small intestines manifested by signs and symptoms such as diarheas and abdominal pain, as manifested by recurrent episodes of gastroenteritis. []), Entity_Matching_Score :0.7275374531745911, (C1262481,

This is a tutorial showing how to extract biomedical and clinical entities and link to medical knowledge base(s) using the scispaCy python library.
I hope this answers some of your questions and you have a great time exploring.
Did you also learn a couple of things about pandas and swifter?

You can find more about scispaCy here

Scispacy | Swifter | Entity Linker | Biomedical Entities | Named Entity Recognition

Written by Wuraola Oyewusi

Follow

347 Followers

Data Scientist | AI Researcher | Pharmacist

More from Wuraola Oyewusi

Wuraola Oyewusi

How to use ScispaCy for Biomedical Named Entity...

scispaCy is a Python package containing spaCy models for processing biomedical,...

3 min read · Aug 11, 2019

139 6

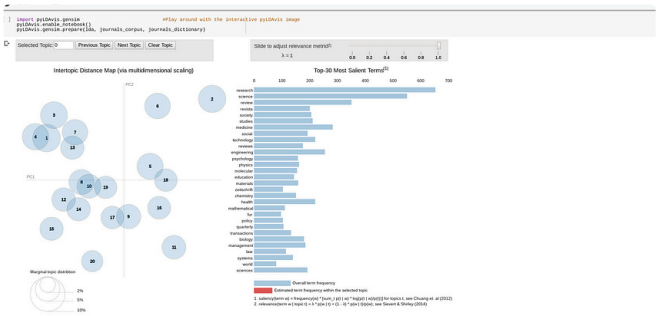
Wuraola Oyewusi

From Pharmacy to Data Science: All the cool things about career...

Tonight, I'm annotating a dataset for custom named entity recognition(ner) and the range...

3 min read · Feb 6, 2020

555 3



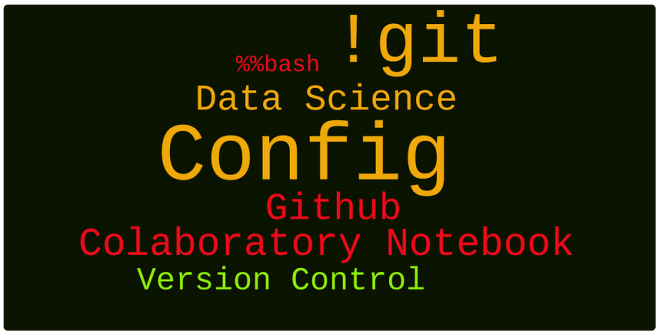
Wuraola Oyewusi

Exploring Topic Modelling with Gensim on the Essential Science...

The Data Set : As stated on the Thomas Reuters website. ‘Essential Science Indicator...

4 min read · Jan 25, 2019

32 1



Wuraola Oyewusi

How to work with Github in Google Colaboratory Notebook

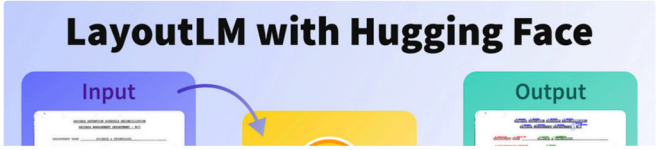
This is one of those things I’m writing because I didn’t find exactly what I wanted when I...

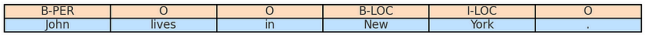
2 min read · Aug 1, 2019

25 1

See all from Wuraola Oyewusi

Recommended from Medium





Wiem Souai in UBIAI NLP

Relationship Extraction with LayoutLM: Transforming Docume...

In the contemporary digital landscape, the ability to swiftly and accurately decipher...

13 min read · Mar 1, 2024

1

Ahmet Münir Kocaman

Mastering Named Entity Recognition with BERT: A...

Introduction

11 min read · Oct 6, 2023

68

Lists

data science and AI

40 stories · 112 saves

Natural Language Processing

1320 stories · 811 saves

PyMuPDF

RAG/LLM and PDF: Enhanced Text Extraction

PyMuPDF delivers enhanced context for RAG environments

5 min read · Mar 18, 2024

58

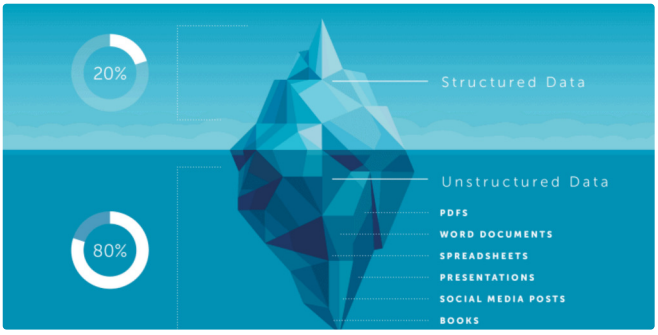
Ravi Chandra Jammalamadaka

Are LLMs Good Spell Checkers?

Recently, I contemplated the following question: Leveraging the capabilities of LLM...

6 min read · Feb 4, 2024

50 1



Johni Douglas Marangon

Building a custom Named Entity Recognition model using spaCy —...

Welcome to the first post about training a custom Named-Entity Recognition (NER) wit...

5 min read · Nov 21, 2023

47

Sanskrutikhedkar

Mastering Information Extraction from Unstructured Text: A Deep...

In today's data-driven landscape, the real treasure often lies buried within unstructure...

5 min read · Oct 17, 2023

453

See more recommendations