MIMIC III NOTE SUMMARIZATION TUTORIAL

AI in Healthcare

MIMIC III Note Sumarization Tutorial

The Problem:

Doctors have very limited time to go over every note on a patient's file and it would be helpful if they could search specific terms and get back only the most relevant information related to that term.

Note:

In this exercise we will focus on patients diagnosed with PTSD who have at some point reported having nightmares.

Code can be found here: https://github.com/nelabdiel/AIinHC/blob/main/MIMIC_III_NLP_Tutorial-Code.ipynb

In this tutorial you will learn to:

- Create your own labels with SpaCy and Displacy

In this tutorial you will learn to:

- Create your own labels with SpaCy and Displacy
- Create a Simple Machine Learning model to extract the most relevant information from all the notes on a patient's file using TF-IDF and Cosine Similarity

In this tutorial you will learn to:

- Create your own labels with SpaCy and Displacy
- Create a Simple Machine Learning model to extract the most relevant information from all the notes on a patient's file using TF-IDF and Cosine Similarity
- Extract the most relevant information for a given keyword from all patient's notes.

Data Processing

Data Processing steps:

- Load all the notes, and all the diagnostics

Data Processing steps:

- Load all the notes, and all the diagnostics

- Find all the admissions (in diagnoses) related to PTSD (icd9 = '**30981**')

Data Processing steps:

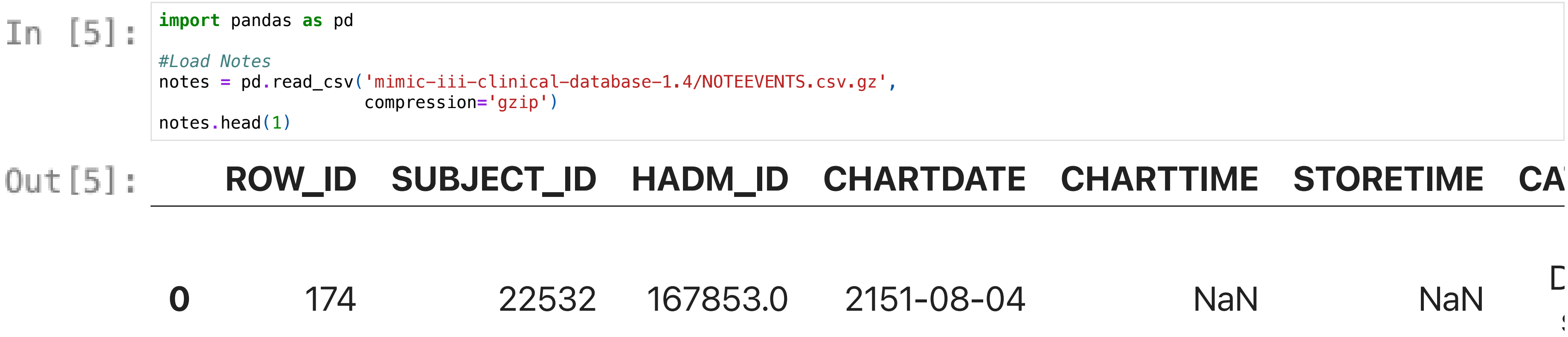
- Load all the notes, and all the diagnostics
- Find all the admissions (in diagnoses) related to PTSD (icd9 = '**30981**')
- Find all the notes corresponding to these admissions

Data Processing steps:

- Load all the notes, and all the diagnostics
- Find all the admissions (in diagnoses) related to PTSD (icd9 = '**30981**')
- Find all the notes corresponding to these admissions
- Find all the notes corresponding to these admissions that mention the word '**nightmares**'

Load all the necessary data

Load all the necessary data



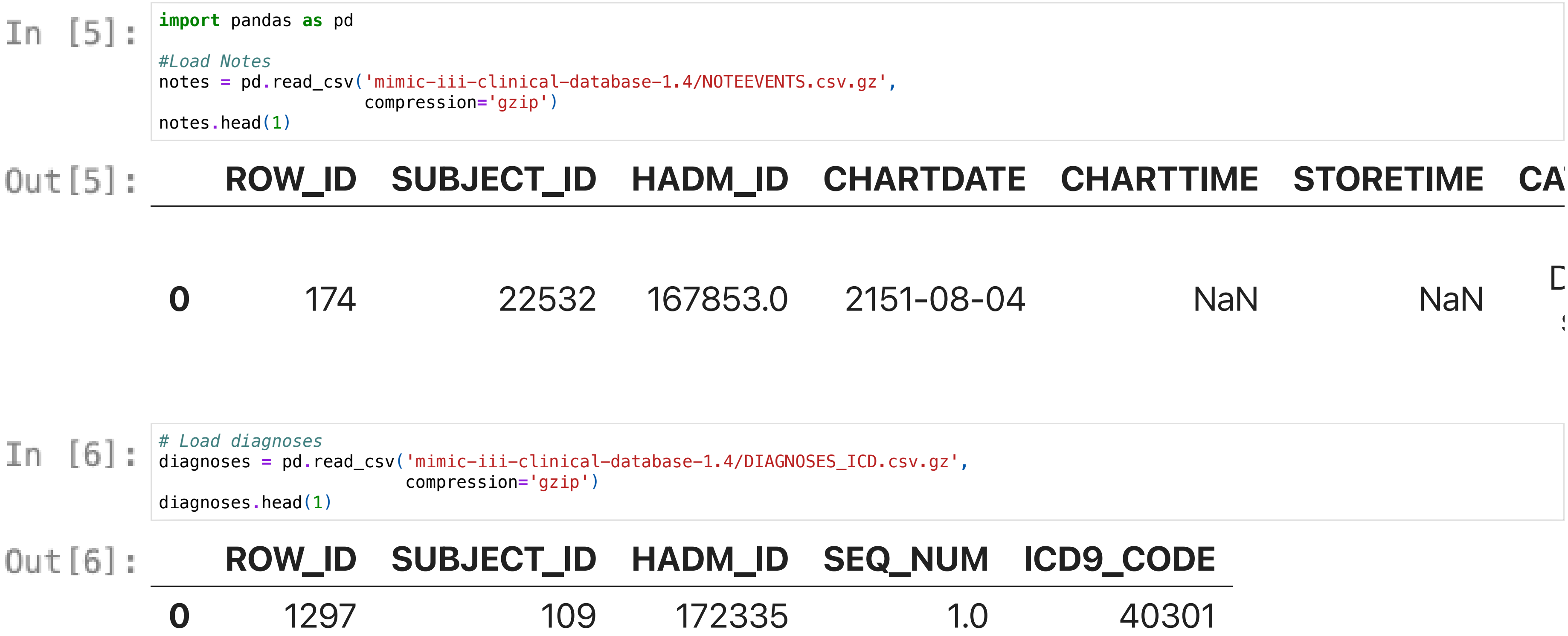
In [5]:

```
import pandas as pd

#Load Notes
notes = pd.read_csv('mimic-iii-clinical-database-1.4/NOTEEVENTS.csv.gz',
                    compression='gzip')
notes.head(1)
```

Out[5]:

	ROW_ID	SUBJECT_ID	HADM_ID	CHARTDATE	CHARTTIME	STORETIME	CA
0	174	22532	167853.0	2151-08-04	NaN	NaN	D

Load all the necessary data



In [5]:

```python
import pandas as pd

#Load Notes
notes = pd.read_csv('mimic-iii-clinical-database-1.4/NOTEEVENTS.csv.gz',
                    compression='gzip')
notes.head(1)
```

Out[5]:

	ROW_ID	SUBJECT_ID	HADM_ID	CHARTDATE	CHARTTIME	STORETIME	CA
0	174	22532	167853.0	2151-08-04	NaN	NaN	D

In [6]:

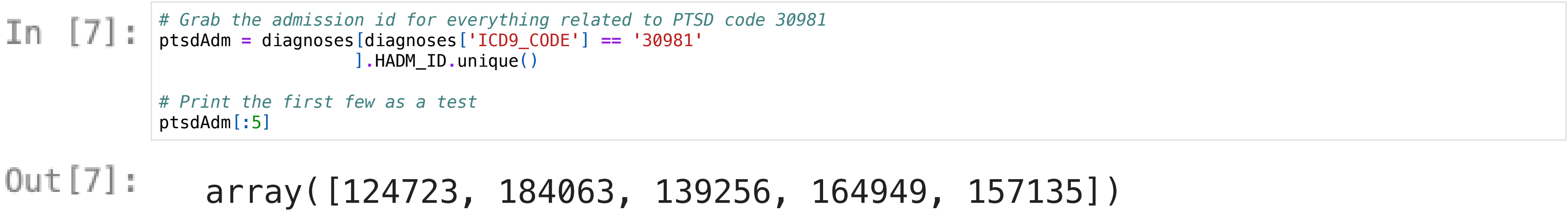
```python
# Load diagnoses
diagnoses = pd.read_csv('mimic-iii-clinical-database-1.4/DIAGNOSES_ICD.csv.gz',
                        compression='gzip')
diagnoses.head(1)
```

Out[6]:

	ROW_ID	SUBJECT_ID	HADM_ID	SEQ_NUM	ICD9_CODE
0	1297	109	172335	1.0	40301

Find all the admissions related to PTSD

Find all the admissions related to PTSD

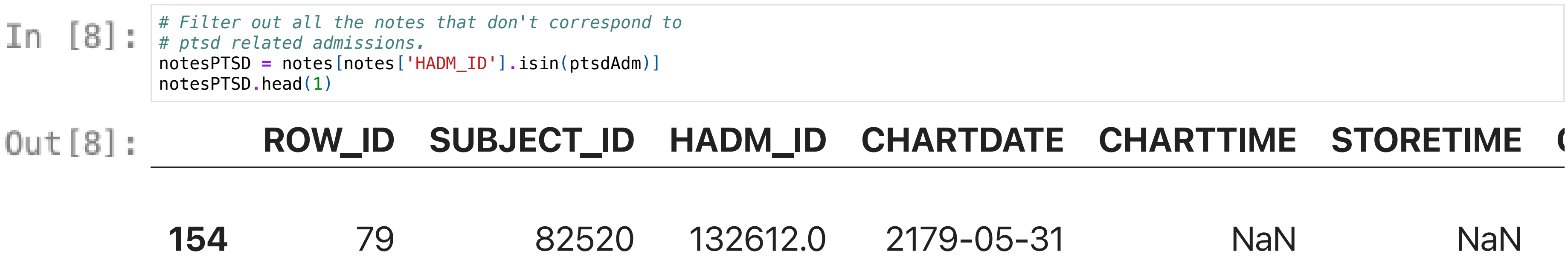


```
In [7]:
# Grab the admission id for everything related to PTSD code 30981
ptsdAdm = diagnoses[diagnoses['ICD9_CODE'] == '30981'
                   ].HADM_ID.unique()

# Print the first few as a test
ptsdAdm[:5]
```

```
Out[7]: array([124723, 184063, 139256, 164949, 157135])
```

Find all the notes related to said admissions

Find all the notes related to said admissions



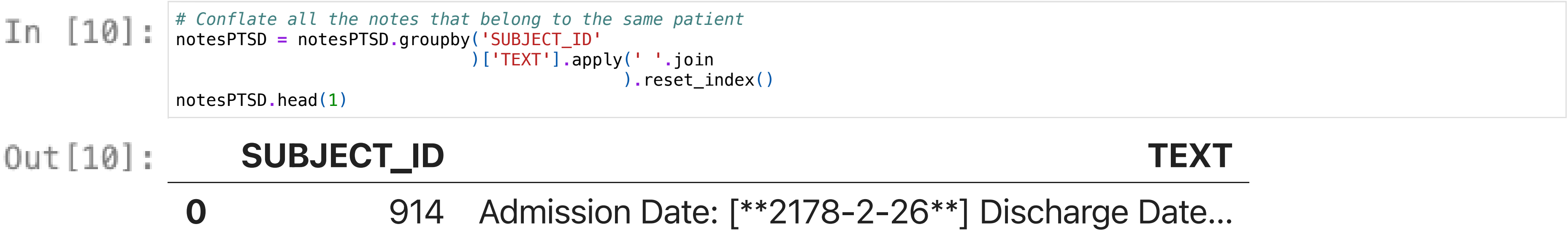
```
In [8]: # Filter out all the notes that don't correspond to
        # ptsd related admissions.
        notesPTSD = notes[notes['HADM_ID'].isin(ptsdAdm)]
        notesPTSD.head(1)
```

Out[8]:

	ROW_ID	SUBJECT_ID	HADM_ID	CHARTDATE	CHARTTIME	STORETIME
154	79	82520	132612.0	2179-05-31	NaN	NaN

Clean Up the notes and group them by patient

Clean Up the notes and group them by patient



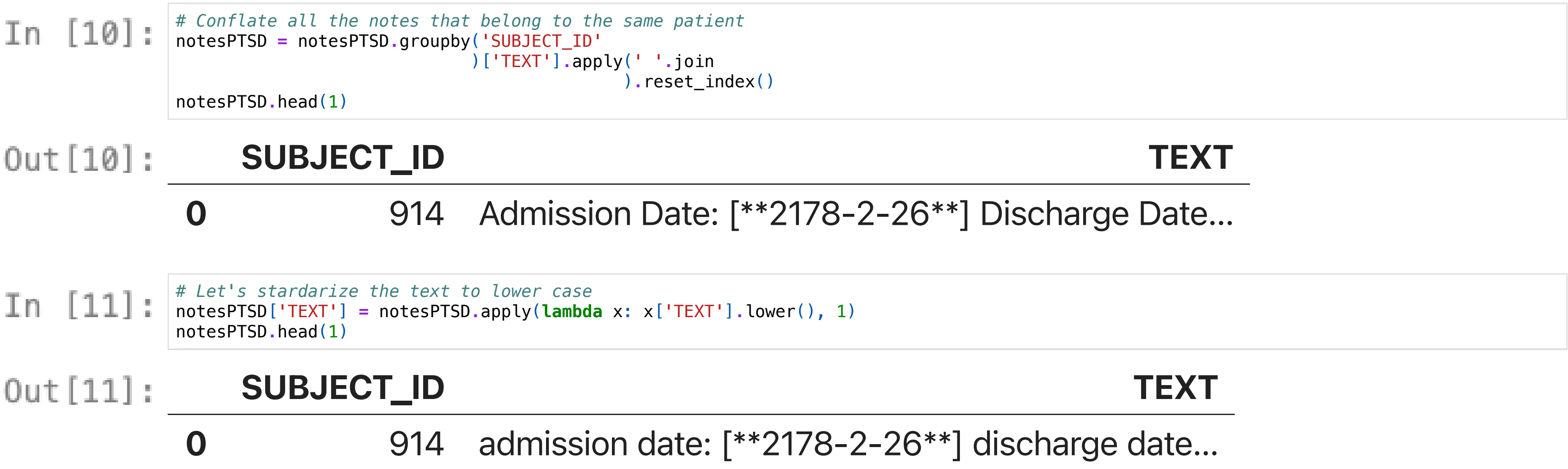
In [10]:

```
# Conflate all the notes that belong to the same patient
notesPTSD = notesPTSD.groupby('SUBJECT_ID'
                             )['TEXT'].apply(' '.join
                                            ).reset_index()
notesPTSD.head(1)
```

Out[10]:

	SUBJECT_ID	TEXT
0	914	Admission Date: [**2178-2-26**] Discharge Date...

Clean Up the notes and group them by patient



In [10]:

```
# Conflate all the notes that belong to the same patient
notesPTSD = notesPTSD.groupby('SUBJECT_ID'
                             )['TEXT'].apply(' '.join
                                            ).reset_index()
notesPTSD.head(1)
```

Out[10]:

	SUBJECT_ID	TEXT
0	914	Admission Date: [**2178-2-26**] Discharge Date...

In [11]:

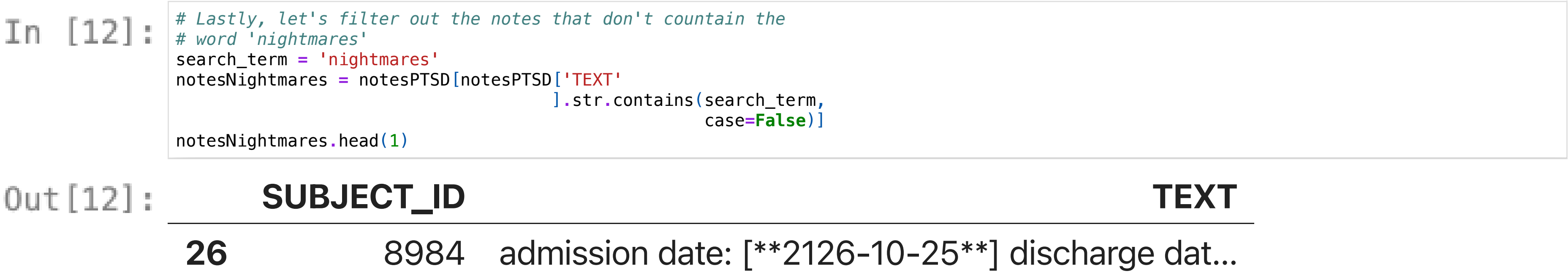
```
# Let's stardarize the text to lower case
notesPTSD['TEXT'] = notesPTSD.apply(lambda x: x['TEXT'].lower(), 1)
notesPTSD.head(1)
```

Out[11]:

	SUBJECT_ID	TEXT
0	914	admission date: [**2178-2-26**] discharge date...

Filter out everything that's does not mention '**nightmares**'

Filter out everything that’s does not mention ’**nightmares**’



In [12]:

```
# Lastly, let's filter out the notes that don't countain the
# word 'nightmares'
search_term = 'nightmares'
notesNightmares = notesPTSD[notesPTSD['TEXT'
                                     ].str.contains(search_term,
                                                    case=False)]
notesNightmares.head(1)
```

Out[12]:

	SUBJECT_ID	TEXT
26	8984	admission date: [**2126-10-25**] discharge dat...

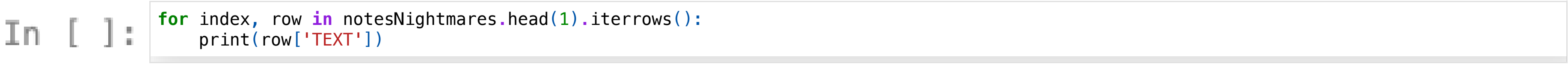
Let’s see what the data looks like

Let’s see what the data looks like

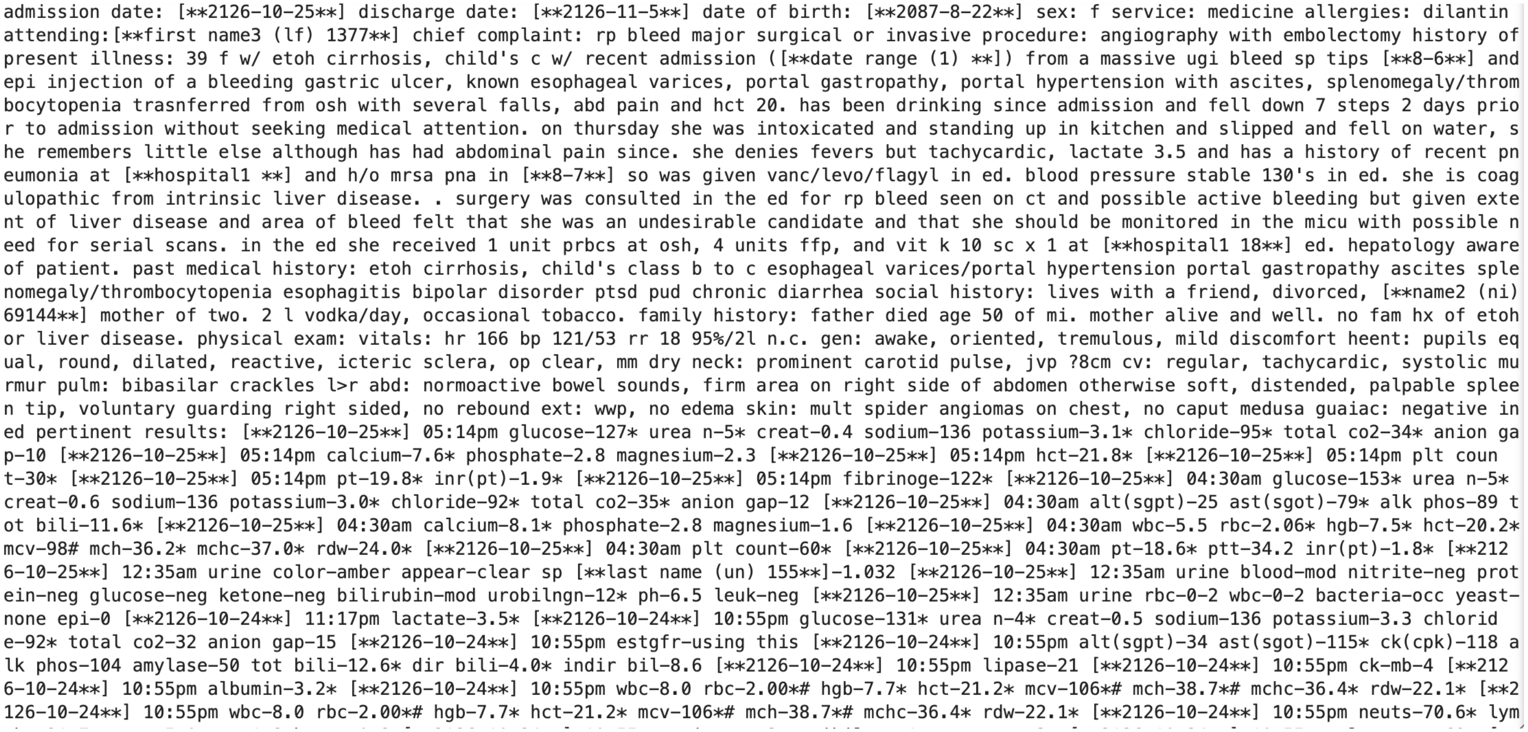
In [ ]:

```
for index, row in notesNightmares.head(1).iterrows():
    print(row['TEXT'])
```

Let’s see what the data looks like



```
In [ ]: for index, row in notesNightmares.head(1).iterrows():
            print(row['TEXT'])
```



```
admission date: [**2126-10-25**] discharge date: [**2126-11-5**] date of birth: [**2087-8-22**] sex: f service: medicine allergies: dilantin
attending:[**first name3 (lf) 1377**] chief complaint: rp bleed major surgical or invasive procedure: angiography with embolectomy history of
present illness: 39 f w/ etoh cirrhosis, child's c w/ recent admission ([**date range (1) **]) from a massive ugi bleed sp tips [**8-6**] and
epi injection of a bleeding gastric ulcer, known esophageal varices, portal gastropathy, portal hypertension with ascites, splenomegaly/throm
bocytopenia trasnferred from osh with several falls, abd pain and hct 20. has been drinking since admission and fell down 7 steps 2 days prio
r to admission without seeking medical attention. on thursday she was intoxicated and standing up in kitchen and slipped and fell on water, s
he remembers little else although has had abdominal pain since. she denies fevers but tachycardic, lactate 3.5 and has a history of recent pn
eumonia at [**hospital1 **] and h/o mrsa pna in [**8-7**] so was given vanc/levo/flagyl in ed. blood pressure stable 130's in ed. she is coag
ulopathic from intrinsic liver disease. . surgery was consulted in the ed for rp bleed seen on ct and possible active bleeding but given exte
nt of liver disease and area of bleed felt that she was an undesirable candidate and that she should be monitored in the micu with possible n
eed for serial scans. in the ed she received 1 unit prbcs at osh, 4 units ffp, and vit k 10 sc x 1 at [**hospital1 18**] ed. hepatology aware
of patient. past medical history: etoh cirrhosis, child's class b to c esophageal varices/portal hypertension portal gastropathy ascites sple
nomegaly/thrombocytopenia esophagitis bipolar disorder ptsd pud chronic diarrhea social history: lives with a friend, divorced, [**name2 (ni)
69144**] mother of two. 2 l vodka/day, occasional tobacco. family history: father died age 50 of mi. mother alive and well. no fam hx of etoh
or liver disease. physical exam: vitals: hr 166 bp 121/53 rr 18 95%/2l n.c. gen: awake, oriented, tremulous, mild discomfort heent: pupils eq
ual, round, dilated, reactive, icteric sclera, op clear, mm dry neck: prominent carotid pulse, jvp ?8cm cv: regular, tachycardic, systolic mu
rmur pulm: bibasilar crackles l>r abd: normoactive bowel sounds, firm area on right side of abdomen otherwise soft, distended, palpable splee
n tip, voluntary guarding right sided, no rebound ext: wwp, no edema skin: mult spider angiomas on chest, no caput medusa guaiac: negative in
ed pertinent results: [**2126-10-25**] 05:14pm glucose-127* urea n-5* creat-0.4 sodium-136 potassium-3.1* chloride-95* total co2-34* anion ga
p-10 [**2126-10-25**] 05:14pm calcium-7.6* phosphate-2.8 magnesium-2.3 [**2126-10-25**] 05:14pm hct-21.8* [**2126-10-25**] 05:14pm plt coun
t-30* [**2126-10-25**] 05:14pm pt-19.8* inr(pt)-1.9* [**2126-10-25**] 05:14pm fibrinoge-122* [**2126-10-25**] 04:30am glucose-153* urea n-5*
creat-0.6 sodium-136 potassium-3.0* chloride-92* total co2-35* anion gap-12 [**2126-10-25**] 04:30am alt(sgpt)-25 ast(sgot)-79* alk phos-89 t
ot bili-11.6* [**2126-10-25**] 04:30am calcium-8.1* phosphate-2.8 magnesium-1.6 [**2126-10-25**] 04:30am wbc-5.5 rbc-2.06* hgb-7.5* hct-20.2*
mcv-98# mch-36.2* mchc-37.0* rdw-24.0* [**2126-10-25**] 04:30am plt count-60* [**2126-10-25**] 04:30am pt-18.6* ptt-34.2 inr(pt)-1.8* [**212
6-10-25**] 12:35am urine color-amber appear-clear sp [**last name (un) 155**]-1.032 [**2126-10-25**] 12:35am urine blood-mod nitrite-neg prot
ein-neg glucose-neg ketone-neg bilirubin-mod urobilngn-12* ph-6.5 leuk-neg [**2126-10-25**] 12:35am urine rbc-0-2 wbc-0-2 bacteria-occ yeast-
none epi-0 [**2126-10-24**] 11:17pm lactate-3.5* [**2126-10-24**] 10:55pm glucose-131* urea n-4* creat-0.5 sodium-136 potassium-3.3 chlorid
e-92* total co2-32 anion gap-15 [**2126-10-24**] 10:55pm estgfr-using this [**2126-10-24**] 10:55pm alt(sgpt)-34 ast(sgot)-115* ck(cpk)-118 a
lk phos-104 amylase-50 tot bili-12.6* dir bili-4.0* indir bil-8.6 [**2126-10-24**] 10:55pm lipase-21 [**2126-10-24**] 10:55pm ck-mb-4 [**212
6-10-24**] 10:55pm albumin-3.2* [**2126-10-24**] 10:55pm wbc-8.0 rbc-2.00*# hgb-7.7* hct-21.2* mcv-106*# mch-38.7*# mchc-36.4* rdw-22.1* [**2
126-10-24**] 10:55pm wbc-8.0 rbc-2.00*# hgb-7.7* hct-21.2* mcv-106*# mch-38.7*# mchc-36.4* rdw-22.1* [**2126-10-24**] 10:55pm neuts-70.6* lym
```

As you can see even for one entry there’s too much information in there still.

As you can see even for one entry there's too much information in there still.

We can get better insights by leveraging Spacy and Displacy

As you can see even for one entry there’s too much information in there still.

We can get better insights by leveraging Spacy and Displacy

In [ ]:

```
# In order to do this we need the following dependencies
import spacy

import en_ner_bc5cdr_md
nlp = en_ner_bc5cdr_md.load()

# Entity Visualizer
from spacy import displacy
```

In [ ]:

```
# highlight the medical terms
for index, row in notesNightmares.head(1).iterrows():
    displacy.render(nlp(row['TEXT']), style="ent", jupyter=True)
```

In [ ]:

```
# highlight the medical terms
for index, row in notesNightmares.head(1).iterrows():
    displacy.render(nlp(row['TEXT']), style="ent", jupyter=True)
```

```
admission date: [**2126-10-25**] discharge date: [**2126-11-5**] date of birth: [ DISEASE **2087-8-22**] sex: f service: medicine allergies DISEASE : dilantin
CHEMICAL attending:[**first name3 (lf) 1377**] chief complaint: rp bleed major surgical or invasive procedure: angiography with embolectomy history of present illness:
39 f w/ etoh cirrhosis DISEASE , child's c w/ recent admission ([**date range (1) **]) from a massive ugi bleed sp tips [**8-6**] and epi injection of a bleeding
DISEASE gastric ulcer DISEASE , known esophageal varices DISEASE , portal gastropathy DISEASE , portal hypertension DISEASE with ascites
DISEASE , splenomegaly/thrombocytopenia DISEASE trasnferred from osh with several falls, abd pain DISEASE and hct 20. has been drinking since admission
and fell down 7 steps 2 days prior to admission without seeking medical attention. on thursday she was intoxicated and standing up in kitchen and slipped and fell on
water, she remembers little else although has had abdominal pain DISEASE since. she denies fevers but tachycardic, lactate CHEMICAL 3.5 and has a history of
recent pneumonia DISEASE at [**hospital1 **] and h/o mrsa pna in [**8-7**] so was given vanc/levo/flagyl CHEMICAL in ed. blood pressure stable 130's in ed. she
is coagulopathic DISEASE from intrinsic liver disease DISEASE . . surgery was consulted in the ed for rp bleed seen on ct and possible active bleeding
DISEASE but given extent of liver disease DISEASE and area of bleed felt that she was an undesirable candidate and that she should be monitored in the micu with
possible need for serial scans. in the ed she received 1 unit prbcs at osh, 4 units ffp, and vit CHEMICAL k 10 sc x 1 at [**hospital1 18**] ed. hepatology aware of
patient. past medical history: etoh cirrhosis DISEASE , child's class b to c esophageal varices/portal hypertension DISEASE portal gastropathy ascites
splenomegaly/thrombocytopenia esophagitis DISEASE bipolar disorder DISEASE ptsd pud chronic diarrhea DISEASE social history: lives with a friend, divorced,
```

This is good but we are interested in information related to nightmares and that is difficult to pick out so far.

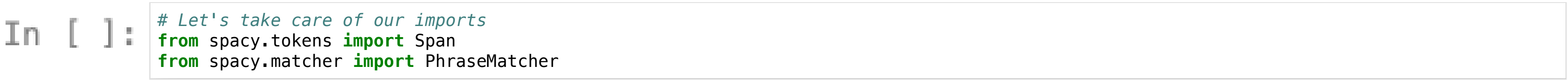
Now that we have preprocessed our data let's move on to the fun parts.

Creating your own labels with Spacy

We will be creating our own label for '**nightmares**'. (Note: This can be used with any search term)

In [ ]:

```
# Let's take care of our imports
from spacy.tokens import Span
from spacy.matcher import PhraseMatcher
```



```
In [ ]: # Let's take care of our imports
        from spacy.tokens import Span
        from spacy.matcher import PhraseMatcher
```

Create the search term or pattern to identify

In [ ]:

```
# Let's take care of our imports
from spacy.tokens import Span
from spacy.matcher import PhraseMatcher
```

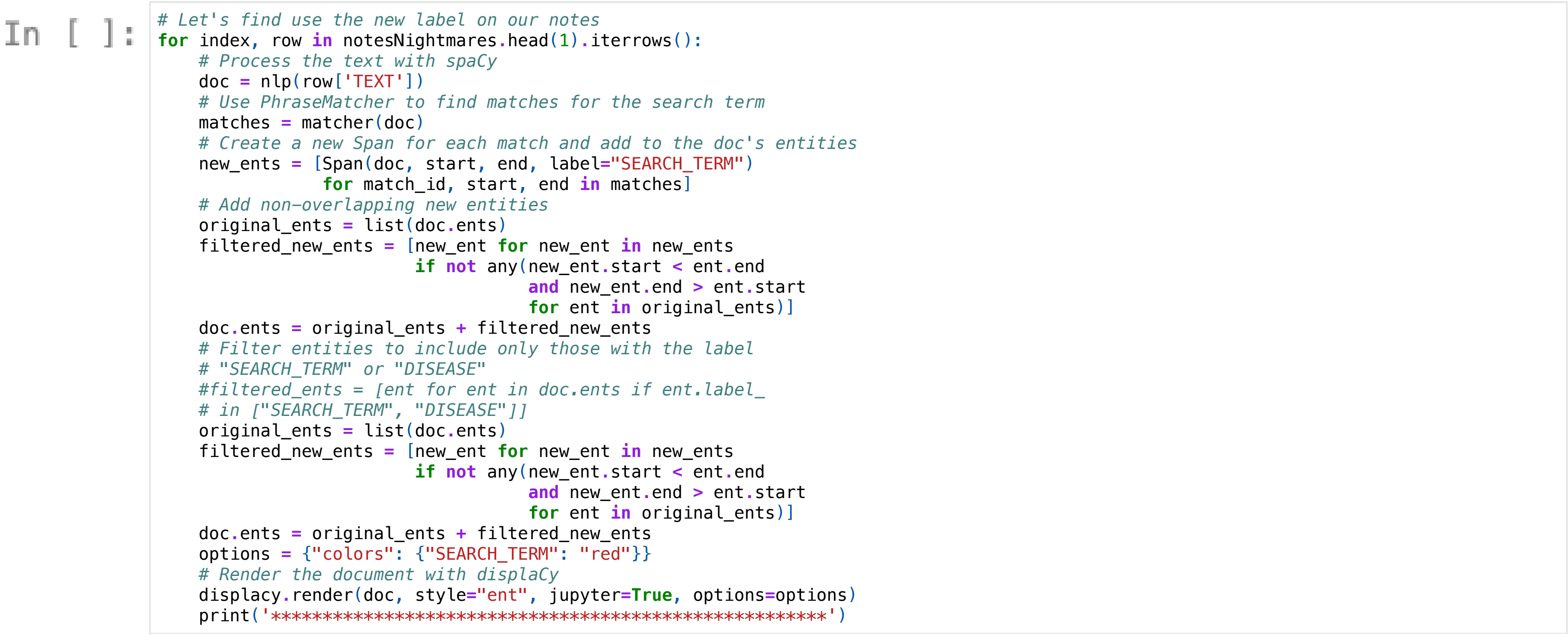
Create the search term or pattern to identify

In [ ]:

```
search_term = 'nightmares'
matcher = PhraseMatcher(nlp.vocab, attr="LOWER")
pattern = [nlp.make_doc(search_term)]
matcher.add("SEARCH_TERM", pattern)
```

Let’s apply this pattern to our data

Let’s apply this pattern to our data

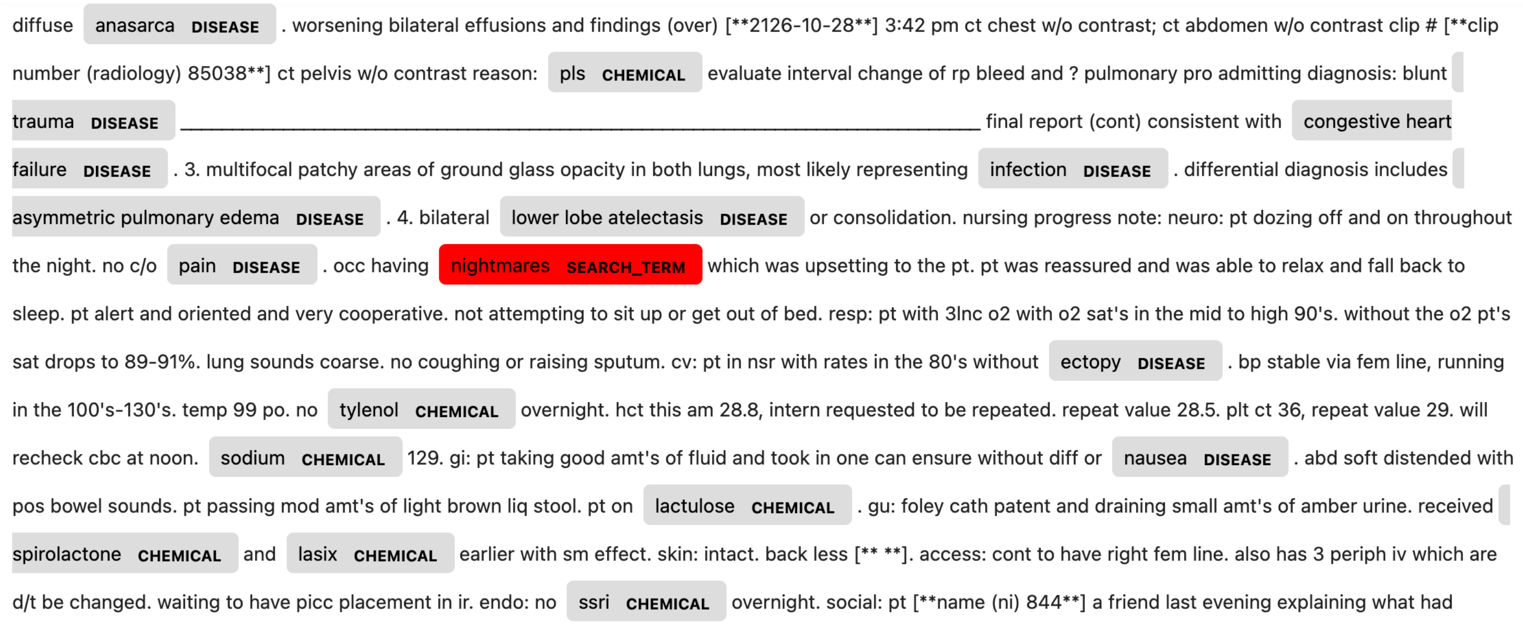


In [ ]:

```
# Let's find use the new label on our notes
for index, row in notesNightmares.head(1).iterrows():
    # Process the text with spaCy
    doc = nlp(row['TEXT'])
    # Use PhraseMatcher to find matches for the search term
    matches = matcher(doc)
    # Create a new Span for each match and add to the doc's entities
    new_ents = [Span(doc, start, end, label="SEARCH_TERM")
                for match_id, start, end in matches]
    # Add non-overlapping new entities
    original_ents = list(doc.ents)
    filtered_new_ents = [new_ent for new_ent in new_ents
                         if not any(new_ent.start < ent.end
                                    and new_ent.end > ent.start
                                    for ent in original_ents)]
    doc.ents = original_ents + filtered_new_ents
    # Filter entities to include only those with the label
    # "SEARCH_TERM" or "DISEASE"
    #filtered_ents = [ent for ent in doc.ents if ent.label_
    # in ["SEARCH_TERM", "DISEASE"]]
    original_ents = list(doc.ents)
    filtered_new_ents = [new_ent for new_ent in new_ents
                         if not any(new_ent.start < ent.end
                                    and new_ent.end > ent.start
                                    for ent in original_ents)]
    doc.ents = original_ents + filtered_new_ents
    options = {"colors": {"SEARCH_TERM": "red"}}
    # Render the document with displaCy
    displacy.render(doc, style="ent", jupyter=True, options=options)
    print('******************************************************')
```

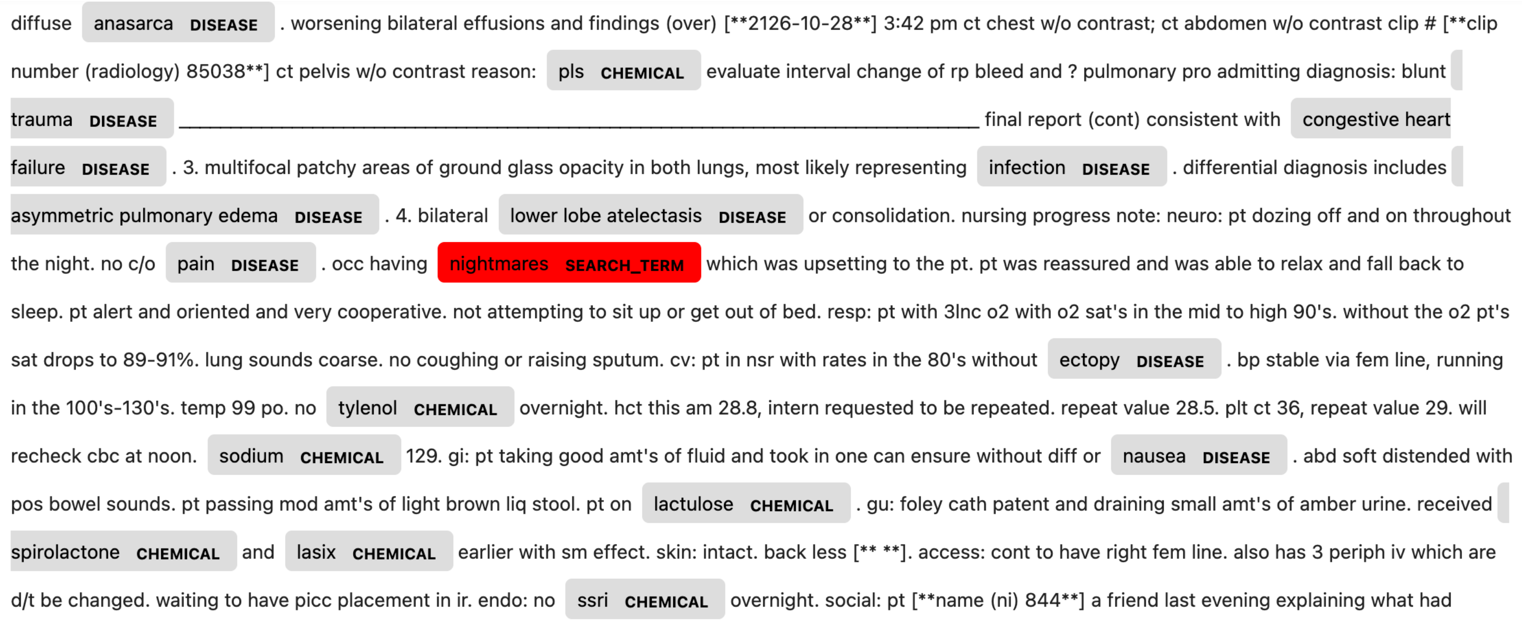
Results

Results



diffuse anasarca DISEASE . worsening bilateral effusions and findings (over) [**2126-10-28**] 3:42 pm ct chest w/o contrast; ct abdomen w/o contrast clip # [**clip number (radiology) 85038**] ct pelvis w/o contrast reason: pls CHEMICAL evaluate interval change of rp bleed and ? pulmonary pro admitting diagnosis: blunt trauma DISEASE ____ final report (cont) consistent with congestive heart failure DISEASE . 3. multifocal patchy areas of ground glass opacity in both lungs, most likely representing infection DISEASE . differential diagnosis includes asymmetric pulmonary edema DISEASE . 4. bilateral lower lobe atelectasis DISEASE or consolidation. nursing progress note: neuro: pt dozing off and on throughout the night. no c/o pain DISEASE . occ having nightmares SEARCH_TERM which was upsetting to the pt. pt was reassured and was able to relax and fall back to sleep. pt alert and oriented and very cooperative. not attempting to sit up or get out of bed. resp: pt with 3lnc o2 with o2 sat's in the mid to high 90's. without the o2 pt's sat drops to 89-91%. lung sounds coarse. no coughing or raising sputum. cv: pt in nsr with rates in the 80's without ectopy DISEASE . bp stable via fem line, running in the 100's-130's. temp 99 po. no tylenol CHEMICAL overnight. hct this am 28.8, intern requested to be repeated. repeat value 28.5. plt ct 36, repeat value 29. will recheck cbc at noon. sodium CHEMICAL 129. gi: pt taking good amt's of fluid and took in one can ensure without diff or nausea DISEASE . abd soft distended with pos bowel sounds. pt passing mod amt's of light brown liq stool. pt on lactulose CHEMICAL . gu: foley cath patent and draining small amt's of amber urine. received spirolactone CHEMICAL and lasix CHEMICAL earlier with sm effect. skin: intact. back less [** **]. access: cont to have right fem line. also has 3 periph iv which are d/t be changed. waiting to have picc placement in ir. endo: no ssri CHEMICAL overnight. social: pt [**name (ni) 844**] a friend last evening explaining what had

Results



diffuse anasarca DISEASE . worsening bilateral effusions and findings (over) [**2126-10-28**] 3:42 pm ct chest w/o contrast; ct abdomen w/o contrast clip # [**clip number (radiology) 85038**] ct pelvis w/o contrast reason: pls CHEMICAL evaluate interval change of rp bleed and ? pulmonary pro admitting diagnosis: blunt trauma DISEASE ____ final report (cont) consistent with congestive heart failure DISEASE . 3. multifocal patchy areas of ground glass opacity in both lungs, most likely representing infection DISEASE . differential diagnosis includes asymmetric pulmonary edema DISEASE . 4. bilateral lower lobe atelectasis DISEASE or consolidation. nursing progress note: neuro: pt dozing off and on throughout the night. no c/o pain DISEASE . occ having nightmares SEARCH_TERM which was upsetting to the pt. pt was reassured and was able to relax and fall back to sleep. pt alert and oriented and very cooperative. not attempting to sit up or get out of bed. resp: pt with 3lnc o2 with o2 sat's in the mid to high 90's. without the o2 pt's sat drops to 89-91%. lung sounds coarse. no coughing or raising sputum. cv: pt in nsr with rates in the 80's without ectopy DISEASE . bp stable via fem line, running in the 100's-130's. temp 99 po. no tylenol CHEMICAL overnight. hct this am 28.8, intern requested to be repeated. repeat value 28.5. plt ct 36, repeat value 29. will recheck cbc at noon. sodium CHEMICAL 129. gi: pt taking good amt's of fluid and took in one can ensure without diff or nausea DISEASE . abd soft distended with pos bowel sounds. pt passing mod amt's of light brown liq stool. pt on lactulose CHEMICAL . gu: foley cath patent and draining small amt's of amber urine. received spirolactone CHEMICAL and lasix CHEMICAL earlier with sm effect. skin: intact. back less [** **]. access: cont to have right fem line. also has 3 periph iv which are d/t be changed. waiting to have picc placement in ir. endo: no ssri CHEMICAL overnight. social: pt [**name (ni) 844**] a friend last evening explaining what had

Much easier to spot our keyword!

Summarizing (NLP)

Summarizing PTSD patient's notes

Summarizing PTSD patient’s notes

First let’s take care of our dependencies

Summarizing PTSD patient’s notes

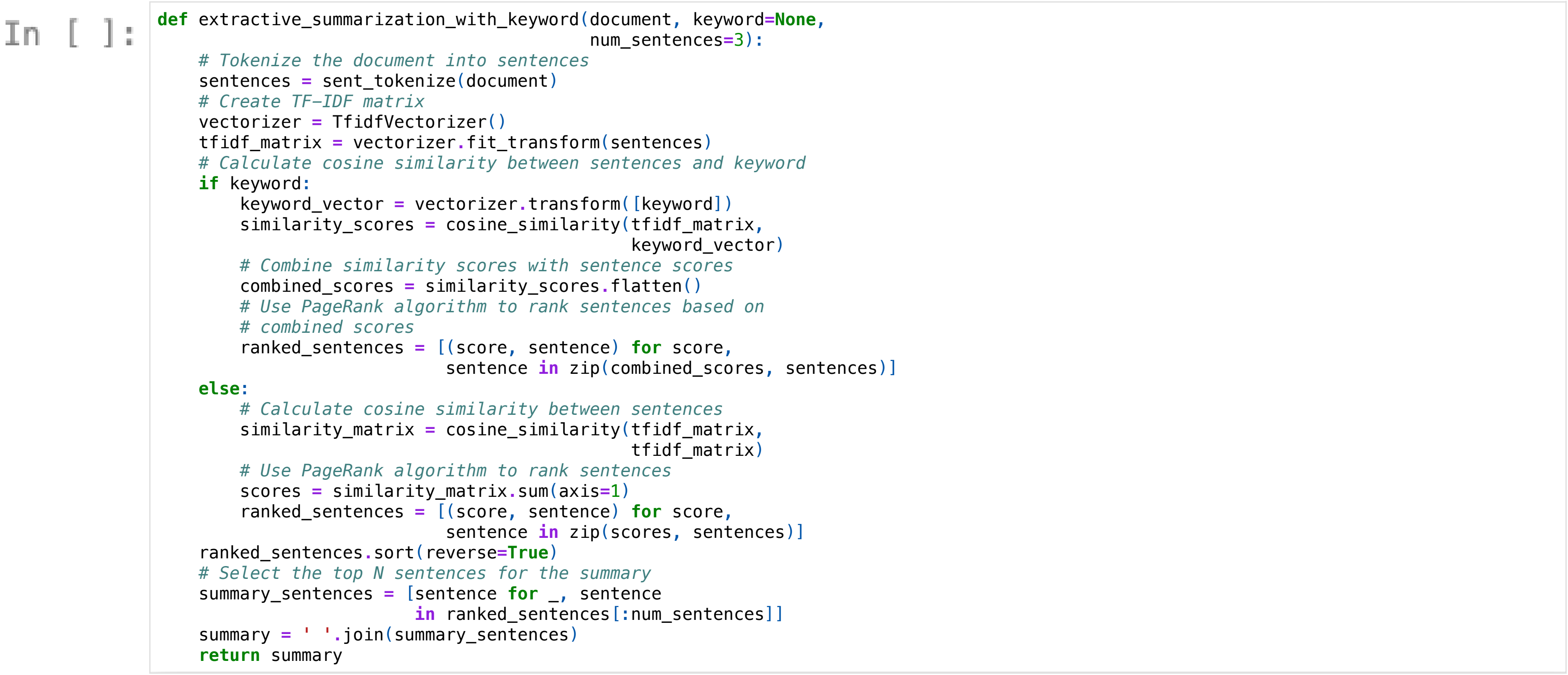
First let’s take care of our dependencies

In [ ]:

```
import re
import string
import nltk
from nltk.tokenize import sent_tokenize

#from IPython.display import HTML
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
```

Let’s create our summarization function

Let’s create our summarization function



In [ ]:

```python
def extractive_summarization_with_keyword(document, keyword=None,
                                          num_sentences=3):
    # Tokenize the document into sentences
    sentences = sent_tokenize(document)
    # Create TF-IDF matrix
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform(sentences)
    # Calculate cosine similarity between sentences and keyword
    if keyword:
        keyword_vector = vectorizer.transform([keyword])
        similarity_scores = cosine_similarity(tfidf_matrix,
                                              keyword_vector)
        # Combine similarity scores with sentence scores
        combined_scores = similarity_scores.flatten()
        # Use PageRank algorithm to rank sentences based on
        # combined scores
        ranked_sentences = [(score, sentence) for score,
                            sentence in zip(combined_scores, sentences)]
    else:
        # Calculate cosine similarity between sentences
        similarity_matrix = cosine_similarity(tfidf_matrix,
                                              tfidf_matrix)
        # Use PageRank algorithm to rank sentences
        scores = similarity_matrix.sum(axis=1)
        ranked_sentences = [(score, sentence) for score,
                            sentence in zip(scores, sentences)]
    ranked_sentences.sort(reverse=True)
    # Select the top N sentences for the summary
    summary_sentences = [sentence for _, sentence
                         in ranked_sentences[:num_sentences]]
    summary = ' '.join(summary_sentences)
    return summary
```

Summarization

Summarization

```
In [ ]:
```

```
# Perform extractive summarization with keyword search
for index, row in notesPTSD.head(3).iterrows():
    summary = extractive_summarization_with_keyword(row['TEXT'])
    # Process the text with spaCy
    doc = nlp(summary)
    # Use PhraseMatcher to find matches for the search term
    matches = matcher(doc)
    # Create a new Span for each match and add to the doc's entities
    new_ents = [Span(doc, start, end, label="SEARCH_TERM")
                for match_id, start, end in matches]
    # Add non-overlapping new entities
    original_ents = list(doc.ents)
    filtered_new_ents = [new_ent for new_ent in new_ents
                         if not any(new_ent.start < ent.end
                                    and new_ent.end > ent.start
                                    for ent in original_ents)]
    doc.ents = original_ents + filtered_new_ents
    # Filter entities to include only those with the label
    # "SEARCH_TERM" or "DISEASE"
    #filtered_ents = [ent for ent in doc.ents if ent.label_
    # in ["SEARCH_TERM", "DISEASE"]]
    #doc.ents = filtered_ents
    original_ents = list(doc.ents)
    filtered_new_ents = [new_ent for new_ent in new_ents
                         if not any(new_ent.start < ent.end
                                    and new_ent.end > ent.start
                                    for ent in original_ents)]
    doc.ents = original_ents + filtered_new_ents
    options = {"colors": {"SEARCH_TERM": "red"}}
    # Render the document with displaCy
    displacy.render(doc, style="ent", jupyter=True, options=options)
    print('************************************************')
```

Results

Results

note is made of opacity **DISEASE** in the right middle and lower lung field, extending from right hilar lesion, which is slightly increased compared to the prior study, probably representing enlarged pulmonary artery in this patient with pe and consolidation noted in the cta study performed on the same day. ct of the chest without and with iv contrast: there has been no significant interval change in the appearance of extensive clot burden extending from the right and left main pulmonary arteries and into more distal segmental and subsegmental branches. imaged portions of the upper abdomen again demonstrate 2 fluid collections within the liver, the larger of the 2 measuring 5.8 x 3.9 cm and located within the right lobe of the liver.

****

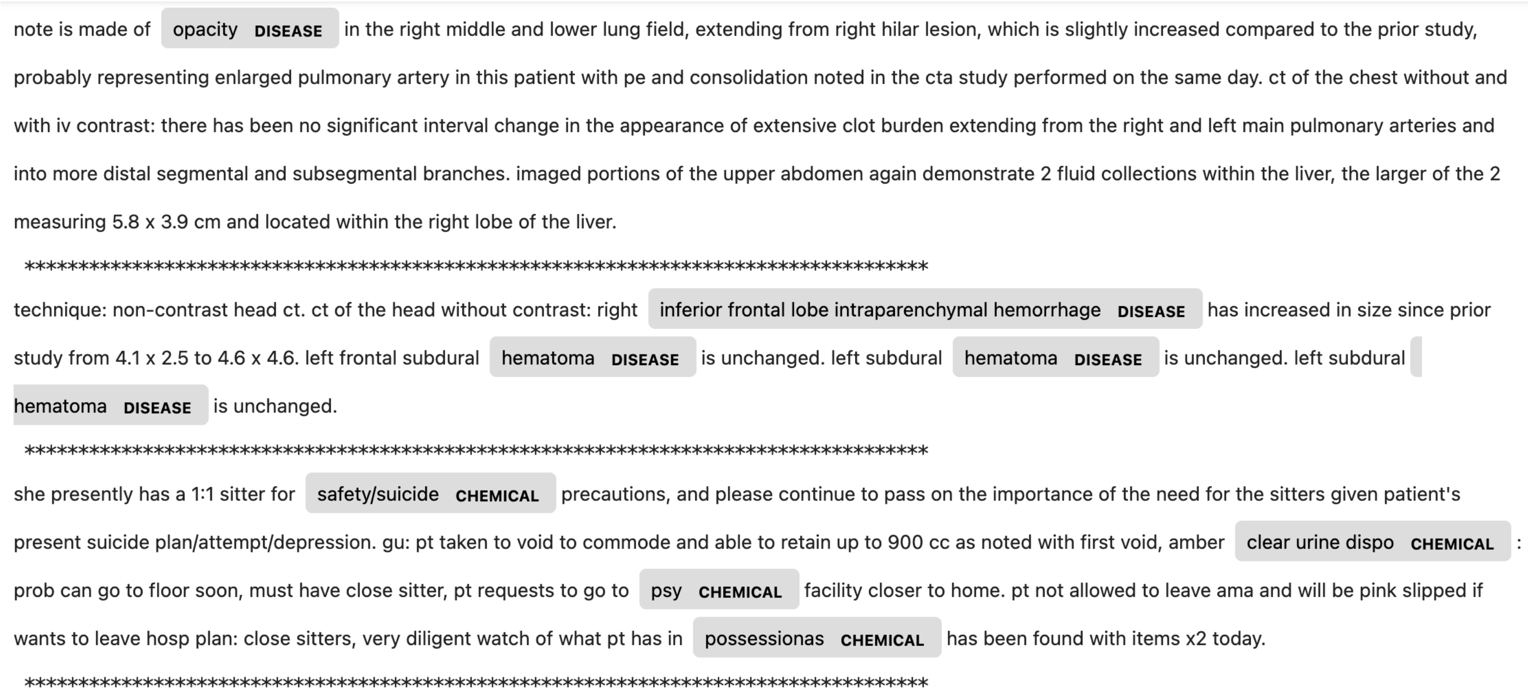
technique: non-contrast head ct. ct of the head without contrast: right inferior frontal lobe intraparenchymal hemorrhage **DISEASE** has increased in size since prior study from 4.1 x 2.5 to 4.6 x 4.6. left frontal subdural hematoma **DISEASE** is unchanged. left subdural hematoma **DISEASE** is unchanged. left subdural hematoma **DISEASE** is unchanged.

****

she presently has a 1:1 sitter for safety/suicide **CHEMICAL** precautions, and please continue to pass on the importance of the need for the sitters given patient's present suicide plan/attempt/depression. gu: pt taken to void to commode and able to retain up to 900 cc as noted with first void, amber clear urine dispo **CHEMICAL** : prob can go to floor soon, must have close sitter, pt requests to go to psy **CHEMICAL** facility closer to home. pt not allowed to leave ama and will be pink slipped if wants to leave hosp plan: close sitters, very diligent watch of what pt has in possessionas **CHEMICAL** has been found with items x2 today.

****

Results



note is made of opacity DISEASE in the right middle and lower lung field, extending from right hilar lesion, which is slightly increased compared to the prior study, probably representing enlarged pulmonary artery in this patient with pe and consolidation noted in the cta study performed on the same day. ct of the chest without and with iv contrast: there has been no significant interval change in the appearance of extensive clot burden extending from the right and left main pulmonary arteries and into more distal segmental and subsegmental branches. imaged portions of the upper abdomen again demonstrate 2 fluid collections within the liver, the larger of the 2 measuring 5.8 x 3.9 cm and located within the right lobe of the liver.

****

technique: non-contrast head ct. ct of the head without contrast: right inferior frontal lobe intraparenchymal hemorrhage DISEASE has increased in size since prior study from 4.1 x 2.5 to 4.6 x 4.6. left frontal subdural hematoma DISEASE is unchanged. left subdural hematoma DISEASE is unchanged. left subdural hematoma DISEASE is unchanged.

****

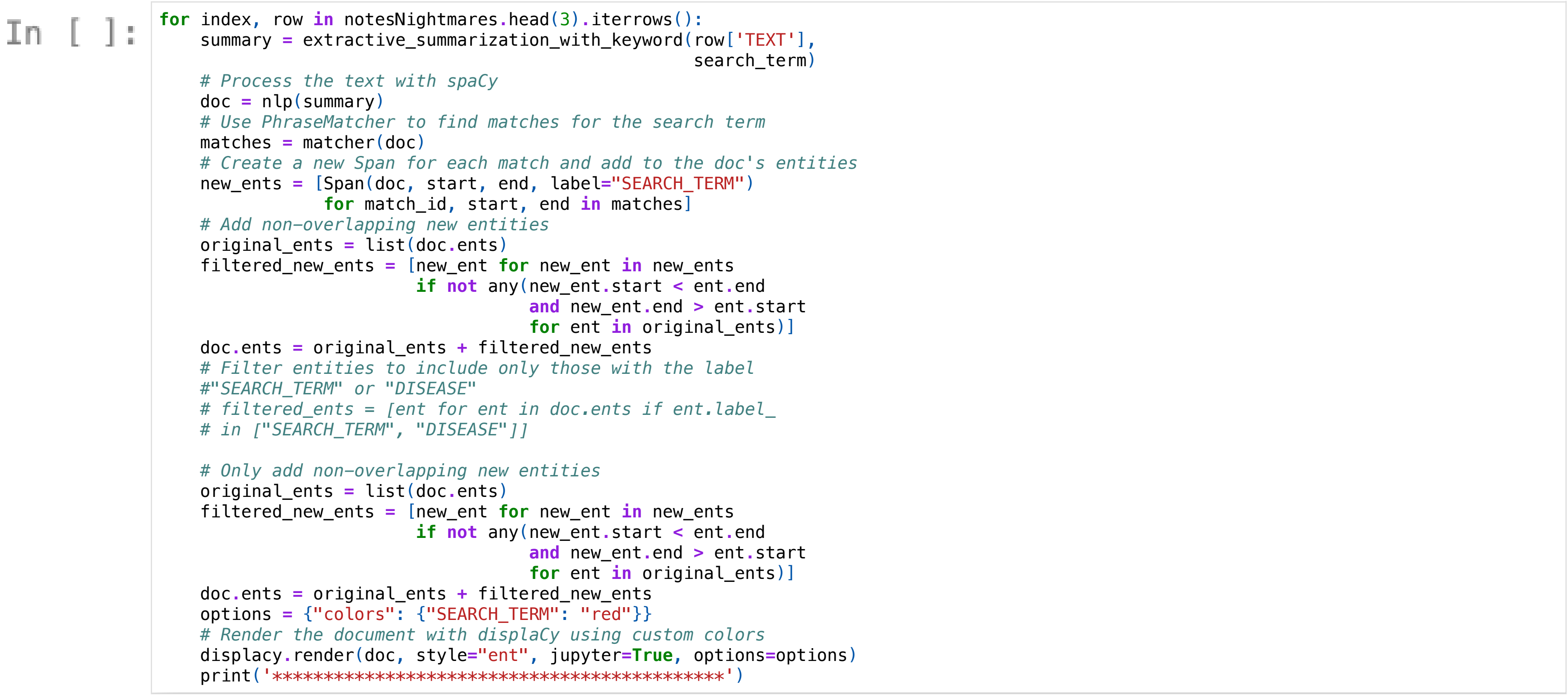
she presently has a 1:1 sitter for safety/suicide CHEMICAL precautions, and please continue to pass on the importance of the need for the sitters given patient's present suicide plan/attempt/depression. gu: pt taken to void to commode and able to retain up to 900 cc as noted with first void, amber clear urine dispo CHEMICAL : prob can go to floor soon, must have close sitter, pt requests to go to psy CHEMICAL facility closer to home. pt not allowed to leave ama and will be pink slipped if wants to leave hosp plan: close sitters, very diligent watch of what pt has in possessionas CHEMICAL has been found with items x2 today.

****

We've summarized the notes of three patients onto 3 sentences for each

Summarizing relevant information based on a Keyword (In our case: ’**nightmares**’)

Summarizing relevant information based on a Keyword (In our case: '**nightmares**')



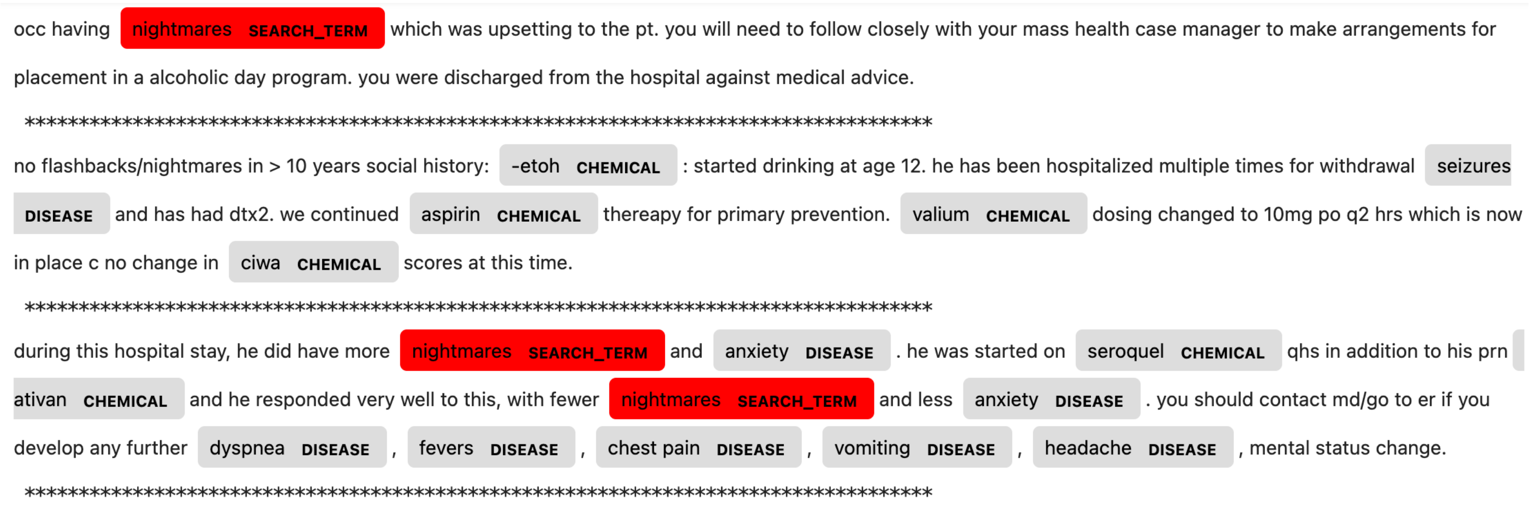
```
In [ ]: for index, row in notesNightmares.head(3).iterrows():
            summary = extractive_summarization_with_keyword(row['TEXT'],
                                                            search_term)
            # Process the text with spaCy
            doc = nlp(summary)
            # Use PhraseMatcher to find matches for the search term
            matches = matcher(doc)
            # Create a new Span for each match and add to the doc's entities
            new_ents = [Span(doc, start, end, label="SEARCH_TERM")
                        for match_id, start, end in matches]
            # Add non-overlapping new entities
            original_ents = list(doc.ents)
            filtered_new_ents = [new_ent for new_ent in new_ents
                                 if not any(new_ent.start < ent.end
                                            and new_ent.end > ent.start
                                            for ent in original_ents)]
            doc.ents = original_ents + filtered_new_ents
            # Filter entities to include only those with the label
            #"SEARCH_TERM" or "DISEASE"
            # filtered_ents = [ent for ent in doc.ents if ent.label_
            # in ["SEARCH_TERM", "DISEASE"]]

            # Only add non-overlapping new entities
            original_ents = list(doc.ents)
            filtered_new_ents = [new_ent for new_ent in new_ents
                                 if not any(new_ent.start < ent.end
                                            and new_ent.end > ent.start
                                            for ent in original_ents)]
            doc.ents = original_ents + filtered_new_ents
            options = {"colors": {"SEARCH_TERM": "red"}}
            # Render the document with displaCy using custom colors
            displacy.render(doc, style="ent", jupyter=True, options=options)
            print('*******************************************')
```

Results

We’ve extracted the relevant sentences with regards to ’nightmares and summarized those’

Results

We've extracted the relevant sentences with regards to 'nightmares and summarized those'



occc having nightmares SEARCH_TERM which was upsetting to the pt. you will need to follow closely with your mass health case manager to make arrangements for placement in a alcoholic day program. you were discharged from the hospital against medical advice.

****

no flashbacks/nightmares in > 10 years social history: -etoh CHEMICAL : started drinking at age 12. he has been hospitalized multiple times for withdrawal seizures DISEASE and has had dtx2. we continued aspirin CHEMICAL thereapy for primary prevention. valium CHEMICAL dosing changed to 10mg po q2 hrs which is now in place c no change in ciwa CHEMICAL scores at this time.

****

during this hospital stay, he did have more nightmares SEARCH_TERM and anxiety DISEASE . he was started on seroquel CHEMICAL qhs in addition to his prn ativan CHEMICAL and he responded very well to this, with fewer nightmares SEARCH_TERM and less anxiety DISEASE . you should contact md/go to er if you develop any further dyspnea DISEASE , fevers DISEASE , chest pain DISEASE , vomiting DISEASE , headache DISEASE , mental status change.

****

Thank you!