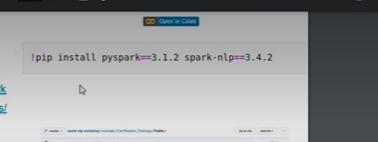
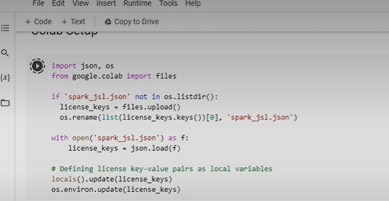
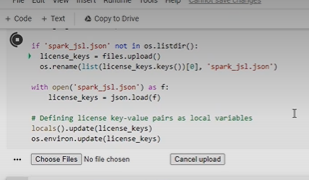
Spark nlp

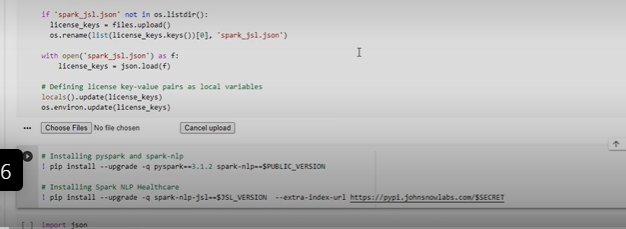




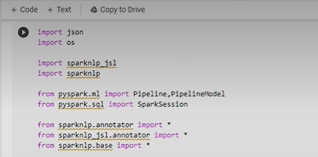




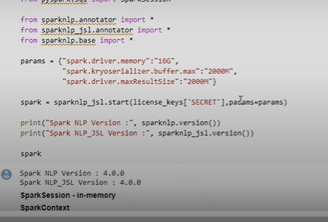
We need secret key



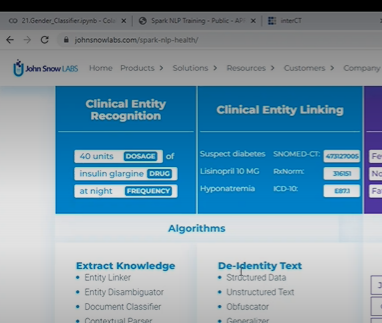
Install all dependencies for spark nlp



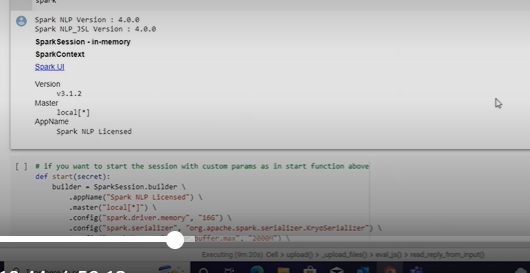
Now we need to start session

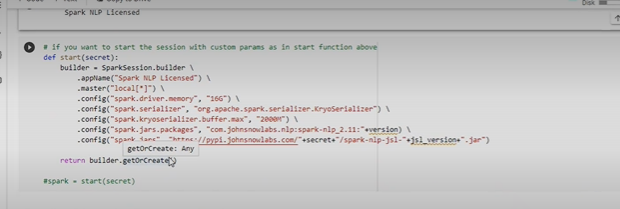


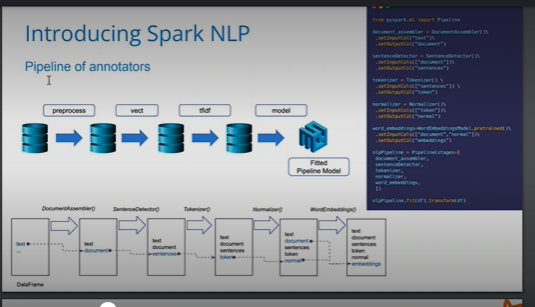
If do not give any parameter it wil take its own parameter



Johnsnowlabs has different pretrained models







Document assembler:

If we have 100 documentwe will assemble at one place: done by document assembler

1st step: document assembler

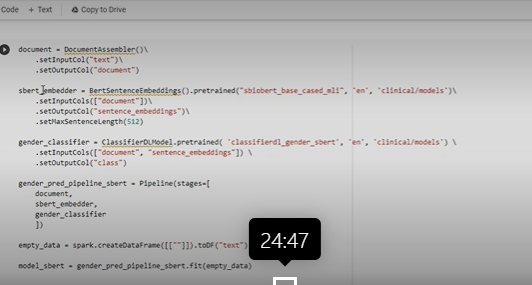
2. detect sentese from doc

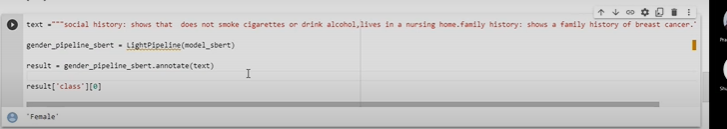
Tokenization

Normalizer( Remove special words, remove special characters )

Tokens into sentence

Sentence to word embeddings



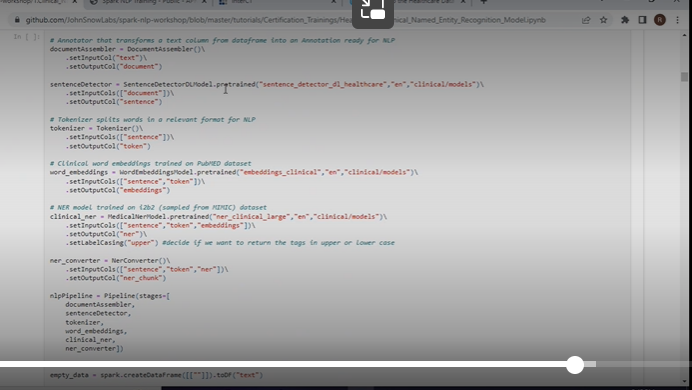


Transfer learning

There is existing model, when we want to apply their model on ur data set, we are not allowed to change architechure like number of neurons number of layers and activation function only few matters can be changed like batch size, epoch and all

Only we can fine tune the model

NER we follow medical NER



NER

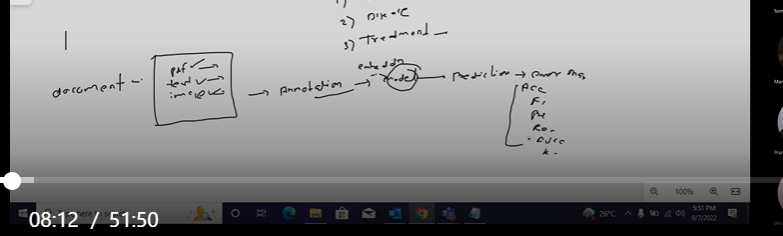
I sould know what kind of entities I sould recognize

For example in this

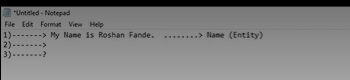
Test

Disease

Treatment

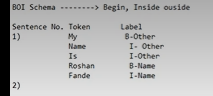


Annotation means label data



How I do annotation

Below- conll model



Above process is called annotation

Different tools available- ner annotator , prodigy

After annotation – embeddings



Other format of labeling

Using spacy



Most of the time we get annotated data set

We will be using their complete architecture and fine tune



If the data has pdf or image formate- we need to convert to text format

OCR will be used to convert image into text



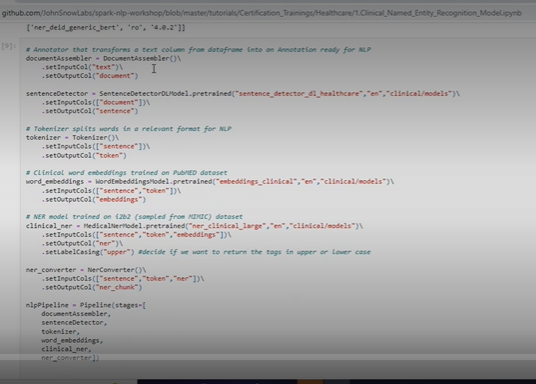
If I want to use image directly – we use layoutLM



For classification-

LAYOUTLM use internally OCR(it extract text- and also ets position of text(start and end index)

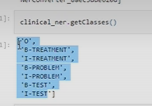
Visualizer to highlight entity



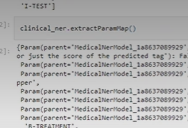
Text: should be in text column

Ner converter: filter which only extract ner chunk not all the entities

How many ner this model recognizes



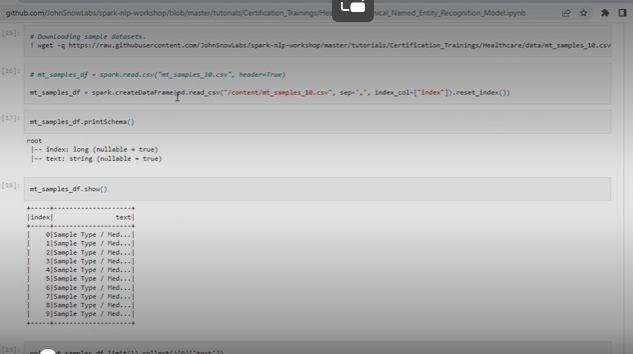
Parameters which they have used while training



Now test ner model

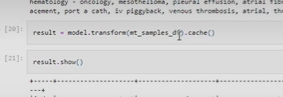
First download data set

Convert data set into spark dataset

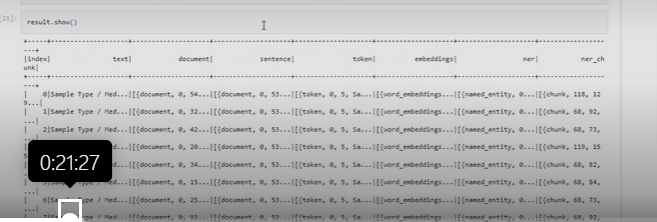


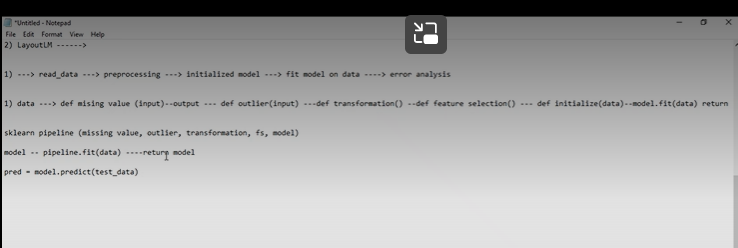
First we fitted model into empty dataset

Now for test we are transforming testing dataset



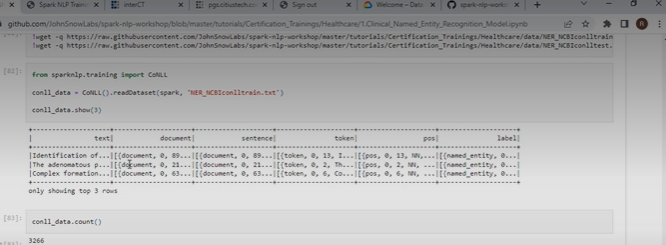
Result





In spark nlp taking data as data frame and giving result as data frame

Custom NER

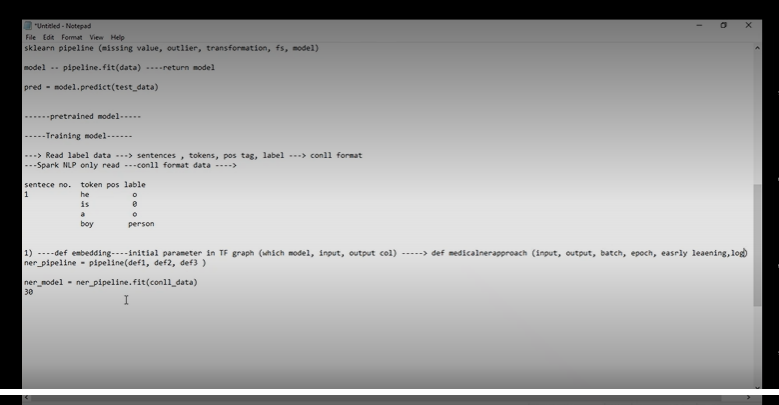


NERDL graph

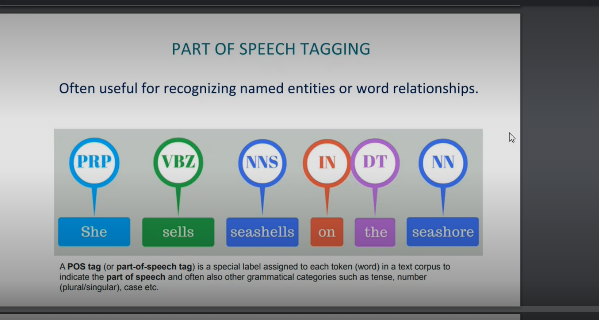
Model which u are sing

Input

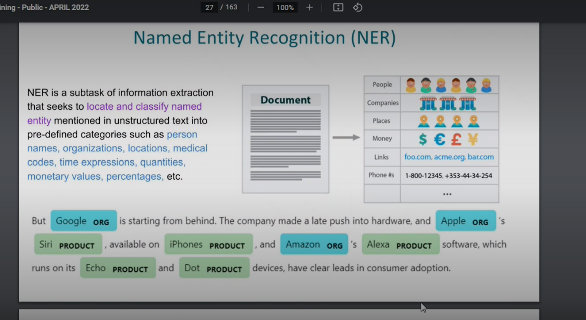
Lable column and its name

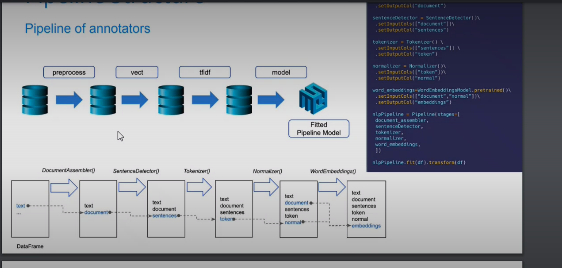


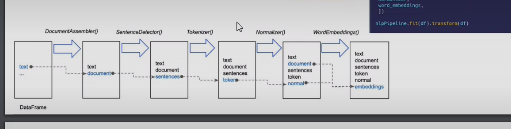
NER MODEL;

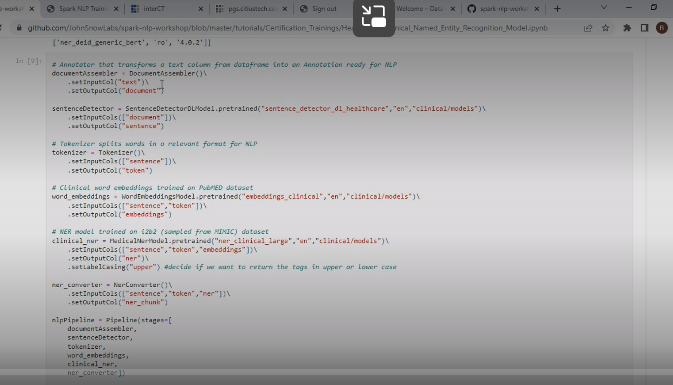


To know relation between





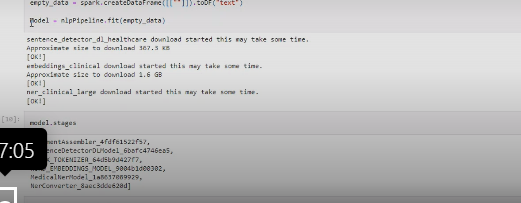


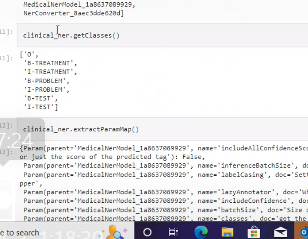


Ner converter : once we fit our model to dataset

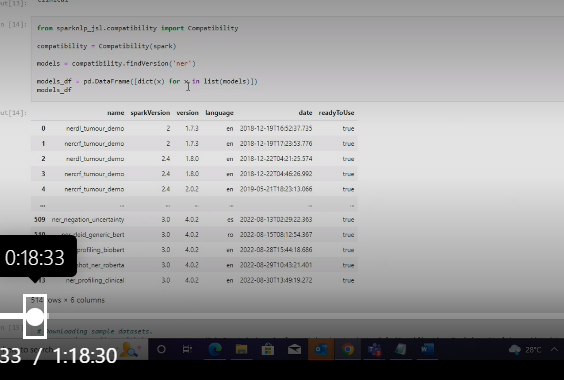
We will get many ner

Using this will convert only specifies in this model





Diff ner models

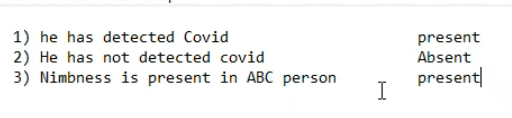


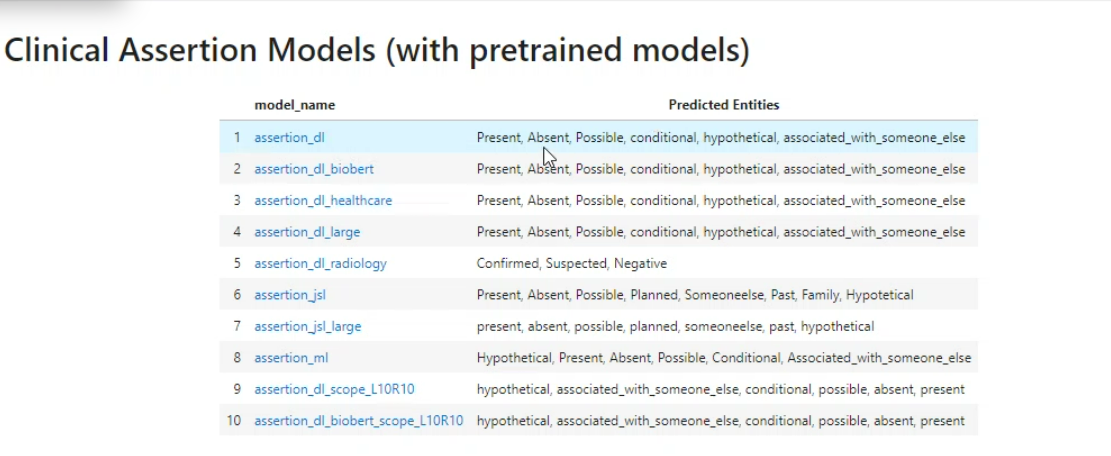
Assertion

Desease is present – means assertion present

Disease not present

Numbness is there --- assertion present

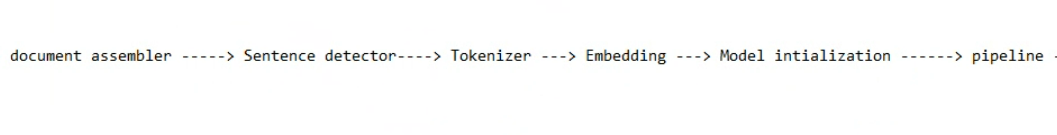


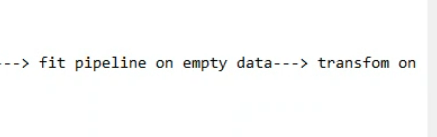


Models and predicted entities or detection assertion (status)

Pipeline

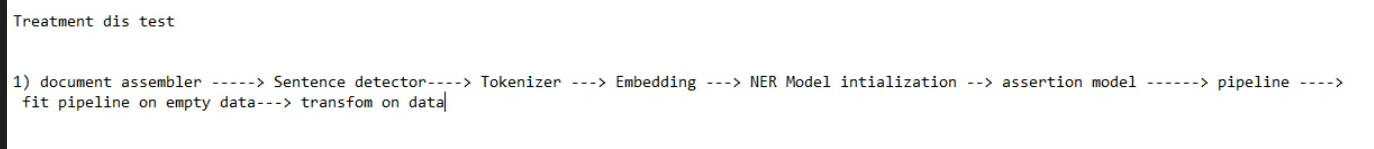
General pipeline





Assertion pipeline

1st I build ner model to detect entities and then assertion model

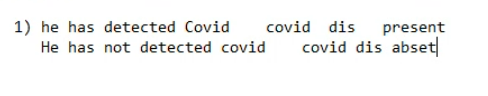


Assertions

Disease present or not

Disease present or not

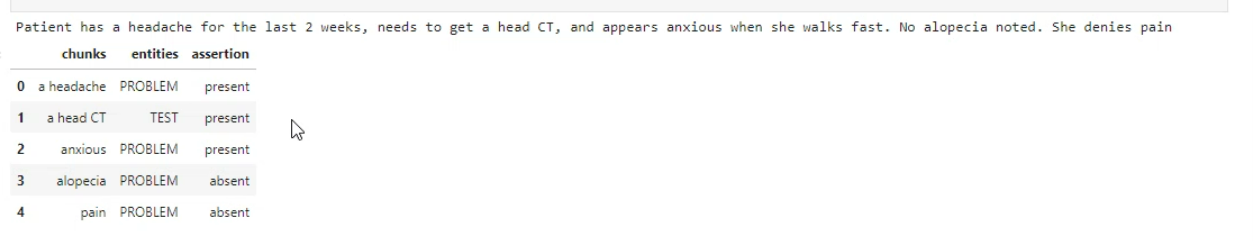
Treatment given or not











[Google Colab](https://colab.research.google.com/github/JohnSnowLabs/spark-nlp-workshop/blob/master/tutorials/Certification_Trainings/Healthcare/17.Graph_builder_for_DL_models.ipynb#scrollTo=mKyHz-v_T7o6)

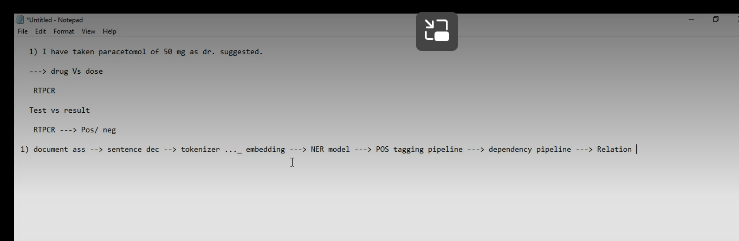
Assertion

\

[Training Custom NER models in SpaCy to auto-detect named entities [Complete Guide] (machinelearningplus.com)](https://www.machinelearningplus.com/nlp/training-custom-ner-model-in-spacy/)

Relation

Drug and dose

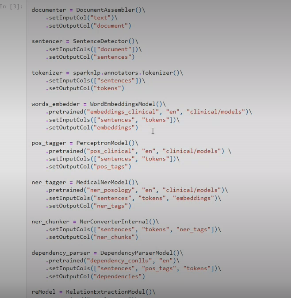


<https://github.com/JohnSnowLabs/spark-nlp-workshop/tree/master/tutorials/Certification_Trainings/Healthcare>

instead of org printing name

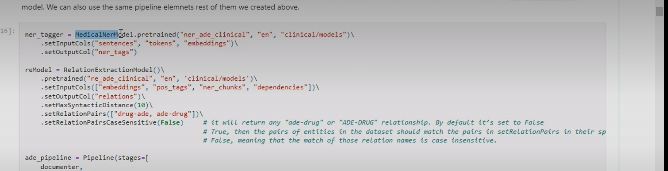


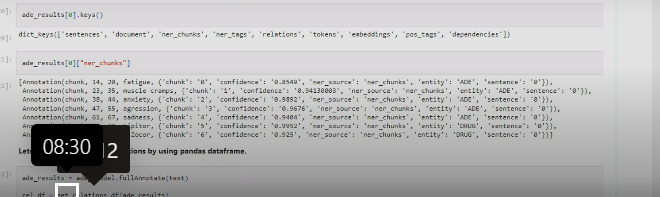
Updating spacy model or retraining model

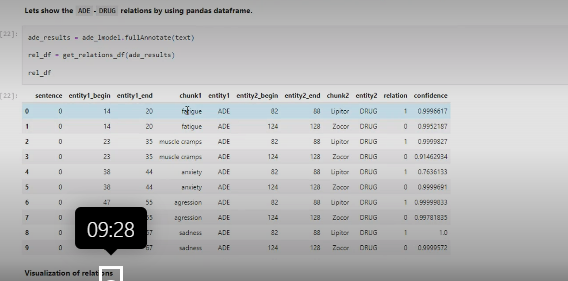


Clinical relationship extraction

Relation between drug and dosage







Use case:

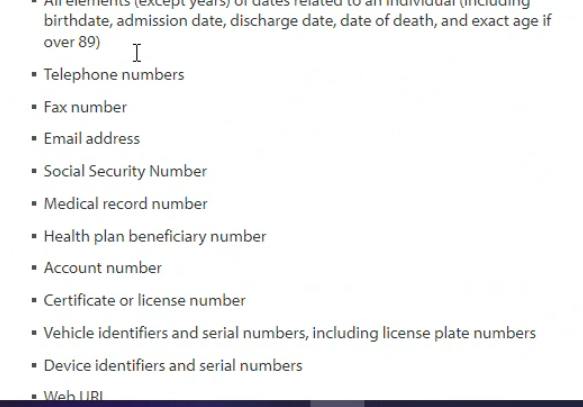
Data de identification:

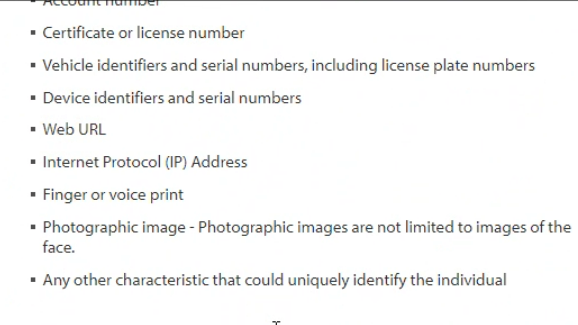
Masking the information which is protected and send other info

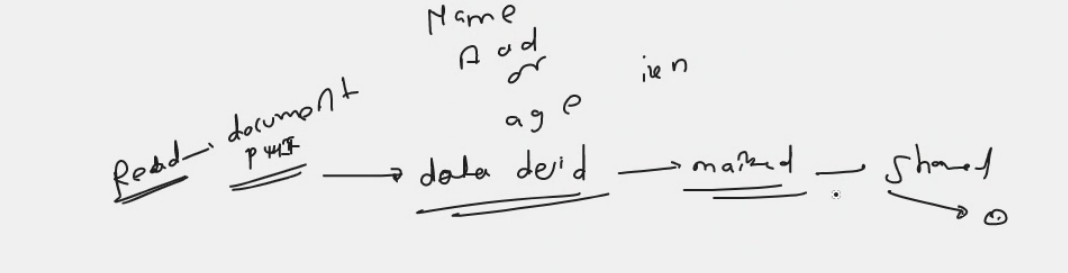


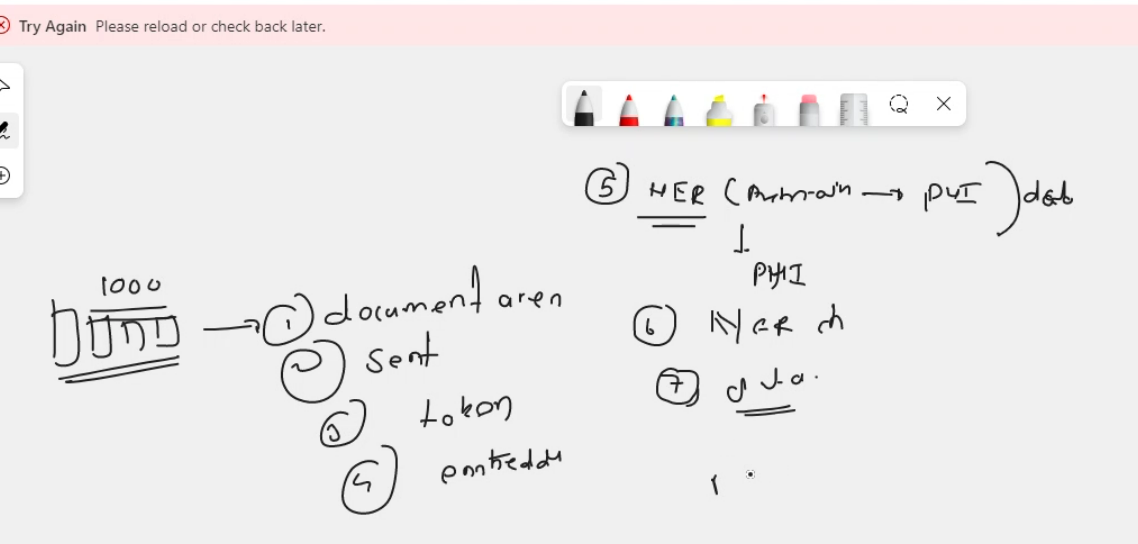
Hippa identifies spme entities as phi

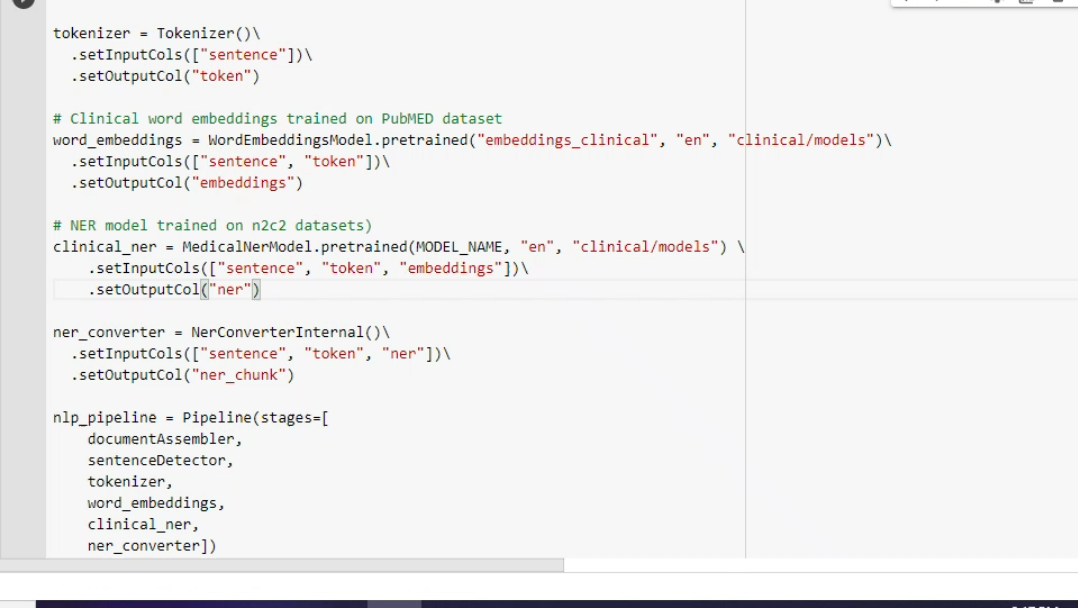
18 entities













Cache()