Natural Language Processing

## Introduction:

* There are many use cases available for nlp reference such as Gmail writing where we get auto completion, spam filters, language translation, customer service chatbot, voice assistance
* Google uses BERT model to answer your questions.
* NLP is a field in computer science and AI that gives machines an ability to understand human language better and to assist in language related tasks.
* Libraries which we will be using are python, spaCy, Gensim, NLTK, sklearn, tensorflow, pytorch, hugging face etc.
* Why nlp is booming now?

1. There are freely trained/pretrained models such as fasttext, Tensorflow hub, GTP3 are available.
2. With the help of transfer learning we can take the model which google has created for use. Tensorflow provides such models.
3. Cheap hardware & cloud resources.
4. Learning resources
5. Huge investment by Big Tech

## Regex for NLP:

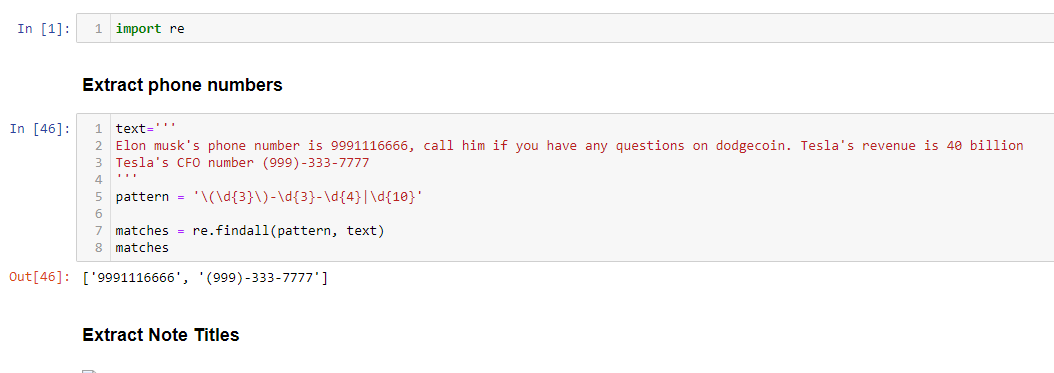
1. Customer service chatbot
2. Information extraction task

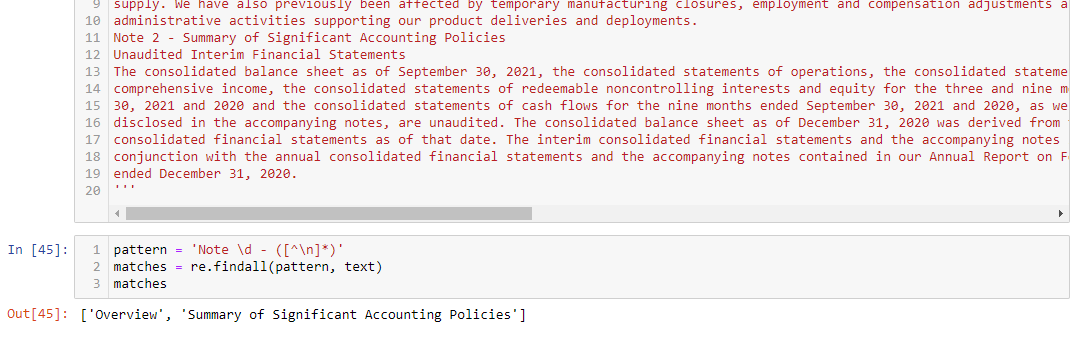
Regular expression – Matching pattern in a text & retrieving key information out of it.

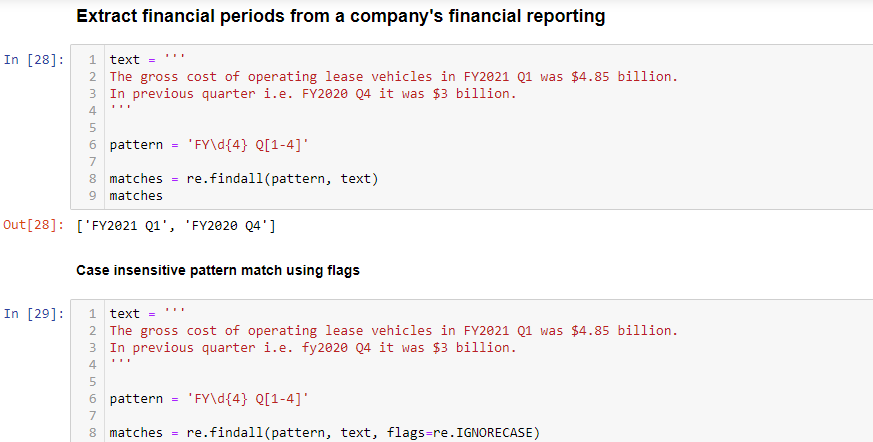
* Chatbot uses pattern matching
* Common patterns to find out
* We need python, git-bash for that.
* (before moving forward we need to learn regular expressions and if \_\_name\_\_==”\_\_main\_\_”)

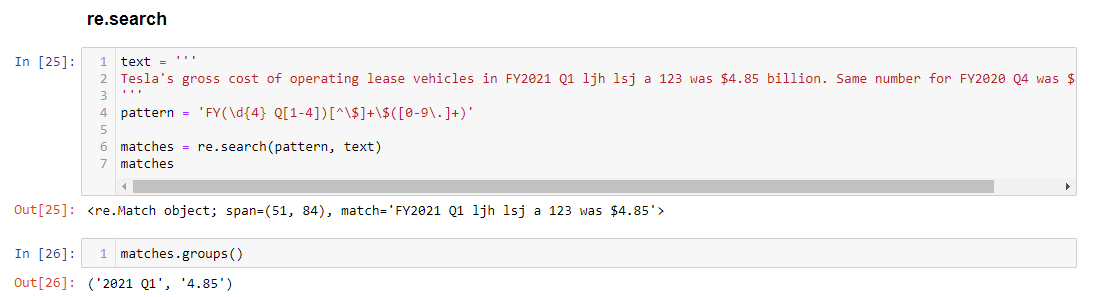
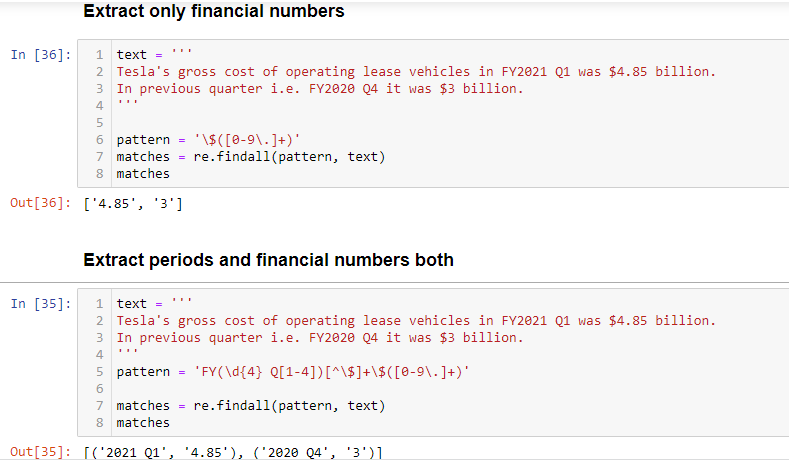
(Regular expressions in python):

All the files are added on github to check.









pattern = '\(\d{3}\)-\d{3}-\d{4}|\d{10}'

pattern = 'Note \d - ([^\n]\*)'

pattern = 'FY\d{4} Q[1-4]'

pattern = '\$([0-9\.]+)'

pattern = 'FY(\d{4} Q[1-4])[^\$]+\$([0-9\.]+)'

pattern = 'FY(\d{4} Q[1-4])[^\$]+\$([0-9\.]+)'

back to Regex in NLP:

* Open git bash and the required directory

Python –m notebook

To launch a jupyter notebook

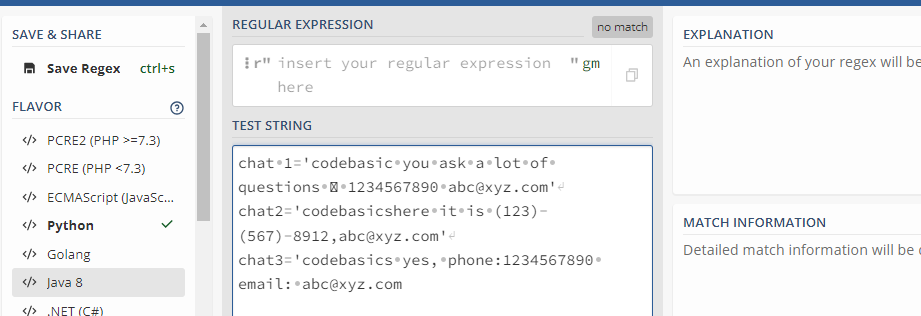
Create a new jupyter notebook for “Regex for NLP”

And in built library:

import re

open regex101.com

and paste a chat as below:



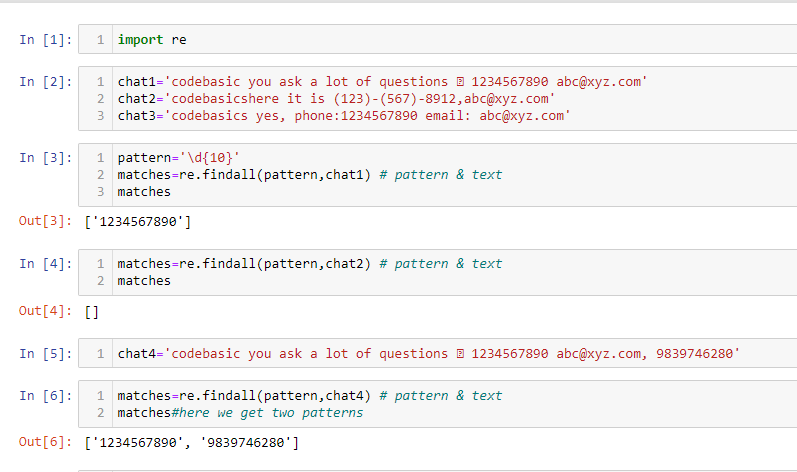
We want to extract 10 consecutive digits’ sequence.

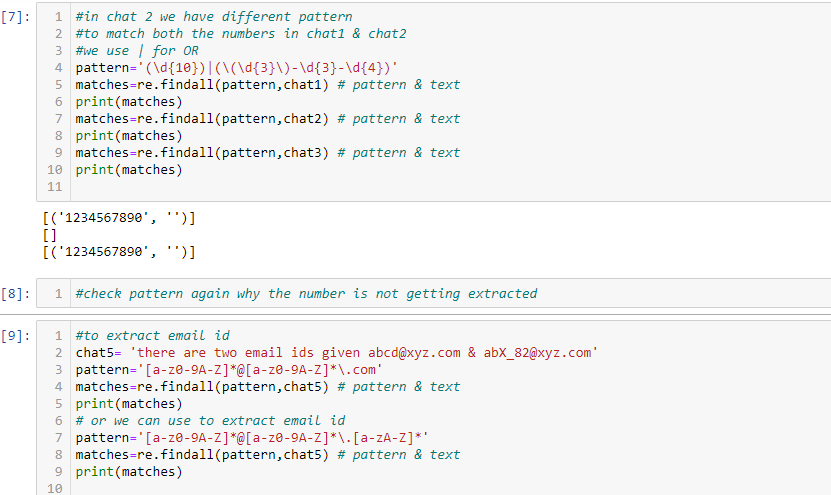
That is we want to extract phone number and email.

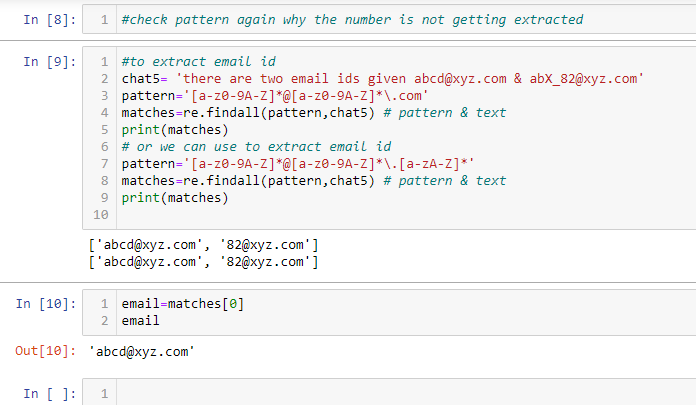
chat 1='codebasic you ask a lot of questions  1234567890 abc@xyz.com'

chat2='codebasicshere it is (123)-(567)-8912,abc@xyz.com'

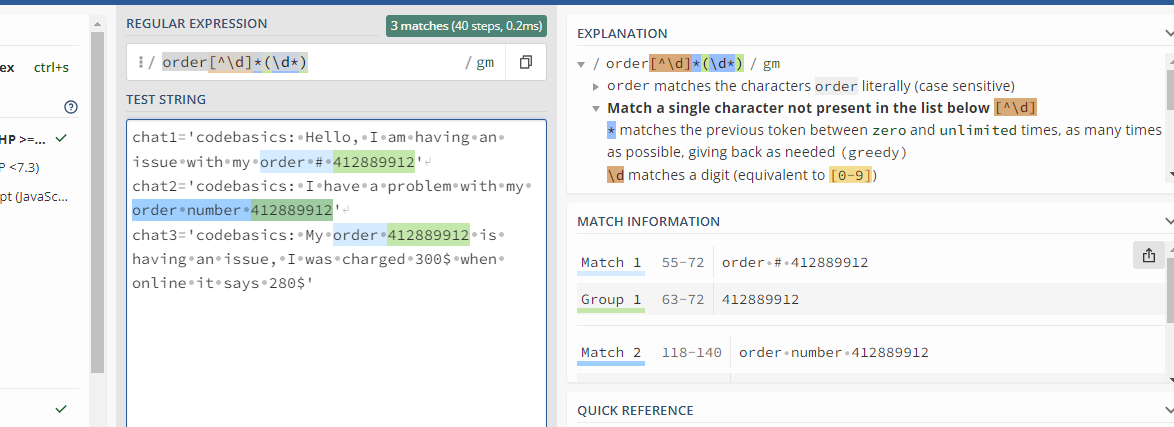
chat3='codebasics yes, phone:1234567890 email: [abc@xyz.com](mailto:abc@xyz.com)

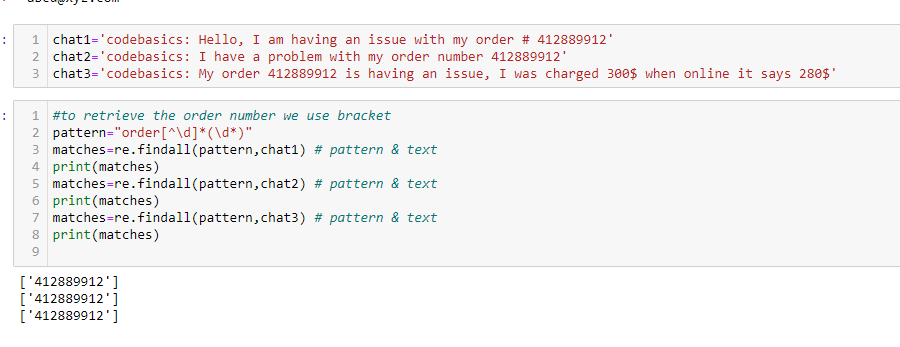






To retrieve the order number





Information Extraction





