NATURAL LANGUAGE PROCESSING LAB ASSIGNMENT - 3 Course Code: CSE3015 Course Title: Natural Language Processing Professor: Prof.V SRIKANTH REDDY Slot: L47+L48 Name: P. Sathwika Reg. No.: 21BCE8118 #1.Implement Regex tagger import re text = ' john is learning python!python python' x= re.sub('python','the latest python', text) print(x) john is learning the latest python!the latest python the latest python #2. Implement Brill tagger? import nltk from nltk.tag import BrillTaggerTrainer from nltk.tag import brill from nltk.corpus import treebank nltk.download('treebank') corpus = treebank.tagged_sents() train_data = corpus[:3000] test_data = corpus[3000:] baseline_tagger = nltk.UnigramTagger(train_data) brill_trainer = BrillTaggerTrainer(baseline_tagger, templates=brill.brill24()) brill_tagger = brill_trainer.train(train_data, max_rules=10) sentence = "This is an example sentence." tokens = nltk.word_tokenize(sentence) tags = brill_tagger.tag(tokens) print("Tagged sentence:", tags)

Tagged sentence: [('This', 'DT'), ('is', 'VBZ'), ('an', 'DT'), ('example', 'NN'), ('sentence', None), ('.', '.')]

[nltk_data] Downloading package treebank to /root/nltk_data...
[nltk data] Package treebank is already up-to-date!

```
#3. Implement Max Entropy Classifier
from scipy.stats import entropy as en
a=[1,2,3,4]
import pandas as pd
a=pd.Series(a)
data=a.value_counts()
en(data)
    1.3862943611198906
#4. Implement NER tagger.
import nltk
nltk.download('punkt')
nltk.download('averaged perceptron tagger')
nltk.download('maxent_ne_chunker')
nltk.download('words')
from nltk import word tokenize
from nltk import pos_tag
from nltk import ne chunk
nltk.download('averaged perceptron tagger')
text= 'sundar pichai is one of the CEO of company Google and was born in Chennai.'
tokens=nltk.word tokenize(text)
print(tokens)
tag=nltk.pos tag(tokens)
print(tag)
ne=nltk.ne chunk(tag)
print(ne)
    ['sundar', 'pichai', 'is', 'one', 'of', 'the', 'CEO', 'of', 'company', 'Google', 'and', 'was', 'born', 'in', 'Chennai', '.']
     [('sundar', 'NN'), ('pichai', 'NN'), ('is', 'VBZ'), ('one', 'CD'), ('of', 'IN'), ('the', 'DT'), ('CEO', 'NNP'), ('of', 'IN'), ('company', 'NN'), ('Google', 'NNP'), ('and', 'CC'), ('wa
     (S
       sundar/NN
       pichai/NN
       is/VBZ
       one/CD
       of/IN
       the/DT
       (ORGANIZATION CEO/NNP)
       of/IN
       company/NN
       (PERSON Google/NNP)
       and/CC
       was/VBD
       born/VBN
       in/IN
       (GPE Chennai/NNP)
       ./.)
     [nltk data] Downloading package punkt to /root/nltk data...
     [nltk_data] Package punkt is already up-to-date!
     [nltk_data] Downloading package averaged_perceptron_tagger to
     [nltk data]
                    /root/nltk data...
     [nltk data] Package averaged perceptron tagger is already up-to-
     [nltk data]
                      date!
     [nltk_data] Downloading package maxent_ne_chunker to
     [nltk data]
                    /root/nltk data...
     [nltk_data] Package maxent_ne_chunker is already up-to-date!
     [nltk data] Downloading package words to /root/nltk_data...
     [nltk_data] Package words is already up-to-date!
```