Extract all named entities? de Empord nill drem niek. tree Impart Tree from . NIK. tokonize Proport Word-Actorize from Milk. tag Pomport pos-tag from nette. Chark import no chunk nItt. download ( 'printet') nith. download ( laveraged - poscoption tagger') Mith. down load ( monxent - ne - churker!) nite. decembord ('words') Sendence 1 = "Paghamar . said on Honday . Hoat WASHINGTON & -= In the water ob a stringuisof about by New York. police roffreers in the 1990s, boxetta E- Lych, the top strued that Agrican-A mericans felt and said the responsibility. for raparing a generations of miscommunications and mistrust fall to law enforcement." tokens = word tokenizer (sentence 1) tage = pos - tag (tokens) re-free- ne-dunk (lags) print (re-tree [:]) ne livee = ne - chunt (pos-tag (word-foxenize (sentenos))) dox Pin re-tree; print(i)

Exorage - 1

# Natural Language Processing Lab Lab 10. Named Entity Recognition

In this lab, you will extract named entities from the given text file using NLTK. You will also recognize entities based on the regular expression patterns.

## **EXERCISE-1**

Extract all named entities from the following text:

Sentence1 = "Rajkumar said on Monday that WASHINGTON -- In the wake of a string of abuses by New York police officers in the 1990s, Loretta E. Lynch, the top federal prosecutor in Brooklyn, spoke forcefully about the pain of a broken trust that African-Americans felt and said the responsibility for repairing generations of miscommunication and mistrust fell to law enforcement."

#### Source Code:

```
import nltk
from nltk.tokenize import word_tokenize
from nltk.tag import pos_tag
from nltk.chunk import ne_chunk

tokens = word_tokenize(sentence1)
tags = pos_tag(tokens)
ne_tree = ne_chunk(tags)
print(ne_tree)

You can create a pipeline too:
ne_tree = ne_chunk(pos_tag(word_tokenize(sentence1)))
```

#### Question-1

 Count and print the number of PERSON, LOCATION and ORGANIZATION in the given sentence.

## **Question-2**

- Observe the results. Does named entity, "police officers" get recognized?.
- Write a regular expression patter to detect this. You will need nltk.RegexpParser class to define pattern
  and parse terms to detect patterns.

### Question-3

- Does the named entity, "the top federal prosecutor" get recognized?.
- Write a regular expression pattern to detect this.

#### **EXERCISE-2**

Extract all named entities from the following text:

sentence2 = "European authorities fined Google a record **\$5.1 billion** on Wednesday for abusing its power in **the mobile phone** market and ordered **the company** to alter its practices"

## Question-1

Observe the output. Does your code recognize the NE shown in BOLD?

DEPT OF DATA SCIENCE BISHOP HEBER COLLEGE (TRICHY

DR. K. RAJKUMAR.

Gla dim. 1 Toyont ofte from Collections Propert Countes. for chunk in reliber: if tradito (chunk, (label)): print ([lawter (label) for label in chunk]) Guestion 12 D word-Alt. word-dokenta (sentence 1) por tag = nltk. pos tagr (word) chunk = neth. ne - chunk (pos tag) Frammon: JUND: JENNS CHUS) Cp = nitt. Regexprasses (grammar) rosult = cp. parse(cheink) ME = [" ", goin (w for w, 6 9n ele). for ele in result . Et isinstan e (ele, nltk. whee)) print (NE) ( grannar = "NP: f < NN < NNS) }". cp = ntk. Ragexplansex (grammar) Yosut = . Cp. parse (ne-tree) NE = [" . " 30 Po (w for W, t: frelse) to ele in result if ixinstance (ele)

Print (NE)

nith wroce)

Write a regular expression that recognizes the entity, "\$5.1 billion"
Detect and print this

# Question-2

Write a regular expression that recognizes the entity, "the mobile phone" and similar to this entity such as "the company"

## **EXERCISE-3**

In this exercise, you will extract all ingredients from the food recipes text file, food\_recipes.txt". For example, the following text shows one food recipe.

BEEF TENDERLOIN STEAKS WITH SMOKY BACON-BOURBON SAUCE Serves: 4

- 1 1/2 cups dry red wine
- 3 cloves garlic
- 1 3/4 cups beef broth
- 1 1/4 cups chicken broth
- 1 1/2 tablespoons tomato paste
- 1 bay leaf

- 1 sprig thyme
- 8 ounces bacon cut into 1/4 inch pieces
- 1 tablespoon flour
- 1 tablespoon butter
- 4 1 inch rib-eye steaks
- 1 tablespoon bourbon whiskey

The ingredients are highlighted with BOLD in the above list.

Extract all Named Entities from the text file and display them.

Reference: https://sites.google.com/site/anu3bis/recipes-main.

DEPT OF DATA SCIENCE (BISHOP HEBER COLLEGE) TRICHY

DR. K. RAJKUMAR

Sentence 2: 11 European authorities fined google a record \$ 5.1 billion on wednesday for alousing its Dover in the mobile phone market and ordered there company to after its posetice! fok = wood\_topenize (centence 2) tagged = nltk. Pos\_tag(tok) Newfree 2 = nltk. ne - Chun k (tagged, binorry = False) point (re-tree 2 [:]) Write a regular expression that reagnizes the entity. Word = nltr- Word\_tokenize (sentence 2) pos-tag=hlfk.fag (wood) churic = nltk. re-churk (pos-tag) grammar = "NP: { CCD) (CDT) & JI) & KNDY Cp = nitle. Regexpranser (grammars) result = cp. parse (chunk) NE = [" ! join (w for w, tir ele) for ele in rosult if isinstance (elg. nHIC. b Tree) Baut (NE) Question -2. word = nlde totente (sentenced) | result = cp. parse (chunk) pos-tag= Nith. pos-tagr(word) WE = [" " foin (w. to w, t chunk = nlft. ne-chumb ( postq) mele) for ele in result 91: Prostance (ele, nH/c ) tree) grannar = "NP: (<pT) <TT) KHY Cp = MIER. Regext Parses (grammas)