**Praktikum NLP dengan Python**

* Editor Python Online

<https://repl.it/languages/python3>

* Contoh source code <https://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk>, <https://www.nltk.org/book/ch08.html>
* **Kode program 1 : sent\_tokenize**

import nltk

nltk.download('punkt')

from nltk.tokenize import sent\_tokenize, word\_tokenize

text="Hello Mr. Smith, how are you doing today? The weather is great, and city is awesome.The sky is pinkish-blue. You shouldn't eat cardboard"

tokenized\_text=sent\_tokenize(text)

print(tokenized\_text)

* **Kode program 2 : word\_tokenize**

tokenized\_word=word\_tokenize(text)

print(tokenized\_word)

* **Kode program 3 : FreqDist**

from nltk.probability import FreqDist

fdist = FreqDist(tokenized\_word)

print(fdist)

* **Kode program 4 : stopword**

from nltk.corpus import stopwords

stop\_words=set(stopwords.words("english"))

print(stop\_words)

* **Kode program 5 : removing stopword**

filtered\_sent=[]

for w in tokenized\_sent:

 if w not in stop\_words:

 filtered\_sent.append(w)

print("Tokenized Sentence:",tokenized\_sent)

print("Filterd Sentence:",filtered\_sent)

* **Kode program 6 : Stemming**

# Stemming

from nltk.stem import PorterStemmer

from nltk.tokenize import sent\_tokenize, word\_tokenize

ps = PorterStemmer()

stemmed\_words=[]

for w in filtered\_sent:

 stemmed\_words.append(ps.stem(w))

print("Filtered Sentence:",filtered\_sent)

print("Stemmed Sentence:",stemmed\_words)

* **Kode program 7 : stemming and Lemmatization**

#Lexicon Normalization

#performing stemming and Lemmatization

from nltk.stem.wordnet import WordNetLemmatizer

lem = WordNetLemmatizer()

from nltk.stem.porter import PorterStemmer

stem = PorterStemmer()

word = "flying"

print("Lemmatized Word:",lem.lemmatize(word,"v"))

print("Stemmed Word:",stem.stem(word))

* **Kode program 8 : POS Tagging**

sent = "Albert Einstein was born in Ulm, Germany in 1879."

tokens=nltk.word\_tokenize(sent)

print(tokens)

nltk.pos\_tag(tokens)

* **Kode program 9 : Parse Tree**

import nltk

nltk.download('punkt')

from nltk.tokenize import word\_tokenize

text = "I shot an elephant in my pajamas"

tokenized\_word=word\_tokenize(text)

print(tokenized\_word)

groucho\_grammar = nltk.CFG.fromstring("""

S -> NP VP

PP -> P NP

NP -> Det N | Det N PP | 'I'

VP -> V NP | VP PP

Det -> 'an' | 'my'

N -> 'elephant' | 'pajamas'

V -> 'shot'

P -> 'in'

""")

parser = nltk.ChartParser(groucho\_grammar)

for tree in parser.parse(tokenized\_word):

 print(tree)

 tree.draw()



* **Kode program 10**

Ubahlah grammar dalam Bahasa Indonesia bahwa sebuah kalimat tersusun dari kata benda dan kata kerja.

**Kalimat 🡪 KataBenda KataKerja**

* **KataBenda 🡪 {ayam, kucing, budi}**
* **KataKerja 🡪 {makan, minum, baca}**

import nltk

from nltk.tokenize import word\_tokenize

text = "ayam makan"

tokenized\_word=word\_tokenize(text)

print(tokenized\_word)

groucho\_grammar = nltk.CFG.fromstring("""

S -> KATA\_BENDA KATA\_KERJA

KATA\_BENDA -> 'ayam' | 'kucing' | 'budi'

KATA\_KERJA -> 'makan' | 'minum' | 'membaca'

""")

parser = nltk.ChartParser(groucho\_grammar)

for tree in parser.parse(tokenized\_word):

 print(tree)

 #tree.draw()

Output :

['ayam', 'makan']

(S (KATA\_BENDA ayam) (KATA\_KERJA makan))

* **Kode program 10**

Kembangkan kode program 9 dengan membuat kalimat sederhana yang tersusun dari Subjek Predikat atau Subjek Predikat Objek

import nltk

from nltk.tokenize import word\_tokenize

text = "saya makan ayam"

tokenized\_word=word\_tokenize(text)

print(tokenized\_word)

groucho\_grammar = nltk.CFG.fromstring("""

S -> SUBJEK PREDIKAT | SUBJEK PREDIKAT OBJEK

SUBJEK -> KATA\_BENDA | KATA\_GANTI | FRASA

PREDIKAT -> KATA\_KERJA

OBJEK -> KATA\_BENDA | KATA\_GANTI

KATA\_GANTI -> 'saya' | 'kami'

KATA\_BENDA -> 'ayam' | 'kucing' | 'buku'

KATA\_KERJA -> 'makan' | 'minum' | 'membaca'

""")

parser = nltk.ChartParser(groucho\_grammar)

for tree in parser.parse(tokenized\_word):

 print(tree)

 #tree.draw()

Output :

['saya', 'makan', 'ayam']

(S

 (SUBJEK (KATA\_GANTI saya))

 (PREDIKAT (KATA\_KERJA makan))

 (OBJEK (KATA\_BENDA ayam)))