

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 : stopwords

```
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()

```

