Artificial Intelligence
(Natural Language Processing: Tokenize, POS Tagging

Prof K R Chowdhary

CSE Dept., MBM University

January 24, 2025
Lecture #8 tok, pos

Tokenize

Python program uses Natural Language Toolkit (NLTK) to tokenize the given text into individual words and then assigns part-of-speech (POS) tags to each word.

Tokenization: The text is split into words using the word_tokenize function. For the given example, the tokens would be something like:

```
"Hello"
"!"
"How"
"are"
"you"
"today"
"?"
"I"
"hope"
"you're"
"having"
"a"
"great"
"time"
"."
```

POS Tagging: The pos_tag function then assigns a part-of-speech label to each token. Some common POS tags that might be seen in the output include:

Parts of Speech (POS)

```
NN (Noun, singular)
VB (Verb, base form)
DT (Determiner)
JJ (Adjective)
PRP (Pronoun)
IN (Preposition)
NNP (Proper noun,
    singular)
VBZ (Verb, 3rd person
    singular present)
UH (Interjection)
```

Original Text:
"Hello! How are you today? I hope you're having a great time."

Tokenized Words and Their
POS Tags:

```
"Hello" -> UH (Interjection)
"!" -> Punctuation (This is
    usually represented as .
    or !)
"How" -> WRB (Wh-adverb, used
    for questions like "how")
"are" -> VBP (Verb, non-3rd
    person singular present)
"you" -> PRP (Pronoun,
    personal)
"today" -> NN (Noun, singular)
"?" -> Punctuation
```

Tokenized Words and Their POS Tags:

```
"I" -> PRP (Pronoun,
    personal)
"hope" -> VB (Verb, base
    form)
"you're" -> VB (Verb, base
    form) + PRP (Pronoun)
"having" -> VBG (Verb,
  gerund/present
  participle)
```

```
"a" -> DT (Determiner)
"great" -> JJ (Adjective)
"time" -> NN (Noun, singular)
"." -> Punctuation
```

In summary, the output of the POS tagging is a list of tuples where each token (word) is paired with its corresponding part-of-speech tag, allowing you to analyze the grammatical structure of the text

[1] Chowdhary, K.R. (2020). Natural Language Processing. In: Fundamentals of Artificial Intelligence. Springer, New Delhi. https://doi.org/10.1007/978-81-322-3972-7_19