

# Python 101

Lec -1

Real Life Applications

thoum

June 4, 2019

# Why Python

One of the reasons we learn Python is the various libraries others have written for Python.

We don't have to start from scratch.

# Why Python

Numpy: Calculation

OpenCV: Image Processing

Request: Web crawling

Keras: Machine Learning

KoNLPY: Korean Natural Language Processing

Ren'Py: Visual Novel Engine

... you name it.

# Today's topic

Parse tweets on twitter, use bag of words model, so that it can be the basis for many possible applications.

# Tweepy

Learning how to use external libraries by reading their documents is important.

How to get twitter live stream

# TL;DR?

I provided a skeleton code with explanations.

[Download](#)

I want more than names and texts

Other fields

Run It!



# Saving?

Look up: saving files

# Bag of Words

## 2007-01-23: State of the Union Address

George W. Bush (2001-)

abandon accountable affordable afghanistan africa aided ally anbar armed army baghdad bless challenges  
chamber chaos choices civilians coalition commanders commitment confident confront congressman constitution corps  
debates deduction deficit deliver democratic deploy dikembe diplomacy disruptions earmarks economy einstein  
elections eliminates expand extremists failing faithful families freedom fuel funding god haven ideology  
immigration impose insurgents iran **iraq** islam julie lebanon love madam marine math medicare moderation  
neighborhoods nuclear offensive palestinian payroll province pursuing **qaeda** radical regimes resolve retreat rieman  
sacrifices science sectarian senate september shia stays strength students succeed sunni tax territories  
**terrorists** threats uphold victory violence violent Waf washington weapons wesley

# Bag of Words Model

Ignore order of the text. Focus on word count.

Simple yet strong; good for document classification.(*e.g.* spam filtering)

What class can we use?

# The Crude Version

We can split our text by spaces, get word count.

We can look at most common words, number of positive words, number of negative words... yada yada to analyze sentiment on that topic.(e.g. Todays weather? Stock? Politicians?)

## Try It!

Get the top 5 most common words related your query.(Don't forget to exclude the query itself; it will *a/ways* be included.)

## For the Future: The Improved Version

Unlike English, Korean vocabs are not simply split by spaces.

Try this: KoNLPy

# Useful References for the Future

Multiprocessing in Python : Useful if you need speed.

Request Module : Useful when interacting with websites.

Time Complexity : Useful if you want to analyze your code.

Baekjoon : Useful if you want to practice some language or algorithms.

Google : Can't stress enough.

StackOverflow : Can't stress enough.