Natural Language Processing Workshop

29 SEPT 2018

* Disclaimer *

What I am sharing with you today is based on my own understanding for the domain.

The journey started when I went on exchange to undertake Natural Language Processing.

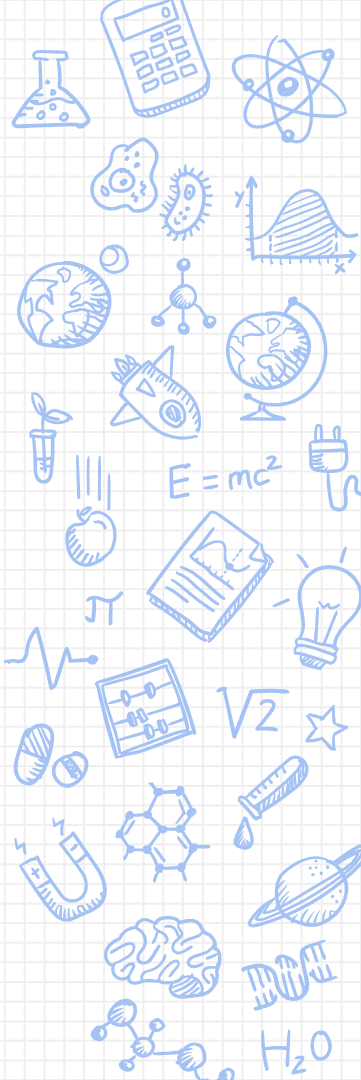
Hope that you takeaway something at the end of the session.

#HELLO!

@ZiQuan #NLP #DataAnalytics #BigData #TextProcessing #WhereGotTime #BIASMU #PythonWorkshopSeries

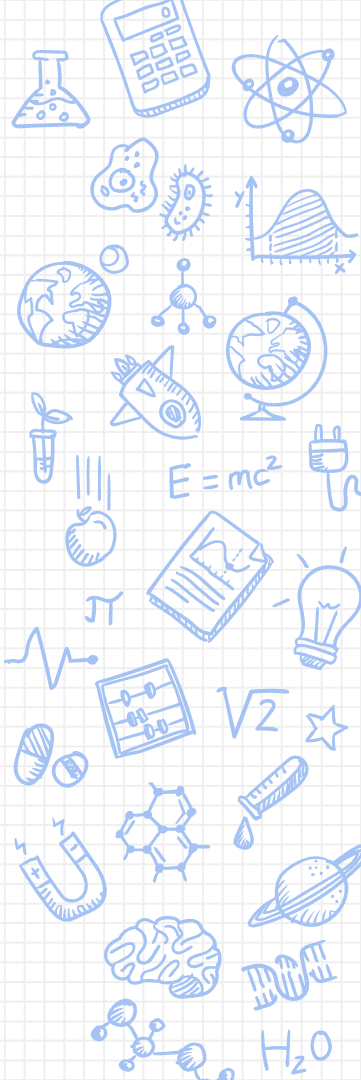
Agenda

- ✓ What is NLP
- ✓ Common NLP Applications
- ✓ NLP Pipeline
- ✓ Resource and Tools
- ✓ Hands-on Session



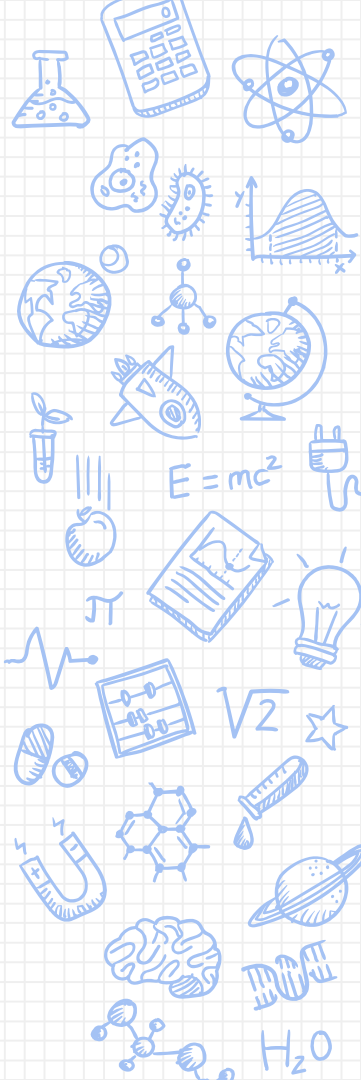
Time Breakdown

Time	Plan
10am – 10.30am	Introduction
10.30am – 11.25am	Hands On Ex 1
11.25am – 11.35am	Break
11.35am – 12.30pm	Hands On Ex 1
12.30pm – 1.30pm	Lunch
1.30pm – 2.30pm	Hands On Ex 2
2.30pm – 2.40pm	Coffee Break
2.40pm – 3.10pm	Hands On Ex 3



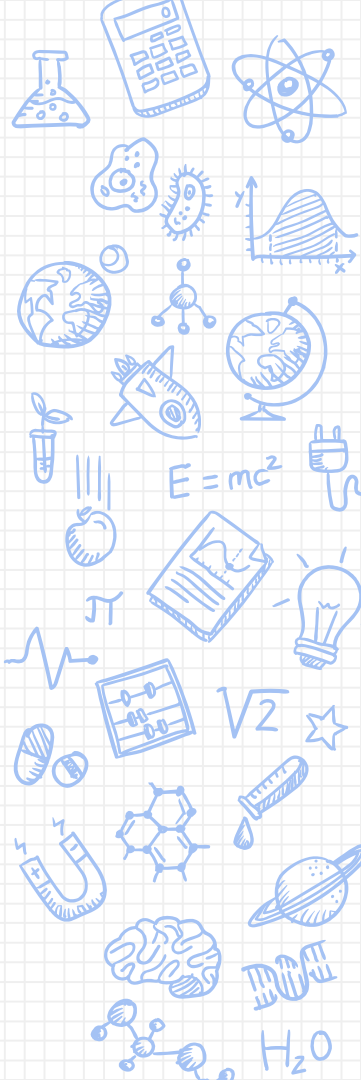
Prerequisites

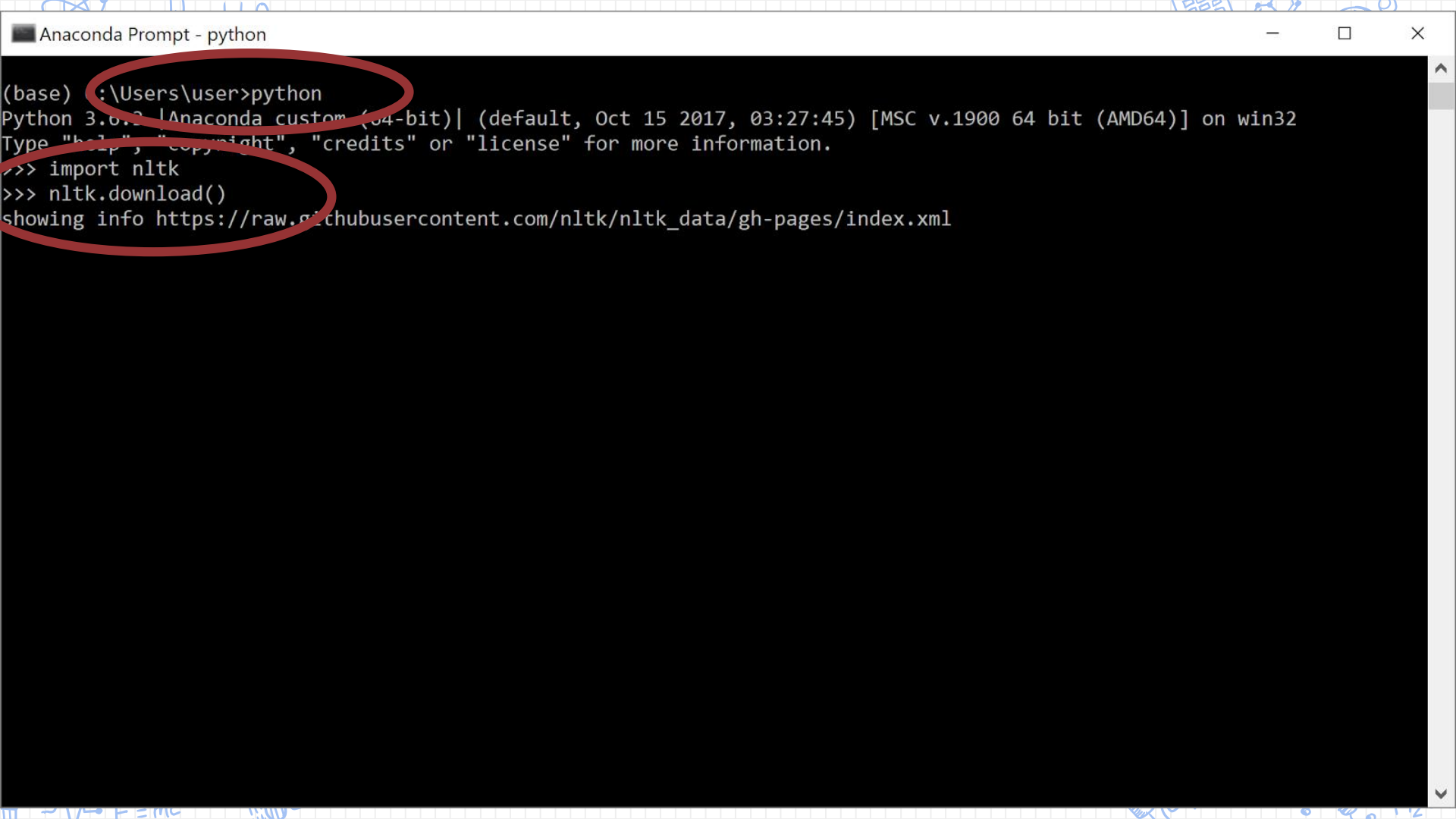
- ✓ Have a Laptop
- ✓ A Passion for Data Analytics
- ✓ Download Anaconda
- ✓ Basic Python Knowledge
- ✓ Basic Statistics Knowledge (STATS101)

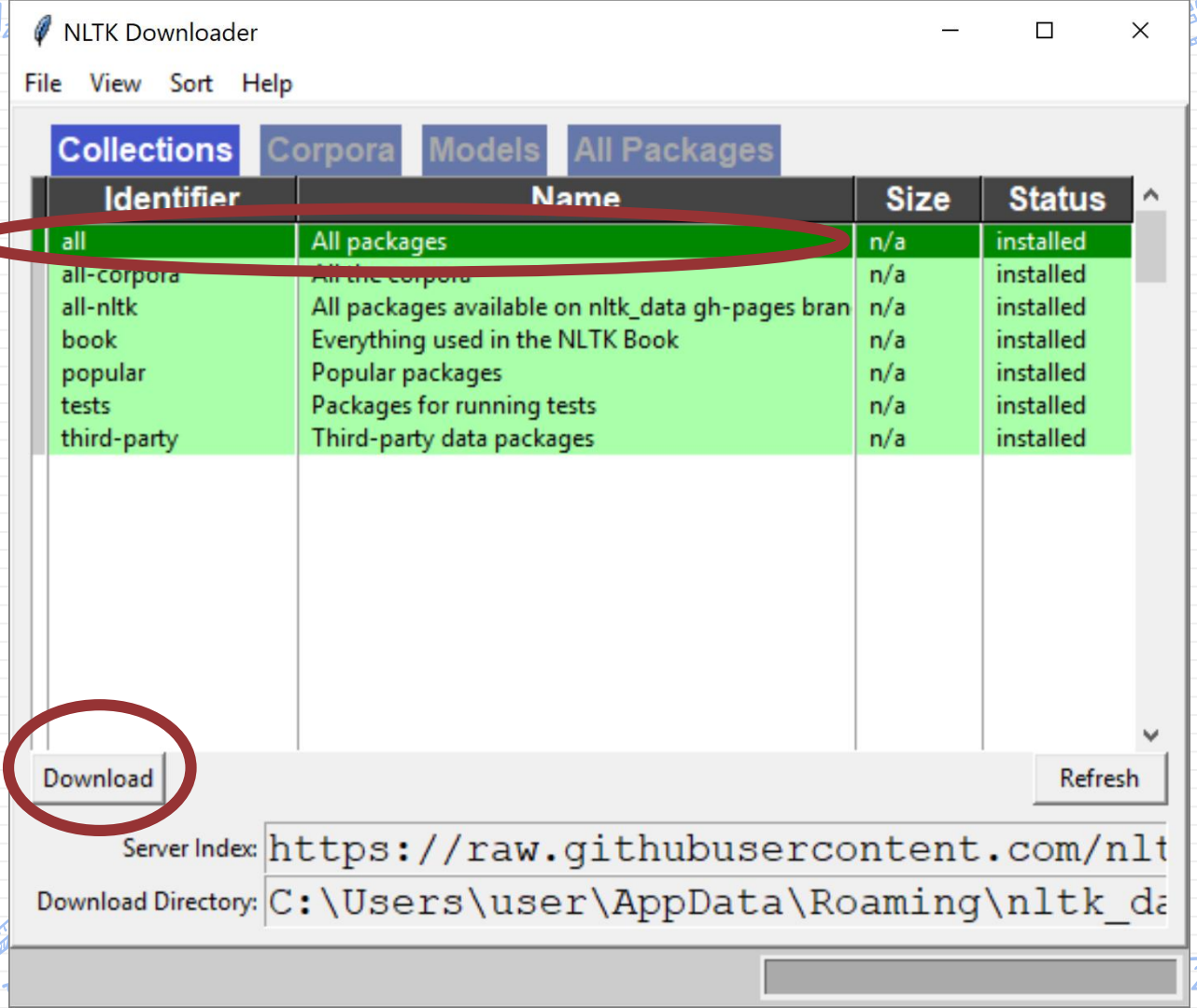


Packages Covered

- ✓ NLTK
- ✓ WordCloud
- ✓ Pandas
- ✓ Matplotlib
- ✓ Scikit Learn




Anaconda Prompt - python
(base) C:\Users\user>python
Python 3.6.3 |Anaconda custom (64-bit)| (default, Oct 15 2017, 03:27:45) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import nltk
>>> nltk.download()
showing info https://raw.githubusercontent.com/nltk/nltk_data/gh-pages/index.xml


NLTK Downloader
File View Sort Help
Collections Corpora Models All Packages
Identifier Name Size Status
all All packages n/a installed
all-corpora n/a installed
all-nltk All packages available on nltk_data gh-pages bran n/a installed
book Everything used in the NLTK Book n/a installed
popular Popular packages n/a installed
tests Packages for running tests n/a installed
third-party Third-party data packages n/a installed
Download
Refresh
Server Index: https://raw.githubusercontent.com/nlt
Download Directory: C:\Users\user\AppData\Roaming\nltk_da

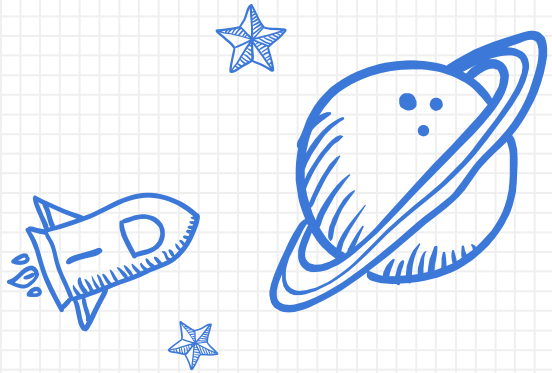
Anaconda Prompt

```
(base) C:\Users\user>conda install wordcloud=1.3.3 -c conda-forge
Solving environment: done

## Package Plan ##

  environment location: D:\Anaconda

  added / updated specs:
    - wordcloud=1.3.3


The following packages will be downloaded:

    package                    |            build
    ---------------------------|-----------------
    blas-1.0                   |              mkl           6 KB
    openssl-1.0.2p             |       hfa6e2cd_0         5.4 MB  conda-forge
    wordcloud-1.3.3            |           py36_0         153 KB  conda-forge
    numpy-1.14.2               |   py36h5c71026_0         3.7 MB
    ------------------------------------------------------------
                                           Total:         9.2 MB

The following NEW packages will be INSTALLED:

    blas:      1.0-mkl
    wordcloud: 1.3.3-py36_0          conda-forge

The following packages will be UPDATED:
```

Download Will Take Some Time

If you encounter any issues with the download

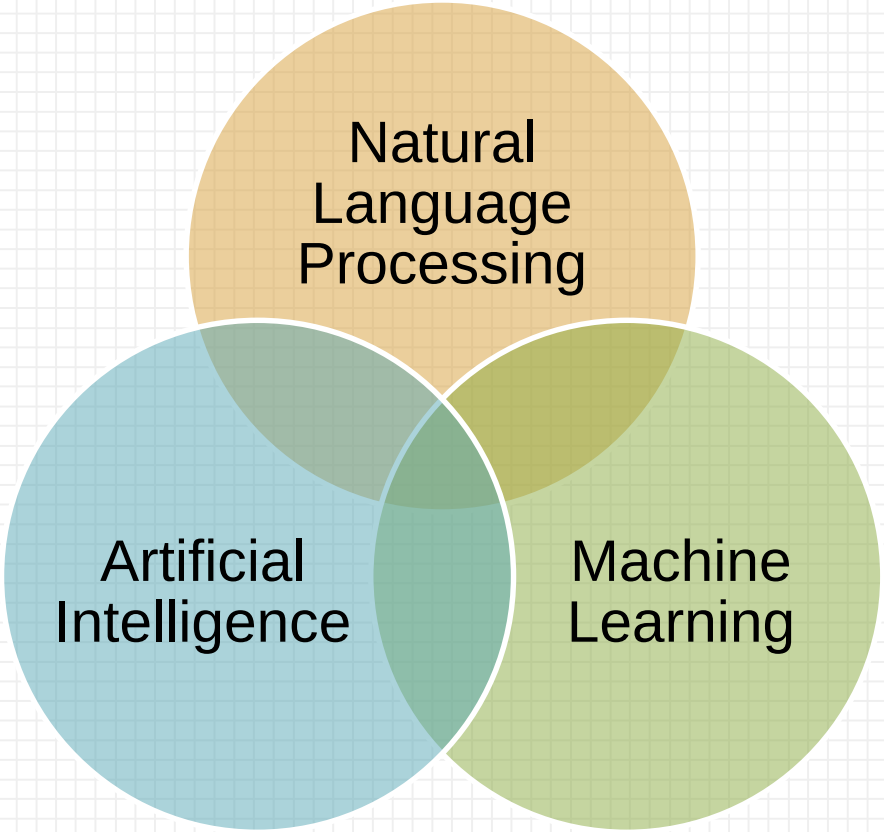
Approach our #friendly @TeachingAssistants

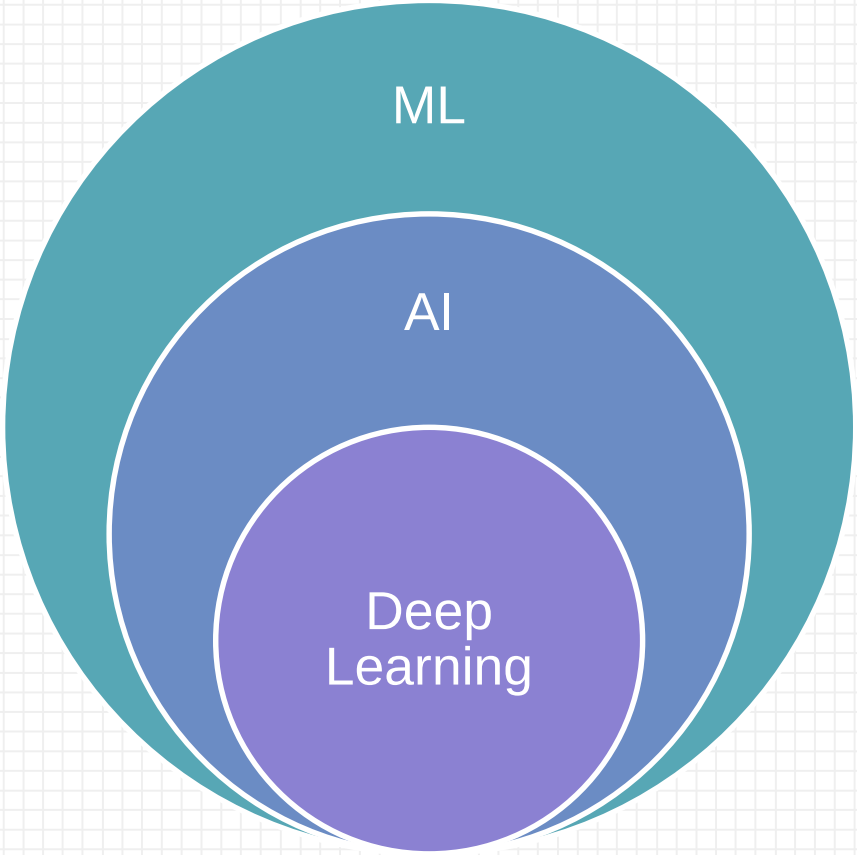


What Does Natural Language Processing Means To You?

What is Natural Language Processing

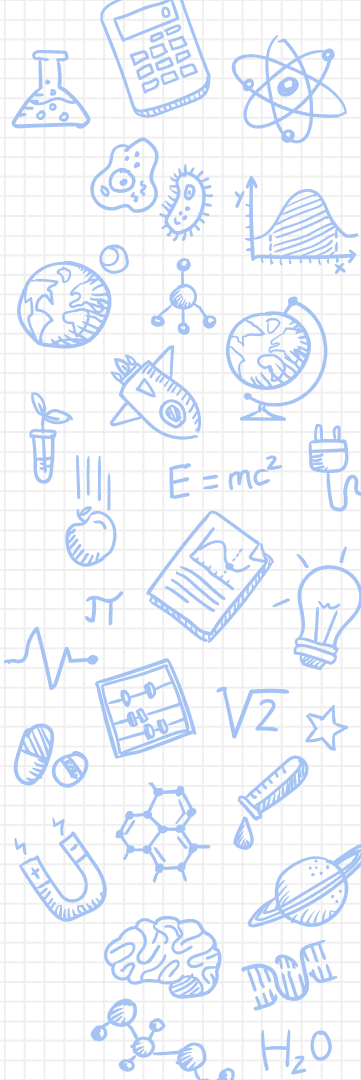
- Natural language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular **how to program computers to process and analyse large amounts of natural language data**. – Wikipedia
- In a way, if we unlock the key to **understanding how language works, we unlock the key to understanding how human brain works**.


Natural Language Processing
Artificial Intelligence
Machine Learning


ML
AI
Deep
Learning

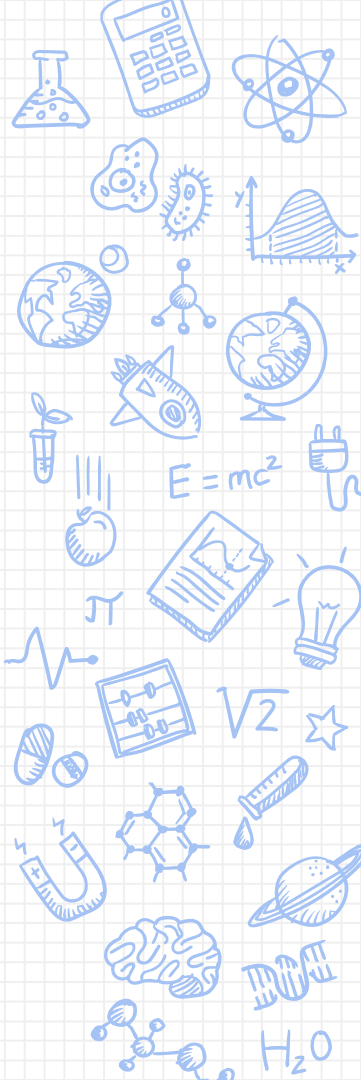
Applications of NLP

1. Text Analytics
2. Question Answering
3. Machine Translation



1. Text Analytics

- Product Marketing Information
- Political Opinion Tracking
- Social Network Analysis
- Buzz Analysis (What is hottest and latest topic right now?)



Product Marketing Information

10

anonymous user
February 2016

"Great location with lots of amenities"

Beautiful hotel in a great location. Enjoyed our stay but found the hotel very busy. Good access to the shopping and casino.

Verified review from Expedia

10

anonymous user
January 2016

"Great Hotel and Experience"

Our stay at this hotel was great we even got upgraded because it was my partners birthday. The room and the pool was amazing.

Powered by wotif

4.0

anonymous user
January 2016

"Amazing architecture, lacks soul"

Recommend a one night stay to experience pool etc, hotel lacks soul and staff appear more focussed on processing rather than serving

Verified review from Expedia

10

anonymous user

"Must stay for anyone coming to singapore"

Our stay was extremely pleasant, staff were lovely and accomodating, we were very luck to receive a room with a view of

Product Marketing Information

Related search

nintendo mario games

View 3+ more

Mario Kart DS

Mario + Rabbids Kingdom...

Tennis

New Super Mario Bros. 2

Super Mario 3D Land

Yoshi

Super Mario RPG

Feedback

Searches related to nintendo switch

nintendo switch **games**

nintendo switch **price**

nintendo switch **bundle**

nintendo switch **best buy**

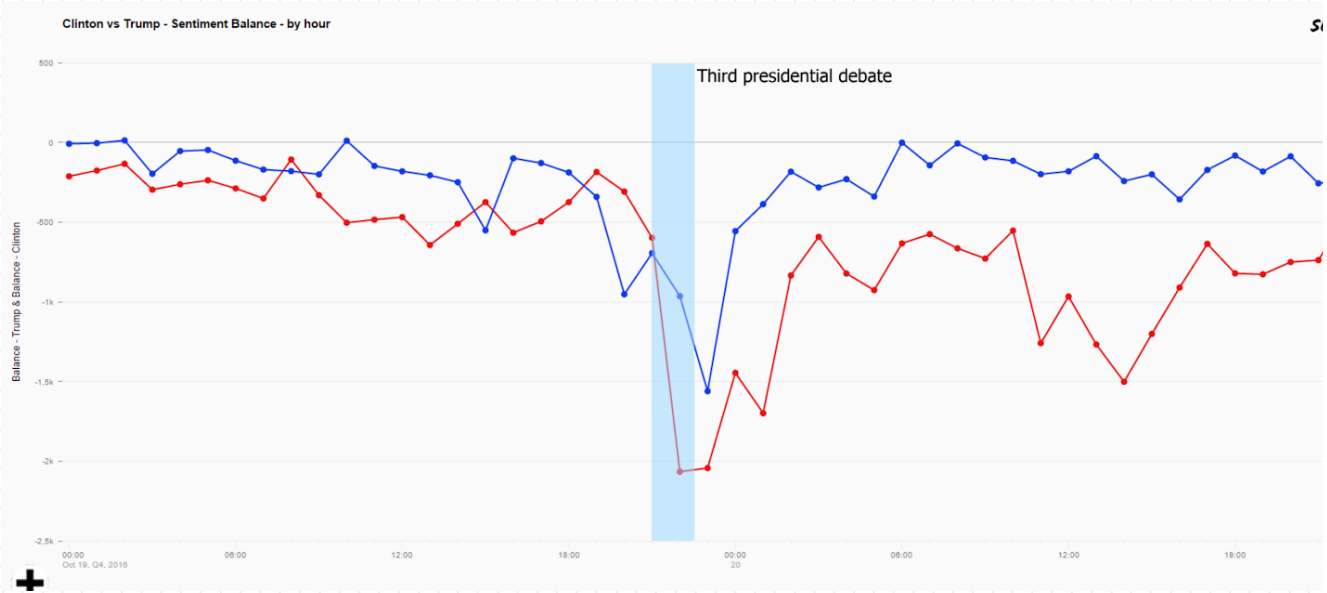
nintendo switch **amazon**

nintendo switch **review**

nintendo switch **walmart**

nintendo switch **eshop**

Opinion Tracking #MAGA



socialmention*

69% strength	6:1 sentiment
46% passion	33% reach
1 minutes avg. per mention	
last mention 6 minutes ago	
66 unique authors	
0 retweets	

Sentiment

positive	37
neutral	96
negative	6

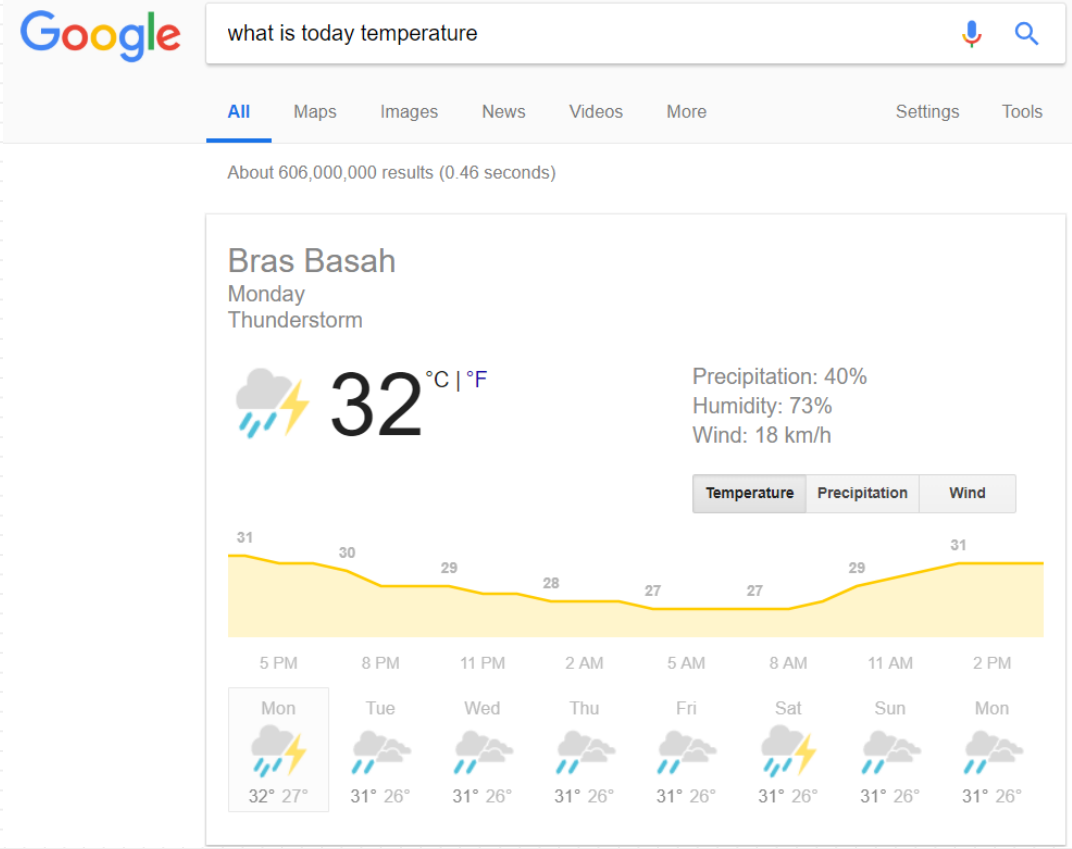
socialmention*

53% strength	1:1 sentiment
46% passion	27% reach
1 minutes avg. per mention	
last mention 5 minutes ago	
54 unique authors	
0 retweets	

Sentiment

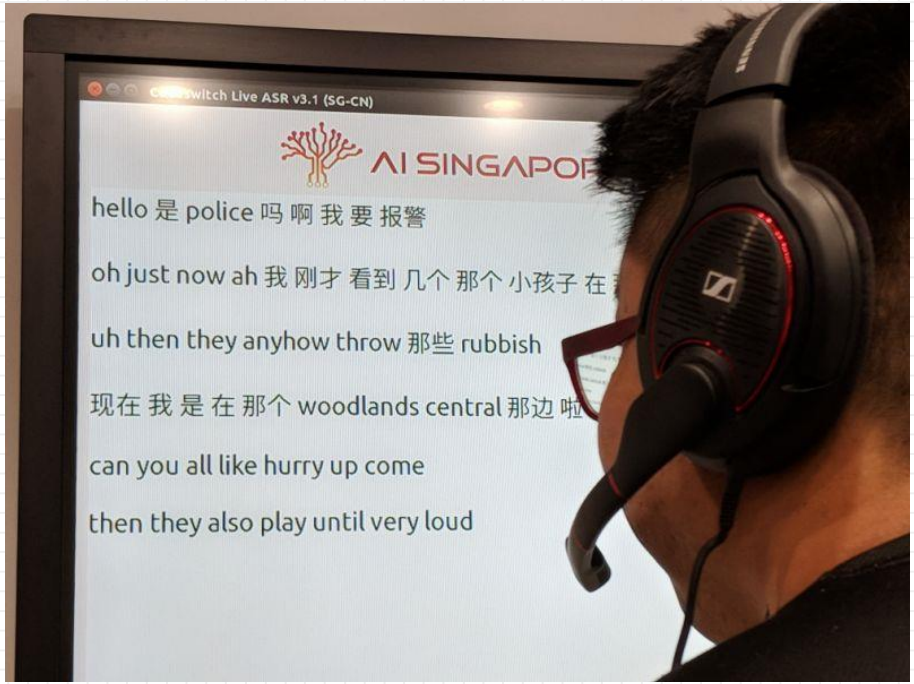
positive	7
neutral	88
negative	10

2. Question Answering



3. Machine Translation

https://sg.news.yahoo.com/speech-recognition-system-can-transcribe-singapore-lingo-real-time-131406725.html



3. Machine Translation

“ **Taiwan wins gold in woman’s 75kg powerlifting in Paralympics**

Beijing, Sept. 14 (CNA) The Chinese Taipei flag was finally raises in the 2008 Paralympic Games in Beijing Sunday, with Lin Tzu-hui winning a gold medal in the women’s 75 kg category of powerlifting event. “

If you were to translate this to Chinese, how would Google Translate fair?

Do you think you can do better or google can do better?

(2012.9.12)

(2014.2.18)

(2015.2.25)

英文 中文 日文 偵測語言

Taiwan wins gold in woman's 75 kg powerlifting in Paralympics
Beijing, Sept. 14 (CNA) The Chinese Taipei flag was finally raised in the 2008 Paralympic Games in Beijing Sunday, with Lin Tzu-hui winning a gold medal in the women's 75 kg category of powerlifting event.

中文(簡體) 英文 中文(繁體) 翻譯

台灣勝金在殘奧會女子75公斤舉重
北京9月14日（CNA）的中華台北國旗終於提出在2008年殘奧會在北京週日與林姿輝贏得舉重賽事的女子75公斤級金牌。

提出修改建議

(2016.2.24)

Taiwan wins gold in woman's 75 kg powerlifting in Paralympics
Beijing, Sept. 14 (CAN) The Chinese Taipei flag was finally raised in the 2008 Paralympic Games in Beijing Sunday, with Lin Tzu-hui winning a gold medal in the women's 75 kg category of powerlifting event.

台湾在女子75公斤举重在残奥会赢得金子
北京9月14日（北京时间）中国台北旗帜在北京星期天的残奥会上终于被提起，林子辉在女子75公斤举重赛事中获得金牌。

提出修改建議

(2016.12.8)

Taiwan wins gold in woman's 75kg powerlifting in Paralympics
Beijing, Sept. 14 (CNA) The Chinese Taipei flag was finally raises in the 2008 Paralympic Games in Beijing Sunday, with Lin Tzu-hui winning a gold medal in the women's 75 kg category of powerlifting event.

台灣勝金在殘奧會女子75公斤級舉重
北京9月14日（CNA）的中華台北國旗終於引發了2008年殘奧會在北京週日與林姿輝贏得舉重賽事的女子75公斤級金牌。

(2017.2.22)

Taiwan wins gold in woman's 75kg powerlifting in Paralympics

Beijing, Sept. 14 (CNA) The Chinese Taipei flag was finally raises in the 2008 Paralympic Games in Beijing Sunday, with Lin Tzu-hui winning a gold medal in the women's 75 kg category of powerlifting event.

台灣在殘奧會女子舉重75公斤級比賽中獲得金牌

（北京時間9月14日）中國台北國旗在周日的北京2008年殘奧會上終於升起，林子輝在女子75公斤級舉重比賽中獲得金牌。

提出修改建議

The Challenges – Lexical Transfer

- Lexical Transfer
  - raise (升起) → 提出 (2008) →抬起 (2012) →抬起 (2014) → 提出 (2015) → 提出 (2016) → 引發 (2017) → 升起 (2018)
  - woman (女子)→婦女(2008)→女子(2012)→女子(2014)→ 女子 (2015) → 女子 (2016)→女子 (2017)→女子 (2018)
  - event (項目) → 事件 (2008) → 事件 (2012) → 事件 (2014) → 賽事 (2015) → 賽事 (2016)→賽事 (2017)→比賽 (2018)
  - Gold (金牌) → 金/金牌 (2008) →黃金/金牌(2012) →金/金牌 (2014) → 金牌/金牌 (2015) → 金/金牌(2016)→ 金牌/金牌(2018)
  - powerlifting (舉重) → powerlifting (2008) → 舉重 (2012) →舉重 (2014) → 舉重 (2015)→舉重(2016)→舉重(2017)→舉重 (2018)
  - Lin Tzu-hui (林資惠) →林姿慧 (2008) →林慈清輝 (2012) →林梓輝 (2014) →林姿輝 (2015)→林姿輝 (2016)→林姿輝 (2017) →林子輝(2018)

suggest

bring up

raise

initiate

Incident

Competition

The Challenges – Structural Transfer

- Taiwan wins gold in woman's 75 kg powerlifting in Paralympics
- 臺灣在殘奧會贏得女子75公斤級舉重項目金牌
- 台灣勝金在婦女的75公斤powerlifting在殘奧會 (2008)
- 台灣勝在殘奧會舉重女子75公斤黃金 (2012)
- 台灣勝金在女子75公斤級舉重殘奧會 (2014)
- 台灣贏得金牌的女子75公斤舉重在殘奧會 (2015)
- 台灣勝金在殘奧會女子75公斤舉重 (2016)
- 台灣勝金在殘奧會女子75公斤級舉重 (2017)
- 台灣在殘奧會女子舉重75公斤級比賽中獲得金牌 (2018)

Taiwan In the Paralympics women's 75kg Powerlifting Won Gold

Try it Yourself !

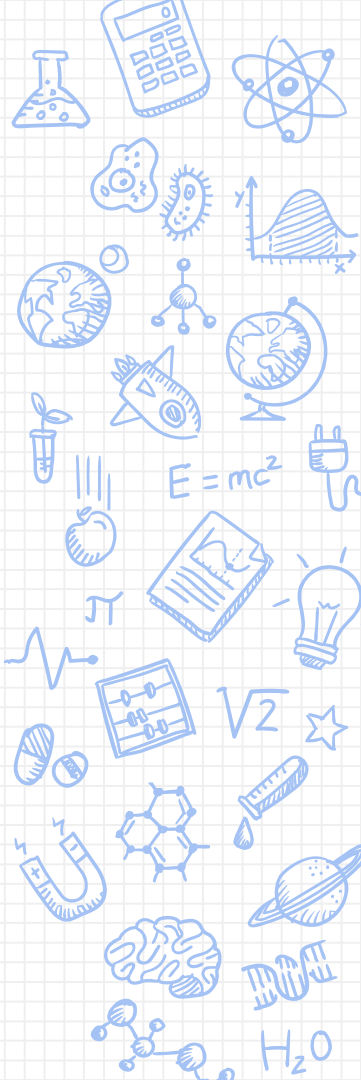
What do you think this sentence means?

“明年北市学童持记名悠游卡搭北捷 票价75折优惠”

My own interpretation:

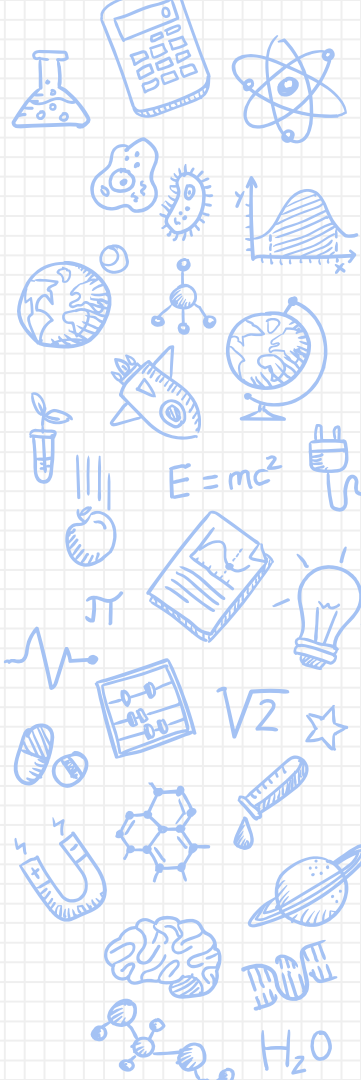
Students from Taipei who posses an EasyCard will enjoy a 25% discount for taking taipei metro next year.

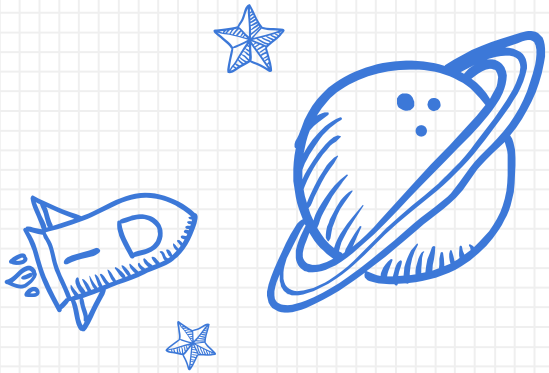
http://news.ltn.com.tw/news/life/breakingnews/2177875



Other Applications of NLP

- Text Analytics
- Question Answering
- Machine Translation
- Text Processing (Understanding/Generation)
- Written Aids (Spelling Checker, Grammar Checker, Style Checker)
- Speech Recognition/Synthesis
- Summarization
- Robo-Adivsor(FinTech), MedWhat, Watson Health, ROSS (Law) ..



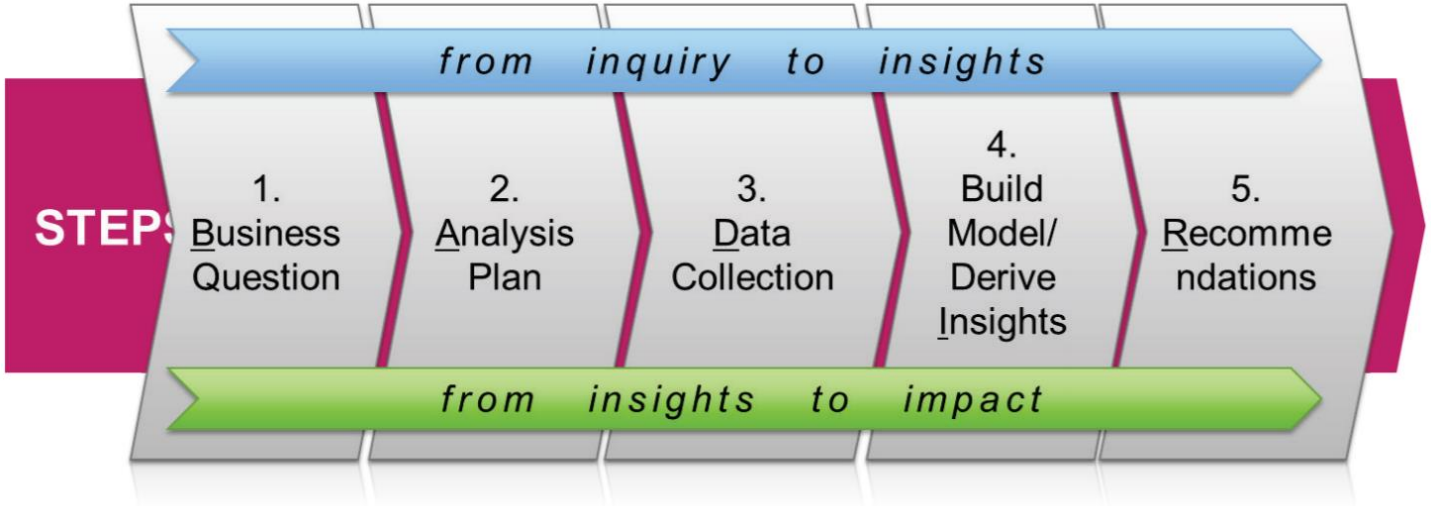


Pipelines

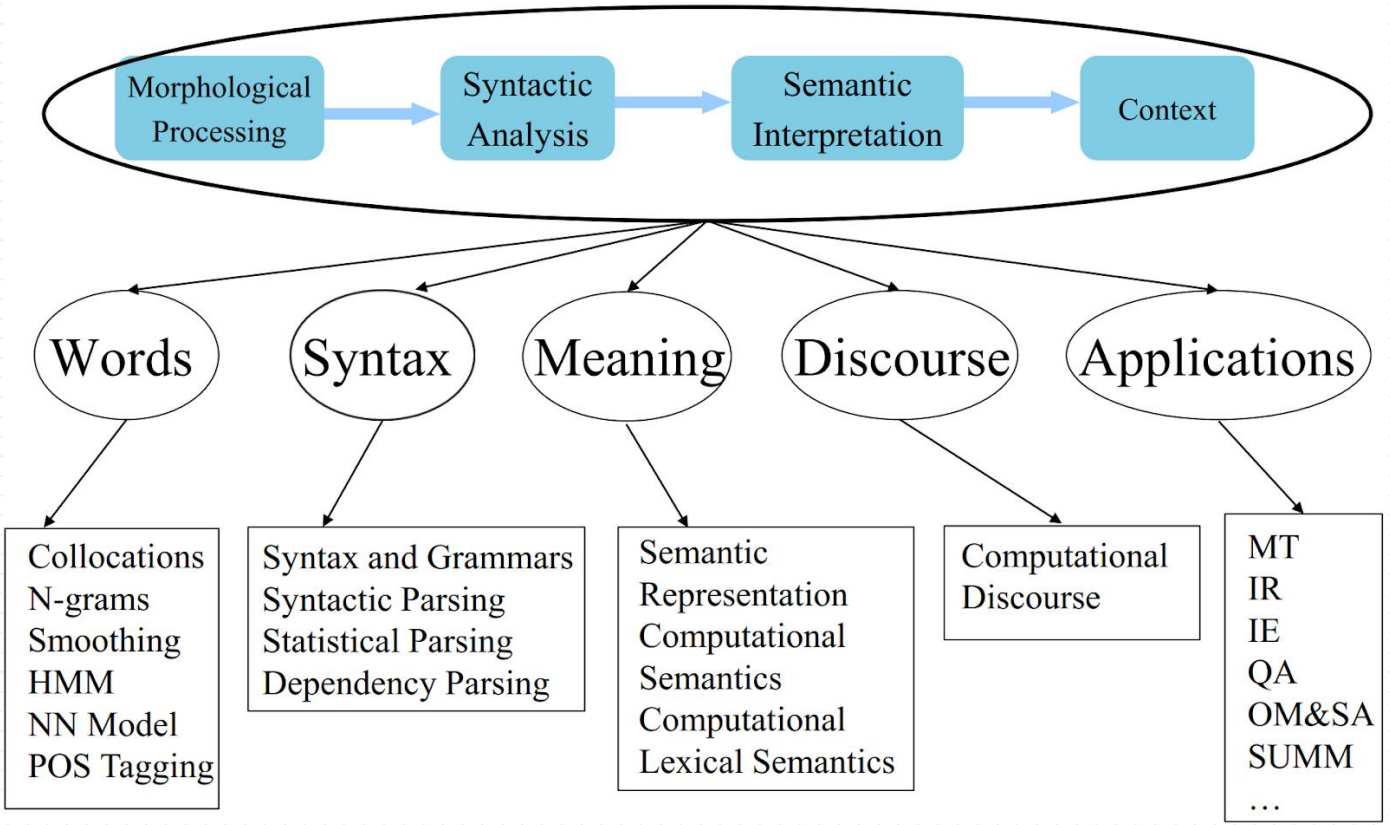
Analytics Pipeline

Effective Predictive Analytics Framework

BADIR™: 5 steps from "data to decisions"™



NLP Pipeline



Morphology & Knowledge of Words

Analysis of Written Language

- morphology ---------- Morphological analyzer

  structure of words

- syntax ---------- Parser

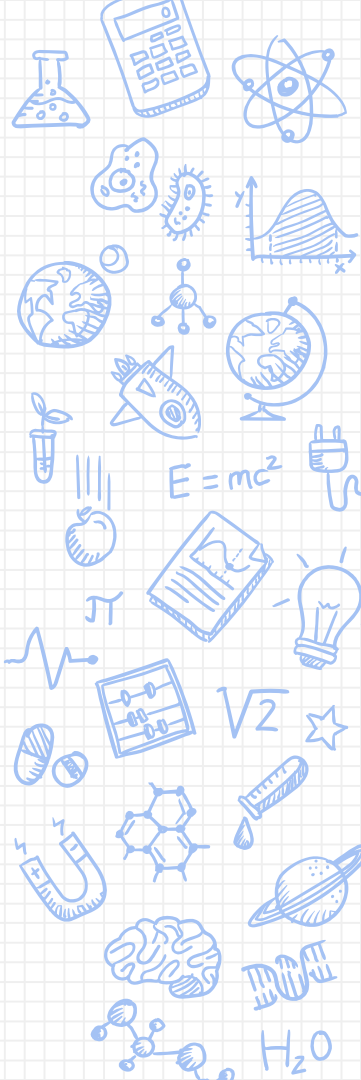
  structure of sentences

- semantics ---------- Semantic interpreter

  meaning of individual sentences

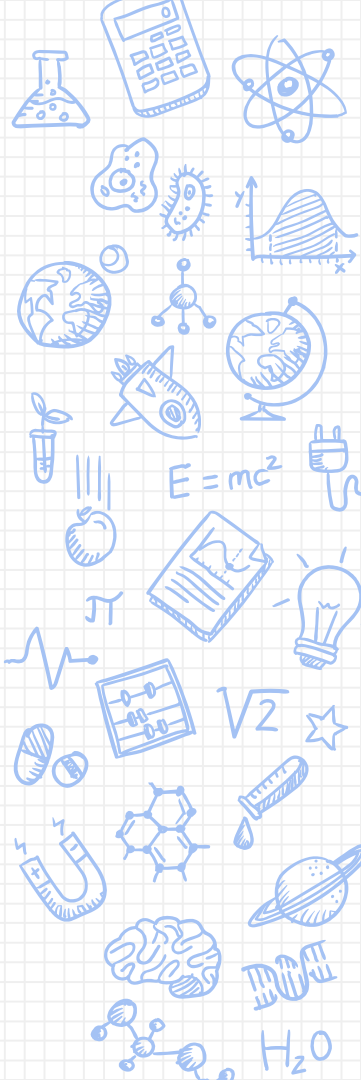
- pragmatics

  how sentences relate to each other



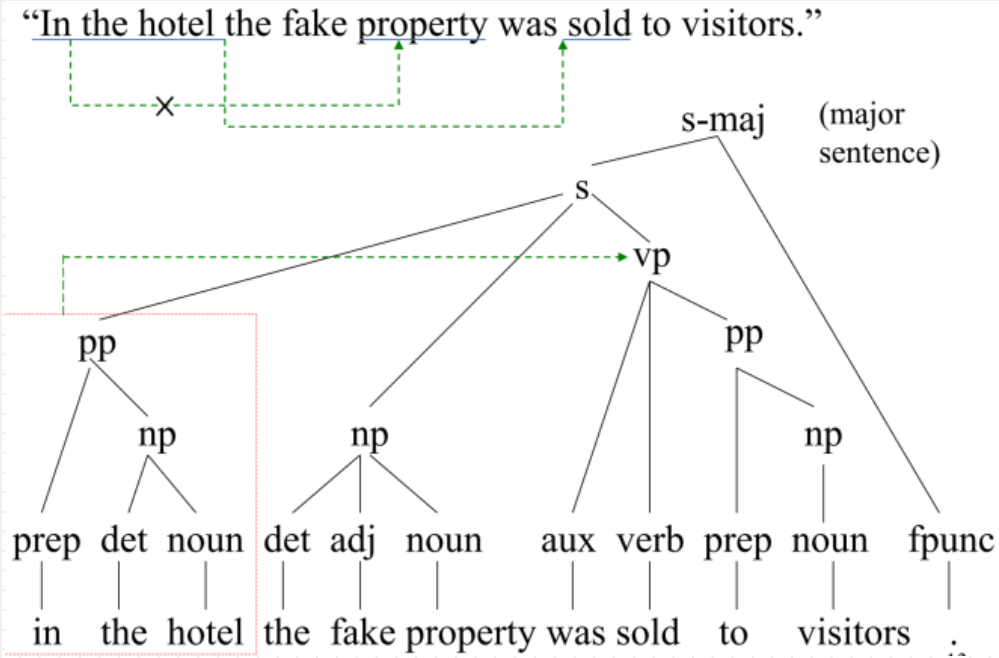
Morphological Analyzer

- word (lexical item)
- dictionary (lexicon)
- morphological analyzer
  - apply morphological rule to finding the roots of words, e.g., going → go, cats → cat
- POS Tagging
  - I (Pronoun) Love (Verb) Cats (Noun)



Syntactic Functions

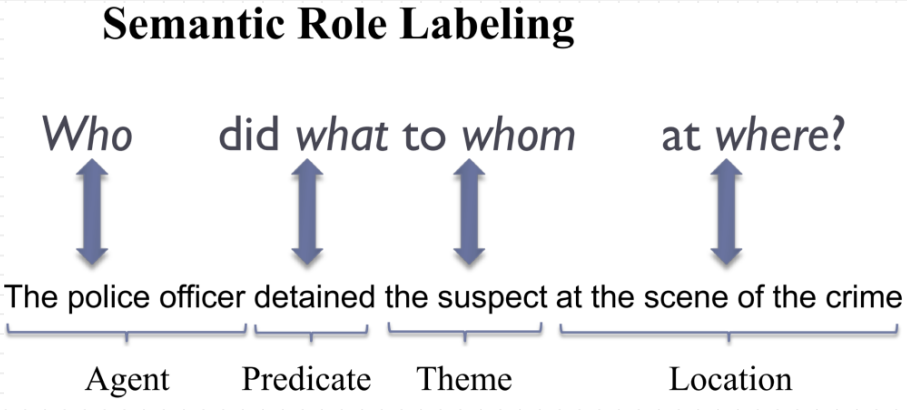
- Syntax
- Subject Verb Agreement
- Context-Free Grammar



Semantic Analysis

- **Agent** : a person or thing who is the doer of an event.
  – *The boy* ran down the street.
- **Patient** : the surface object of the verb in a sentence.
  – He opened *the door.*
- **Instrument** : an inanimate thing that an agent uses to implement an event.
  – The cook cut the cake with *a knife.*
- **Goal** : thing toward which an action is directed
  – He threw the book at *me.*
- **Beneficiary** : a referent which is advantaged or disadvantaged by an event.
  – John sold the car for *a friend*

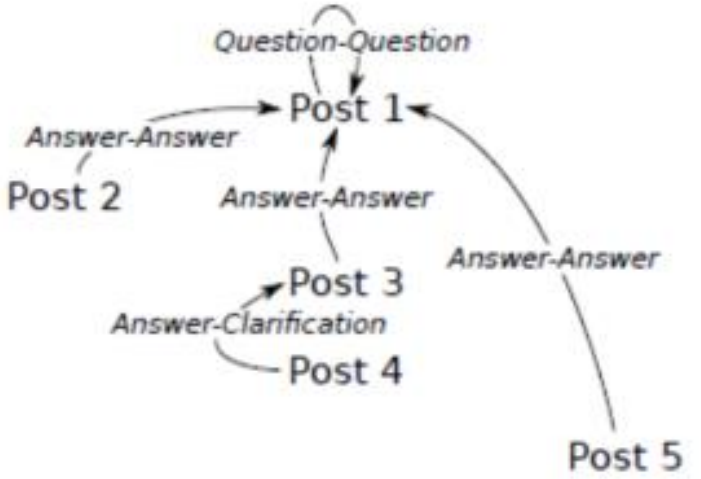
Semantic Analysis



- PropsBank
- FrameNet

Pragmatics Analysis

- Pragmatics is the area of studies that goes beyond the study of the meaning of a sentence and tries to explain what the speaker really is expressing
- The resolution of anaphoric relations is crucial to the task of information extraction



Discourse Analysis

- **Motivation: Information Extraction**
  - Information Extraction
  - Summarization
  - Conversation Agent
- **Result**: one event can cause a following event
  - John hid Bill's car keys. He was very upset about it."
- **Explanation**: a previous event causing one event.
  - John hid Bill's car keys. He was drunk."
- **Parallel**: both events happening at the same time
  - John hid Bill's car keys. Bill was sleeping."
- **Elaboration**: Detailed elaboration of an event.
  - John hid Bill's car keys. He put it in his bag."
- **Occasion**: change of state:
  - John hid Bill's car keys. He found it ten minutes later."

Discourse Analysis

(**S1**) Bill was drunk.

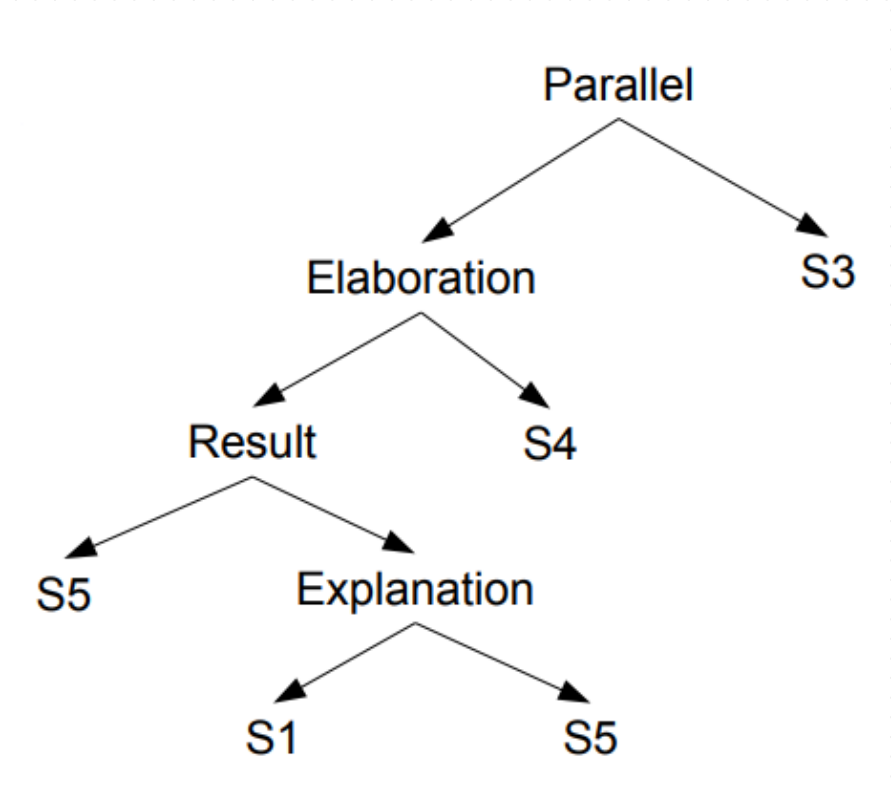
(**S2**) John hid Bill's car keys.

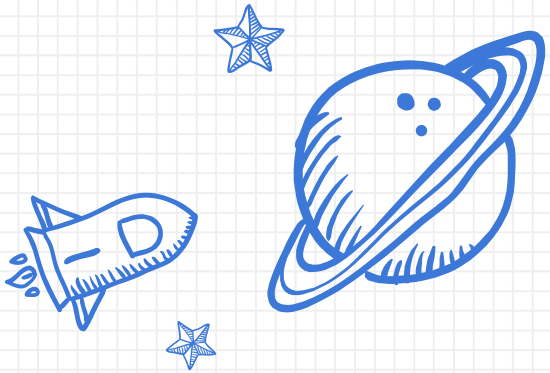
(**S3**) (While) Bill was sleeping.

(**S4**) He put it in his bag.

(**S5**) Bill was very upset about it.

(**S6**) Bill found it ten minutes later.

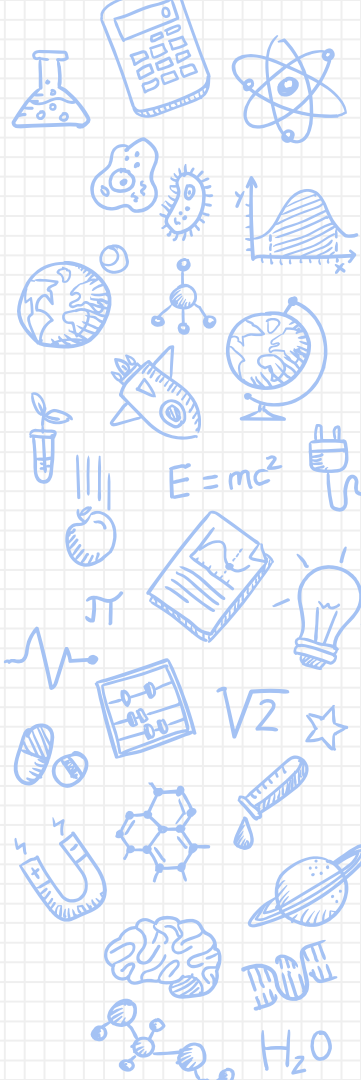




Why is NLP Difficult?

Challenges in NLP

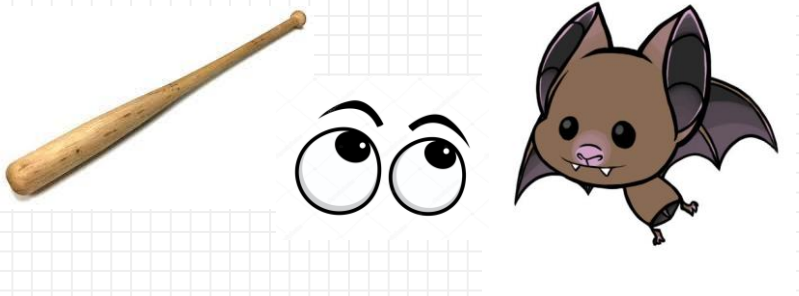
- Ambiguity
  - Lexical
  - Syntactic
  - Semantic
  - Pragmatic
- Irony
- Negations
- Unstructured Data / Abbreviations

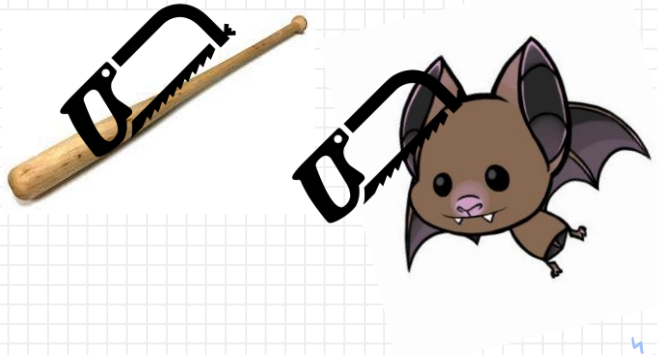


Lexical Ambiguity

I saw bats!

- A single term (Lexical Token) have multiple forms or meaning (Verb vs Noun)





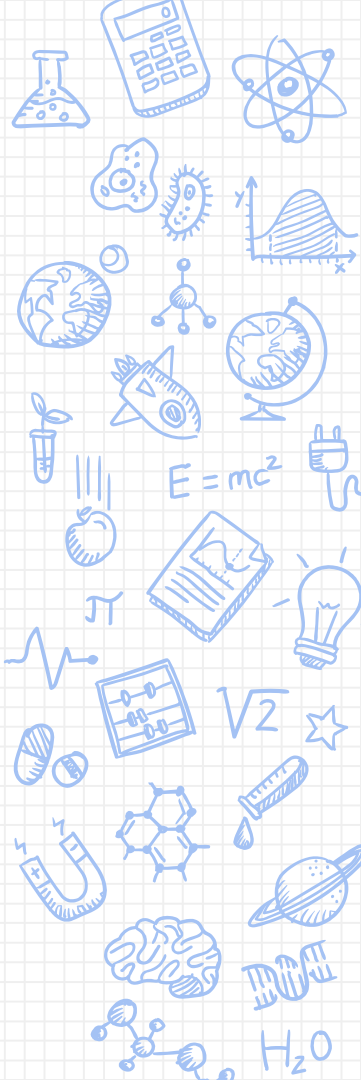
Bank : River Bank vs Money Bank

Current : (Noun or Adjective)

Syntactic Ambiguity

Sammy ate the cookies *on the floor*

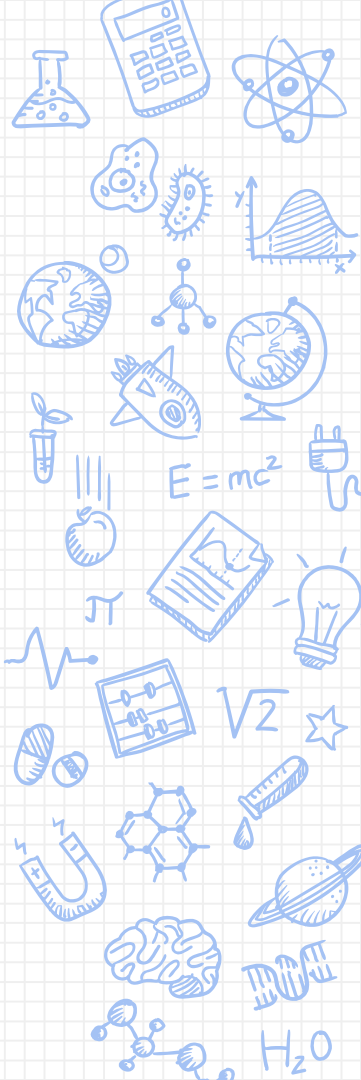
- (Did Sammy pick up the cookies on the floor and eat them or did he sit on the floor while eating the cookies?)
- Prepositional phrase attachment



Semantic Ambiguity

- When meaning of the words themselves can be misinterpreted. Even after the Syntax and Meaning of the individual words have been resolved.

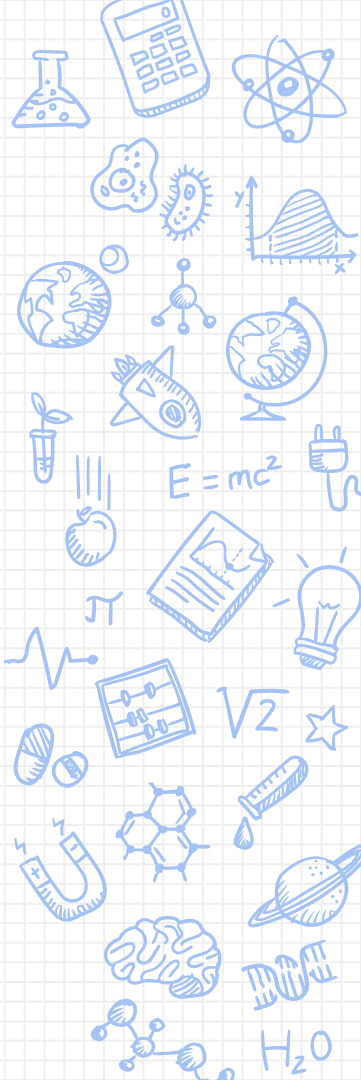
- The car hit the pole while it was moving.
  - One way of understanding is that the car (while moving) hit the pole.
  - Another is that the car hit the pole as the pole moves.



Pragmatics Ambiguity

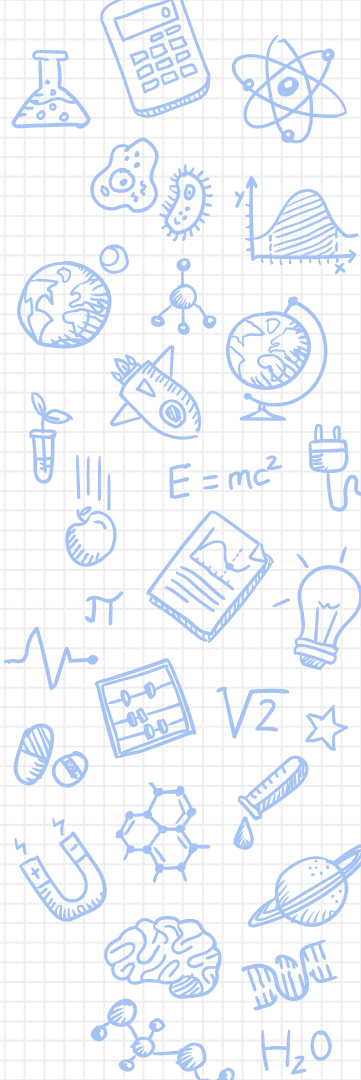
- One of the hardest Task in NLP. Occurs when statement is not specific, and information are missing and required to be inferred.

- **Son (Running Late) : Help me go to the room to see if my wallet is there. Faster, I am running late.**
- **Mom (Went to the room and came out) : Yeah they are on your table.**
- This clearly shows that the mother is falling short of the expectation of the son, as she does not understand the pragmatics of the situation. The missing information is that the son need the calculator and expect the mom to take it for him.

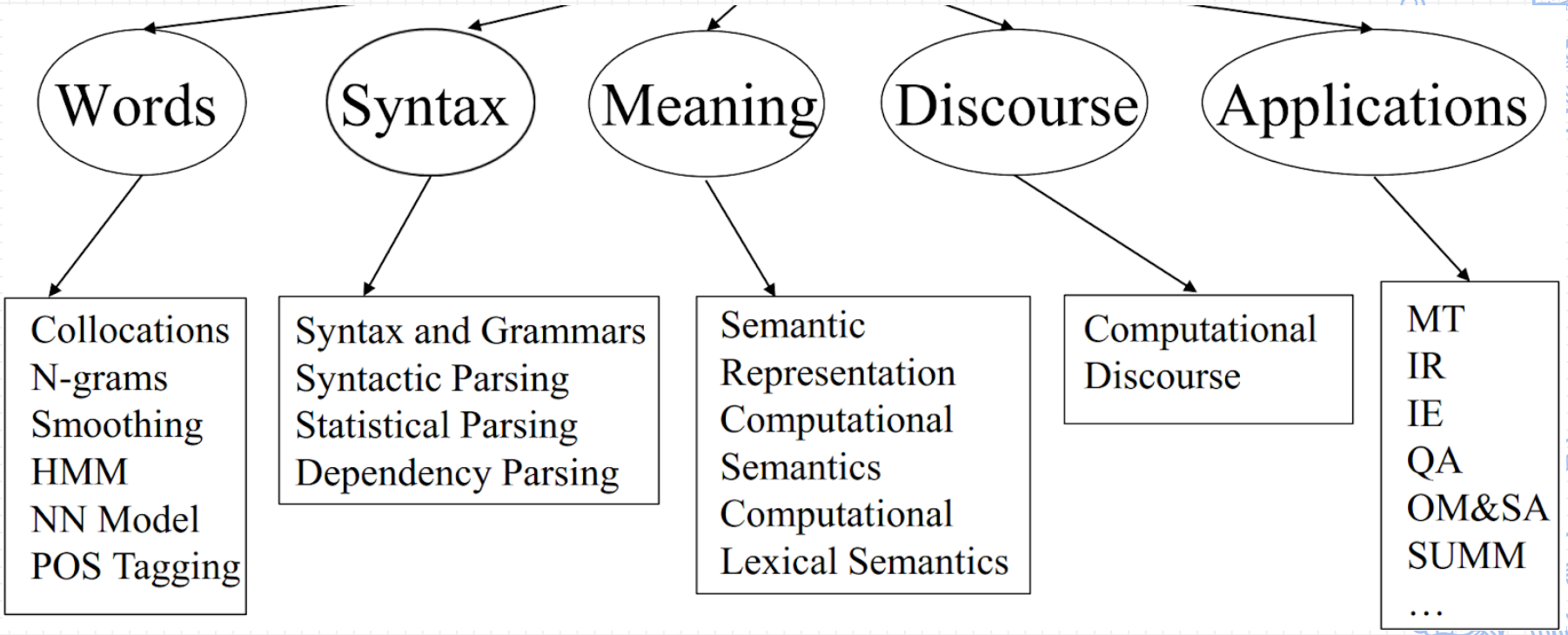


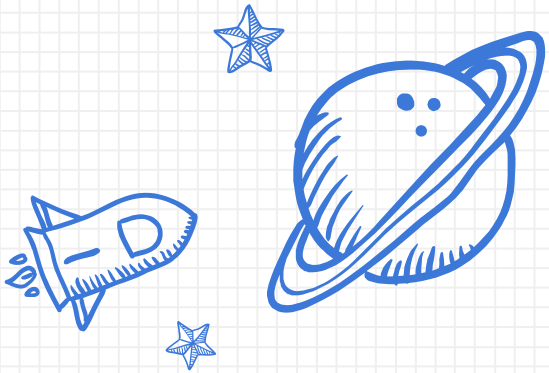
Challenges in NLP

- Irony
  - Iphone XS have the BEST PRICE!!!! AWESOME!
- Negation
  - I don't really not like the idea of having no Home Button.
- Un-structured Data/Abbreviations
  - 666, 87, 4896
  - Lol, ggwp, gr8, 5Q
- Noises
  - ^.^ emoticons, lots of punctuations
- Slang
  - Jialat, shiok, sibei

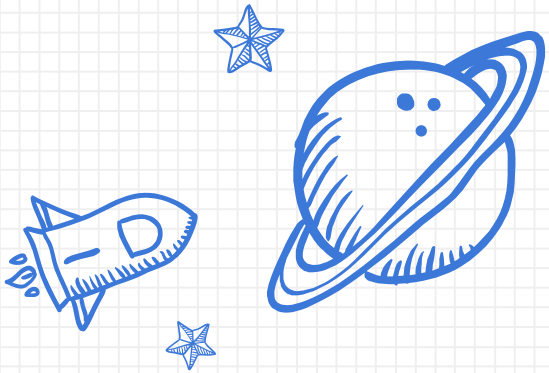


NLP Pipeline





Hands-on Exe 1



Key Concepts

Regex

- Aka. Regular Expression
- Strings with a special Syntax
- Allow us to match pattern with other strings

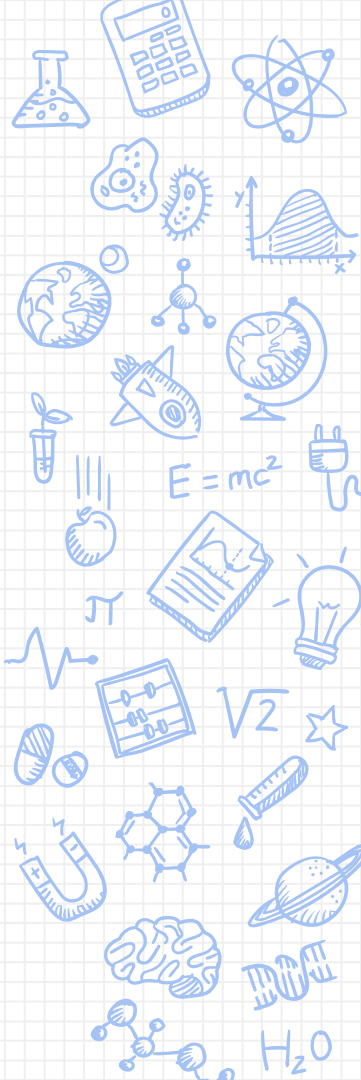
pattern	matches	example
\w+	Word	“Crazy” “Rich”
\d	Digit	9
\s	Space	“ “
\S	Not space	‘no_space”
[A-Z] or [a-z]	grouping	‘ABC’ ‘abcd’

https://www.rexegg.com/regex-quickstart.html

Corpus

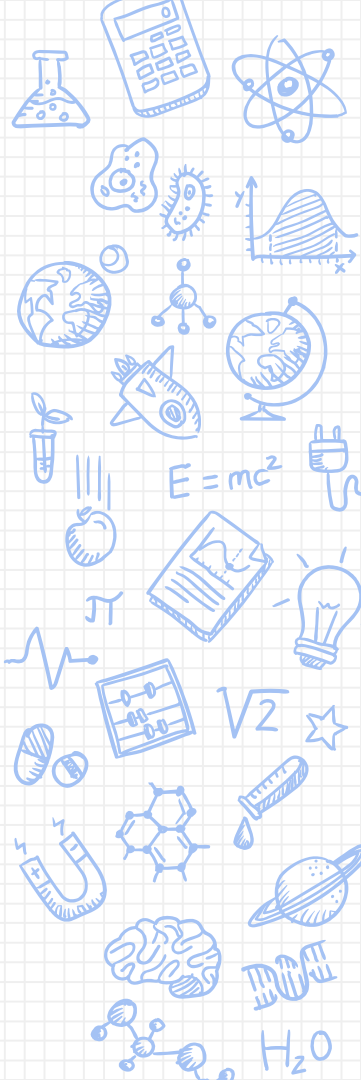
- A large set and structured set of text

- Stopwords
- Brown Corpus
- Sentiment Corpus - http://sentiwordnet.isti.cnr.it/
- Finance Corpus - https://sraf.nd.edu/data/
- Emoticon Corpus - http://saifmohammad.com/WebPages/lexicons.html



Stop words

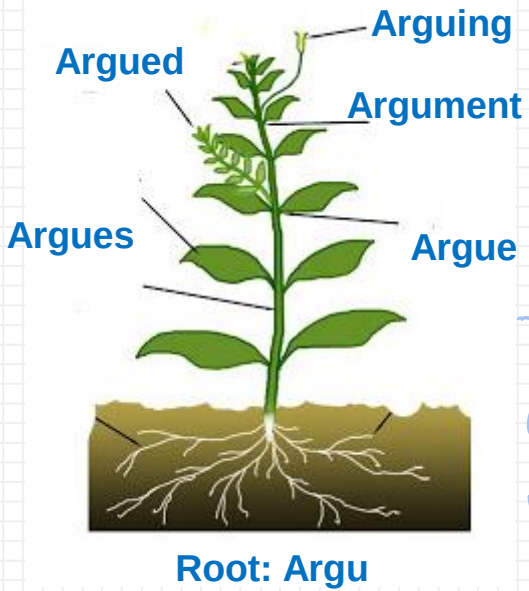
- Stop words are words which are filtered out before or after processing of natural language data (text)
- refers to the most common words in a language
- https://www.ranks.nl/stopwords
- You define your own stop word list



Stemming

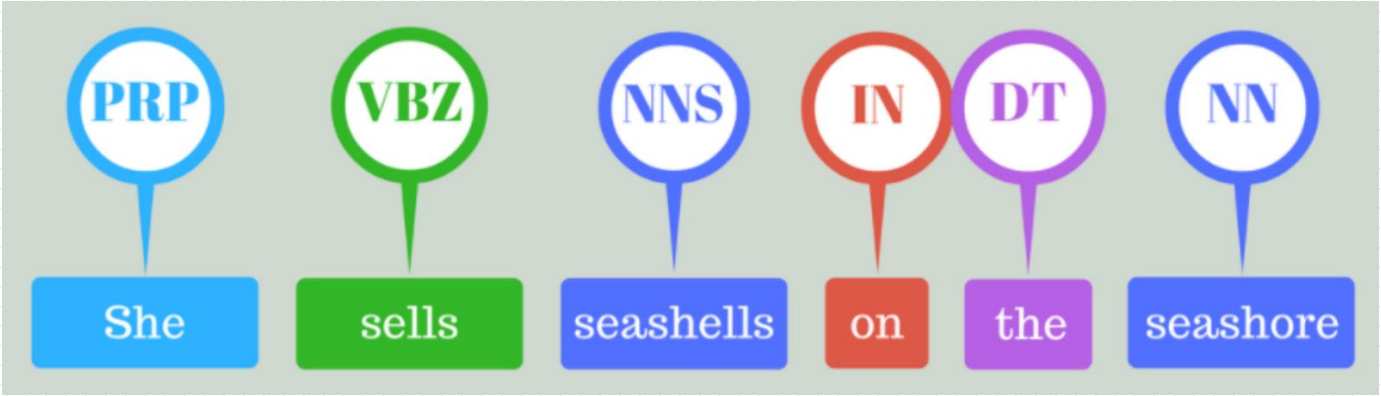
- http://www.nltk.org/howto/stem.html

- Argu – argue, arguing, argued
- Usually Noun and Verb
- Snowball can ignore stop words stemming



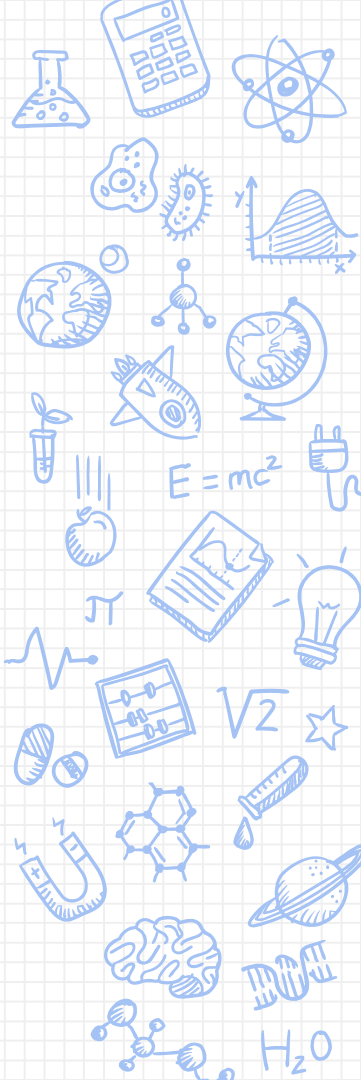
POS Tagging

- 8 (ish) traditional parts of speech
- Noun, verb, adjective, preposition, adverb, article, interjection, pronoun, conjunction



POS Tagging

N	noun	chair, bandwidth, pacing
V	verb	study, debate, munch
ADJ	adjective	purple, tall, ridiculous
ADV	adverb	unfortunately, slowly
P	preposition	of, by, to
PRO	pronoun	I, me, mine
DET	determiner	the, a, that, those



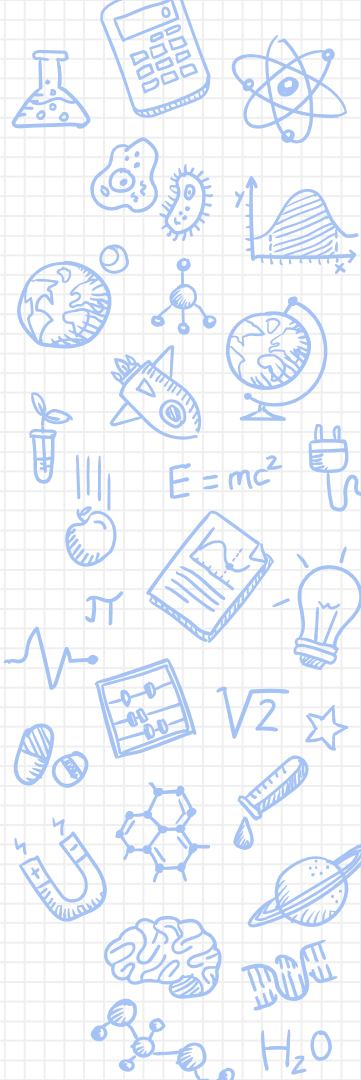
Penn TreeBank POS Tagset

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two, three*	TO	"to"	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential 'there'	*there*	VB	verb, base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb, past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb, gerund	*eating*
JJ	adjective	*yellow*	VBN	verb, past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb, non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb, 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, singular	*IBM*	$	dollar sign	$
NNPS	proper noun, plural	*Carolinas*	#	pound sign	#
PDT	predeterminer	*all, both*	"	left quote	' or "
POS	possessive ending	*'s*	"	right quote	' or "
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one's*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

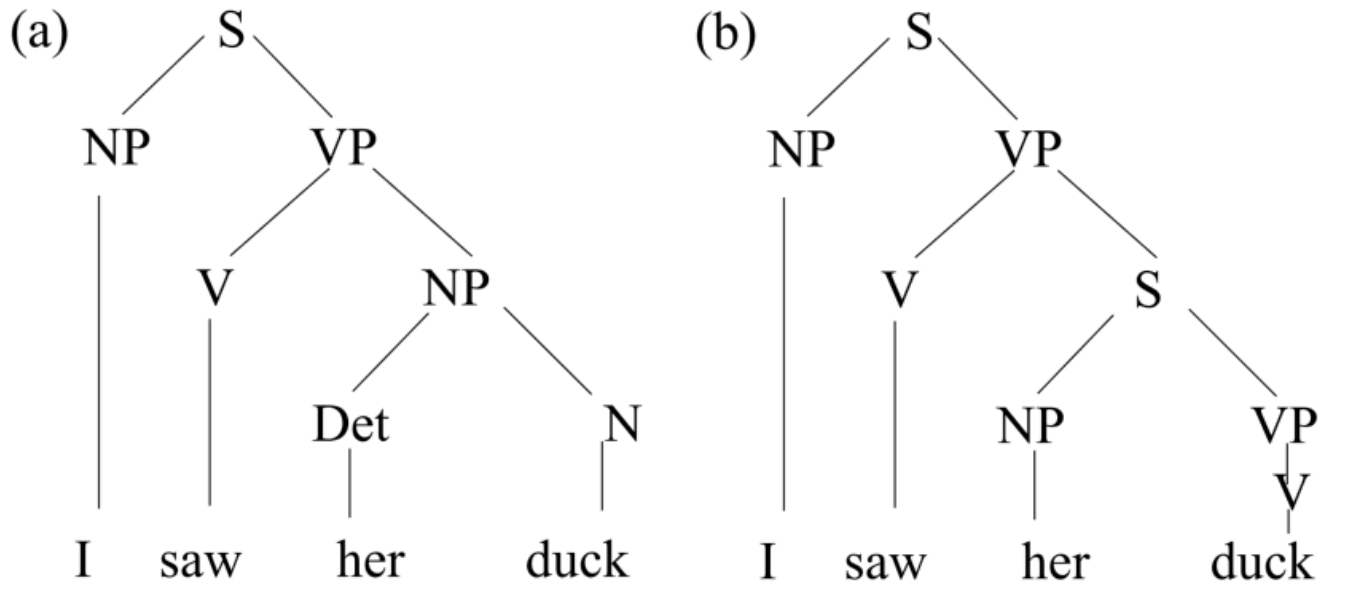
https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

Penn TreeBank POS Tagset

- Words often have more than one POS: back
- The back door = JJ
- On my back = NN
- Win the voters back = RB
- Promised to back the bill = VB
- The POS tagging problem is to determine the POS tag for a particular instance of a word.



Syntax and Grammar



top-down and bottom-up approaches.
CYK algorithm
Earley parser

major sentence

Part (a)

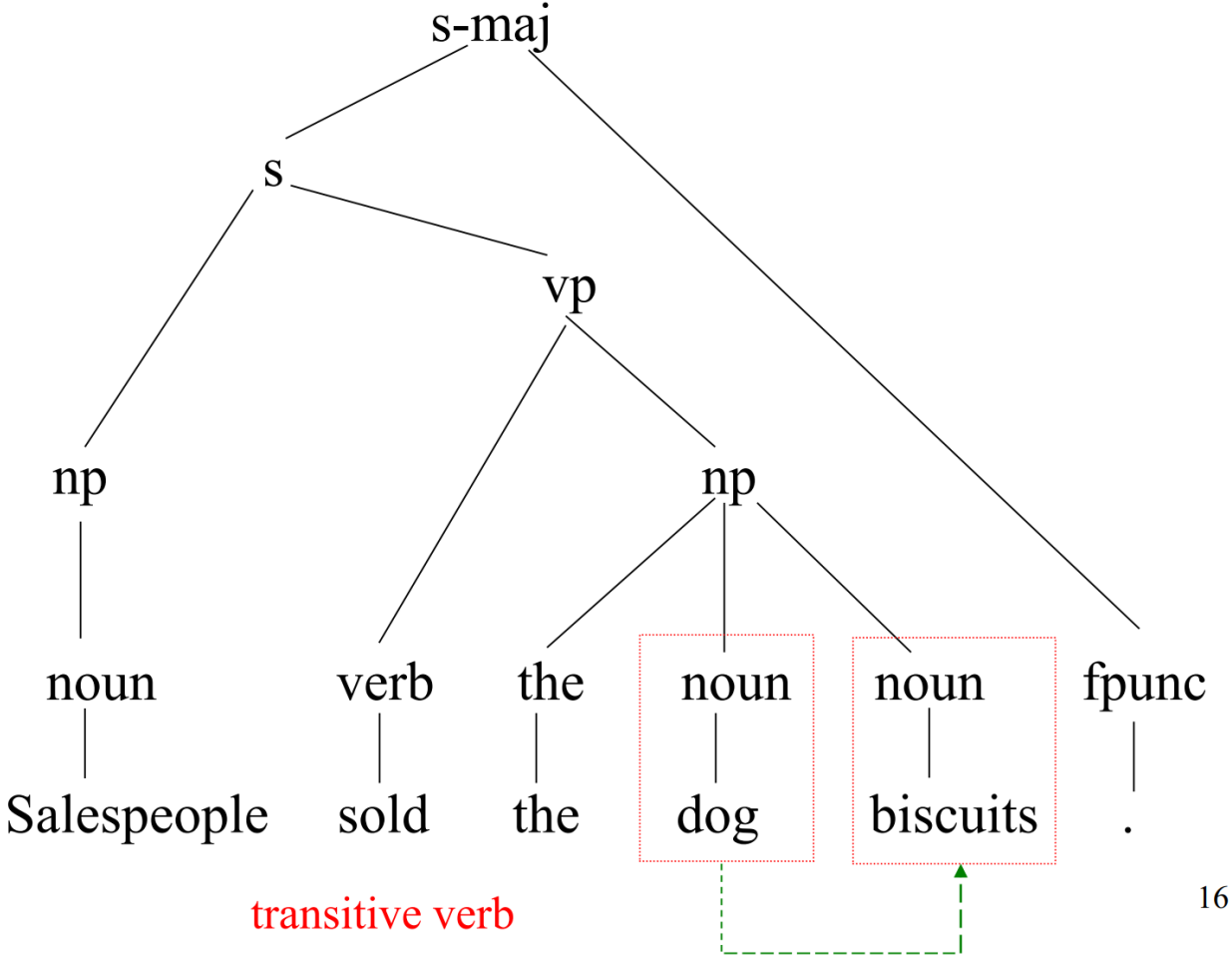
s-maj	→	s fpunc	det	→	the
s	→	np vp	noun	→	dog
vp	→	verb	verb	→	ate
np	→	det noun	fpunc	→	.

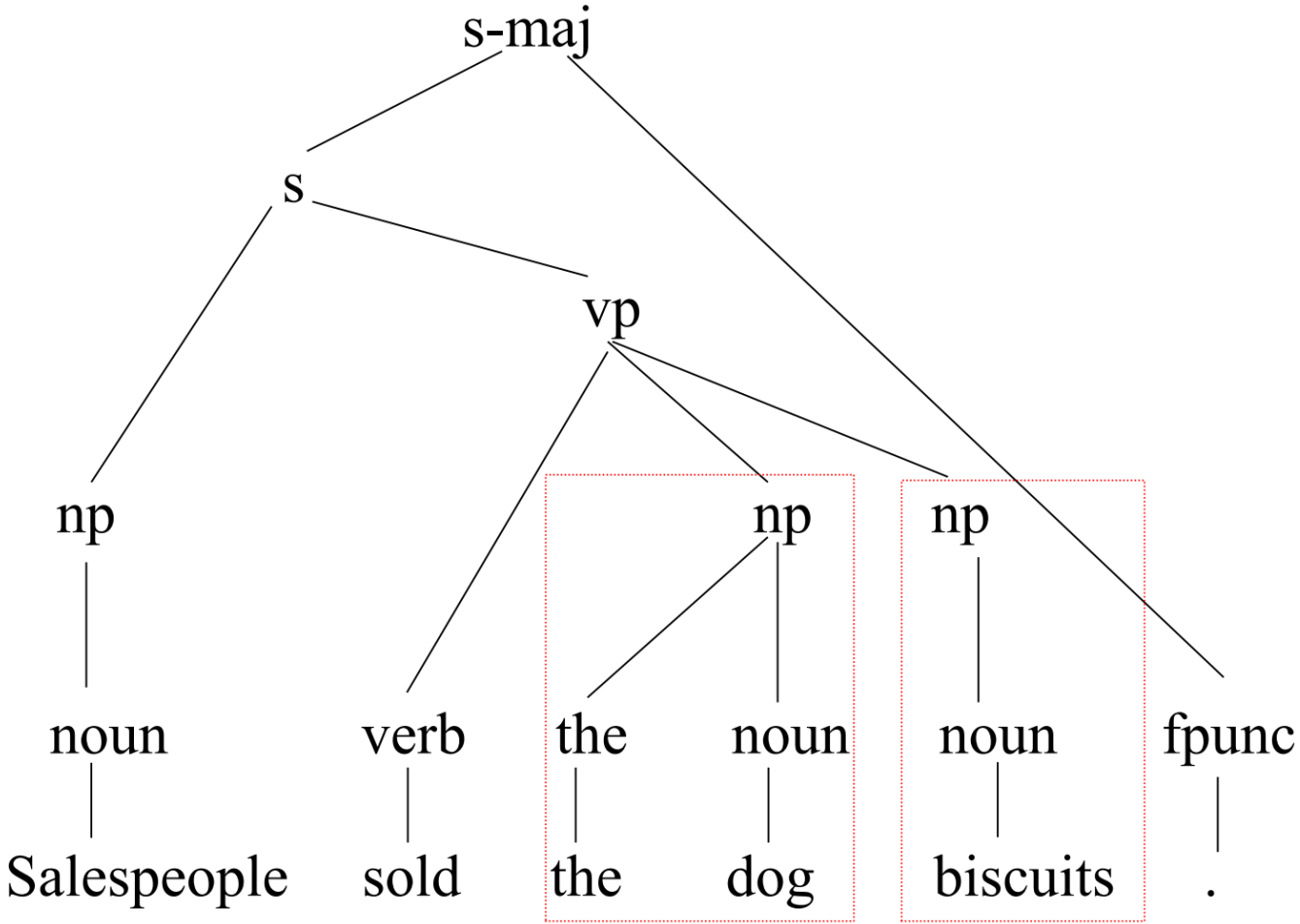
ambiguous grammar

Part (b)

vp	→	verb np	noun	→	salespeople
vp	→	verb np np	verb	→	sold
np	→	det noun noun	noun	→	biscuits
np	→	noun			

“Salespeople sold the dog biscuits.”

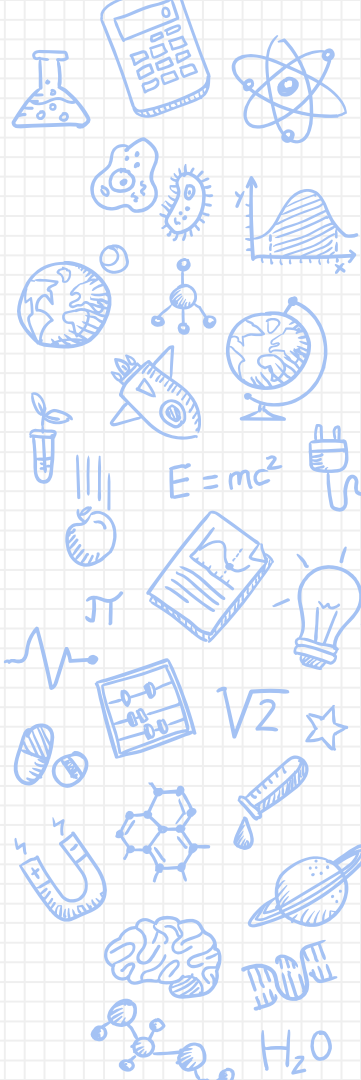


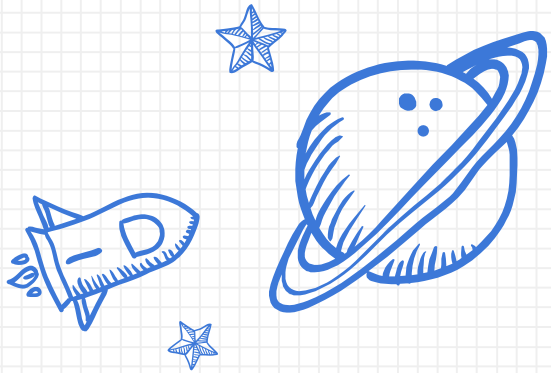


ditransitive verb

Lemmatization

- Lemmatization is similar to stemming but it brings context to the words. So it goes a steps further by linking words with similar meaning to one word.
- It takes consideration of the POS, else default POS Type is NN.

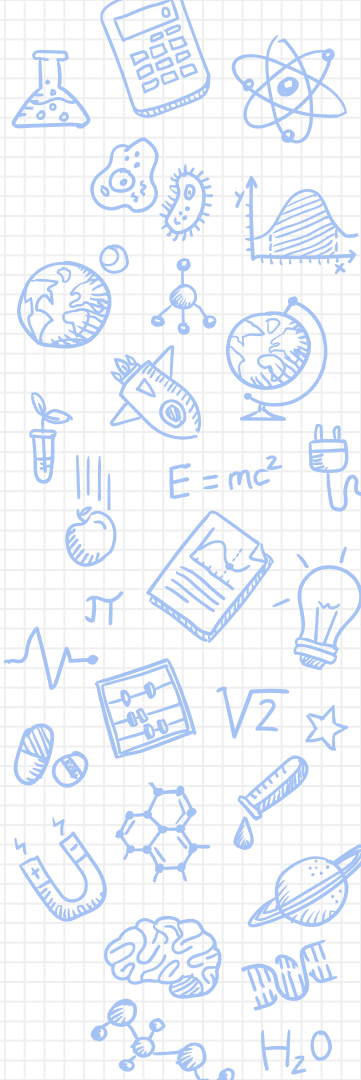




Hands-on Exe 2

Data Set

- Shakespeare (A text file I grab randomly from the internet)
- SemEval-2017 Task 5 - Fine-Grained Sentiment Analysis on Financial Microblogs and News
  - http://alt.qcri.org/semeval2017/task5/

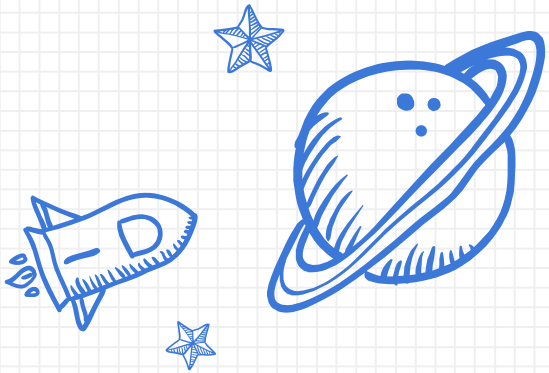


TASK – SemEval 2017 Task 5

Example:
Tweet: $TSLA if $249.84 breaks we see $245 then $2
Cashtag: $TSLA

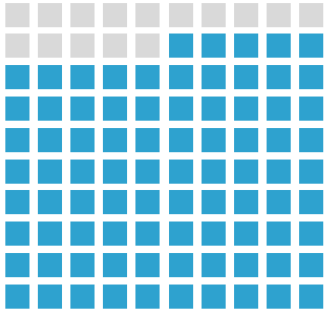
Predict Sentiment Score = -0.519

- Financial Microblog
  - ➔ StockTwits
- Cashtag
  - ➔ $AAPL = Apple Inc.
- Sentiment Analysis
  - ➔Bullish/Bearish
- Fine-Grained Sentiment Analysis
  - ➔ Score between -1 to 1



Hands-on Exe 3

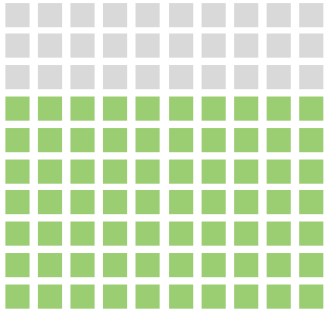
Evaluation Method



F1

3-Classes

Classify into
Bullish/Bearish/Neutral
Calculate
Micro-average F1 &
Macro-average F1



MSE

Fine-Grained Score

Calculate
Mean Squared Error

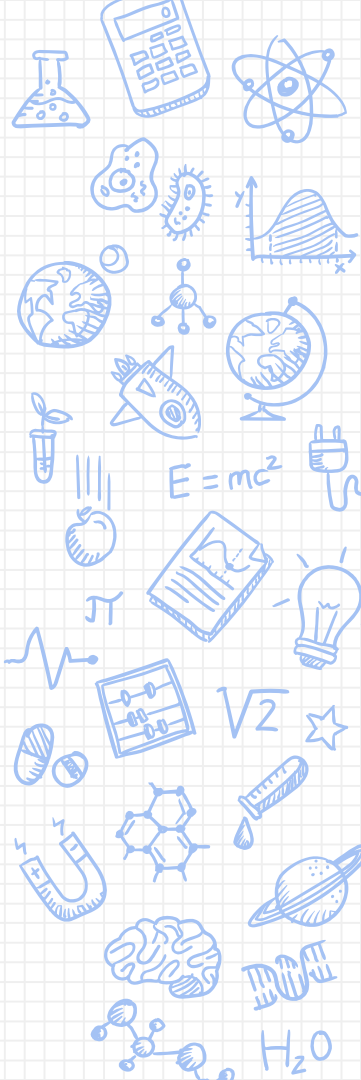
https://en.wikipedia.org/wiki/Precision_and_recall

Corpus:

http://www.nlg.csie.ntu.edu.tw/nlpresource/NTUSD-Fin/

Bag of Words

- Basic method to find topics in a text
- Need to first create tokens using tokenization
- Then count the tokens
- The more frequent the word appear the more important it might be
- Good way to determine words that are significant



Bag of Words Example

“ The cat is in the box”

“ The cat likes the box”

“ The box is over the cat”

Features	Count
The	6
Box	3
Cat	3
Is	2
In	1
Like	1
Over	1

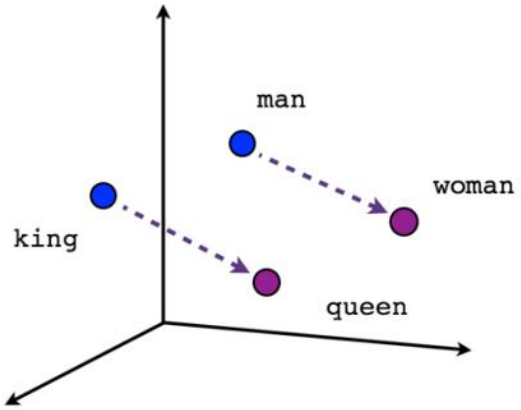
https://towardsdatascience.com/introduction-to-word-embeddings-4cf857b12edc

Count Vectorizer Example

```
>>> from pandas import DataFrame
>>> from sklearn.feature_extraction.text import CountVectorizer
>>> docs = ["You can catch more flies with honey than you can with vinegar.",
...         "You can lead a horse to water, but you can't make him drink."]
>>> vect = CountVectorizer(min_df=0., max_df=1.0)
>>> X = vect.fit_transform(docs)
>>> print(DataFrame(X.A, columns=vect.get_feature_names()).to_string())
   but  can  catch  drink  flies  him  honey  horse  lead  make  more  than  to  vinegar  water  with  you
0    0    2      1      0      1    0      1      0     0     0     1     1   0        1      0     2    2
1    1    2      0      1      0    1      0      1     1     1     0     0   1        0      1     0    2
```

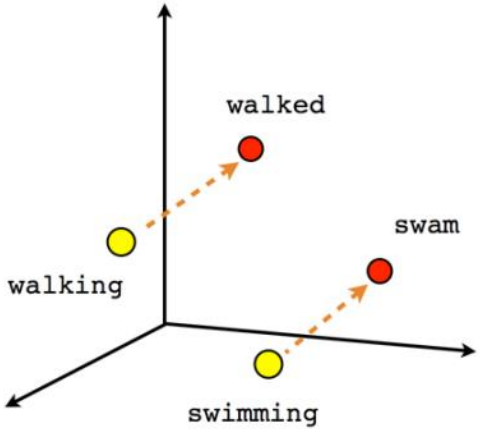
Google Word2Vec

https://code.google.com/archive/p/word2vec/



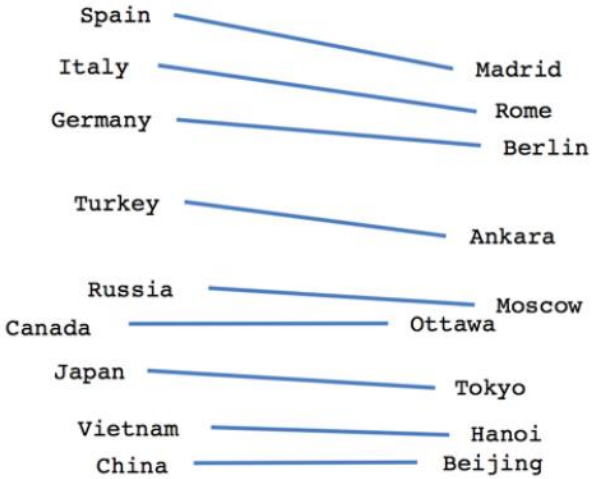
King – Male + Female = ?

Male-Female

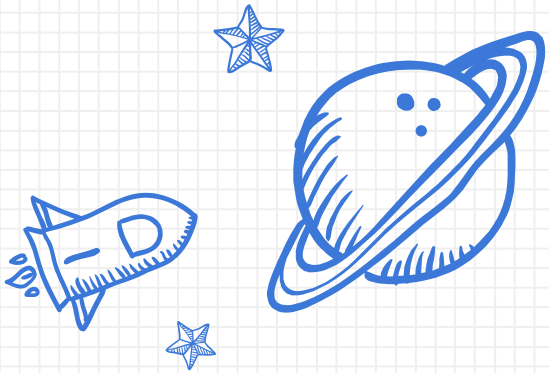


Walking +

Verb tense



Country-Capital



Resources

Resource and Tools

- Text Analytics
  - Sem-Eval - http://alt.qcri.org/semeval2018/index.php?id=tasks
- Question Answering
  - CLEF Question Answering Track
    - QALD: Question Answering over Linked Data
    - BioASQ: Biomedical semantic indexing
  - Allen Institute
    - https://www.kaggle.com/c/the-allen-ai-science-challenge

Conference / Research Papers

- The Association for Computational Linguistics site
  - http://www.aclweb.org
- The ACL Anthology (A Digital Archive of Research Papers in Computational Linguistics)
  - http://www.aclweb.org/anthology-new/
- Language Technology World http://www.lt-world.org/
- Language Resource Listings on the Web http://nlp.stanford.edu/links/statnlp.html
- The Linguistic Data Consortium http://www.ldc.upenn.edu/

Resource and Tools

- WordNet: http://wordnet.princeton.edu/
- Translation:
  - http://babelfish.yahoo.com/,
  - http://translate.google.com/
- Question Answering: http://www.answerbus.com/
- Summarization: http://newsblaster.cs.columbia.edu/
- Online concordancing:
  - http://corpus.leeds.ac.uk/internet.html

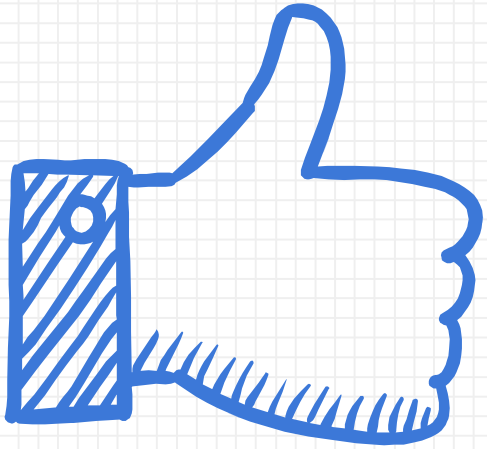
#Toolkits

- NLTK (http://www.nltk.org/)
- The Stanford CoreNLP Natural Language Processing
- Toolkit (http://nlp.stanford.edu/software/)
- OpenNLP (http://opennlp.apache.org/)
- DeepNL (Deep Learning for Natural Language Processing)
- (https://github.com/attardi/deepnl)

#GOOGLEISYOURBESTFRIEND

https://github.com/Lab41/sunny-side-up/wiki/Learning-Resources-for-NLP,-Sentiment-Analysis,-and-Deep-Learning

A big THANKS to @SMUBIA #PythonWorkshop



THANKS!

Any questions?