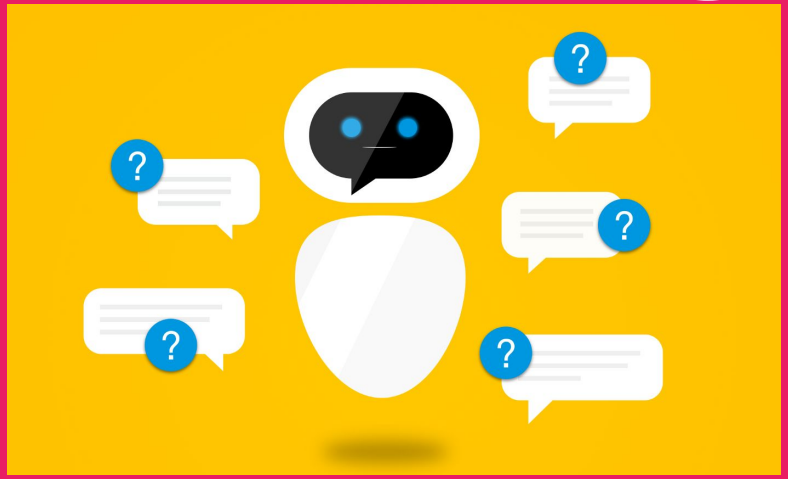
Natural Language Processing



Basics

What?

- Understanding + analyzing lang.
  - Speech to text
- Complex variables
  - Slang, slurs, regional dialects

Examples

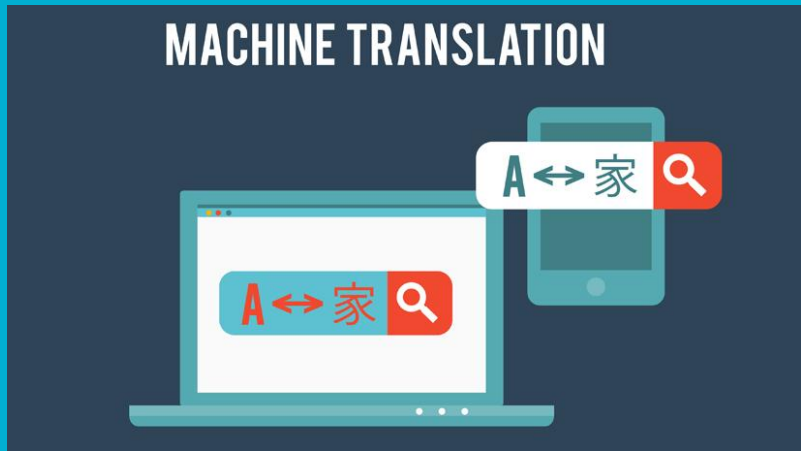
- Siri - Apple
- Galaxy - Samsung
- Watson - IBM
- Cortana - Windows
- Echo Dot - Amazon

SAMSUNG

Galaxy

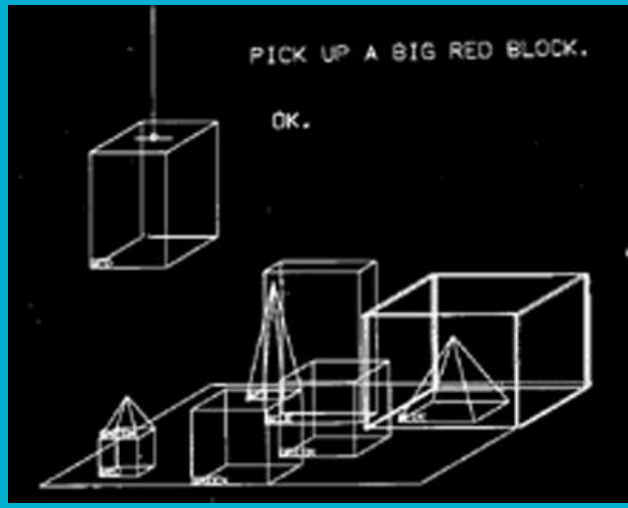
Examples (cont.)

- Machine Translation
  - Google Translate
- CHATBOTS!!!!
  - Applications interacting with users for better experience



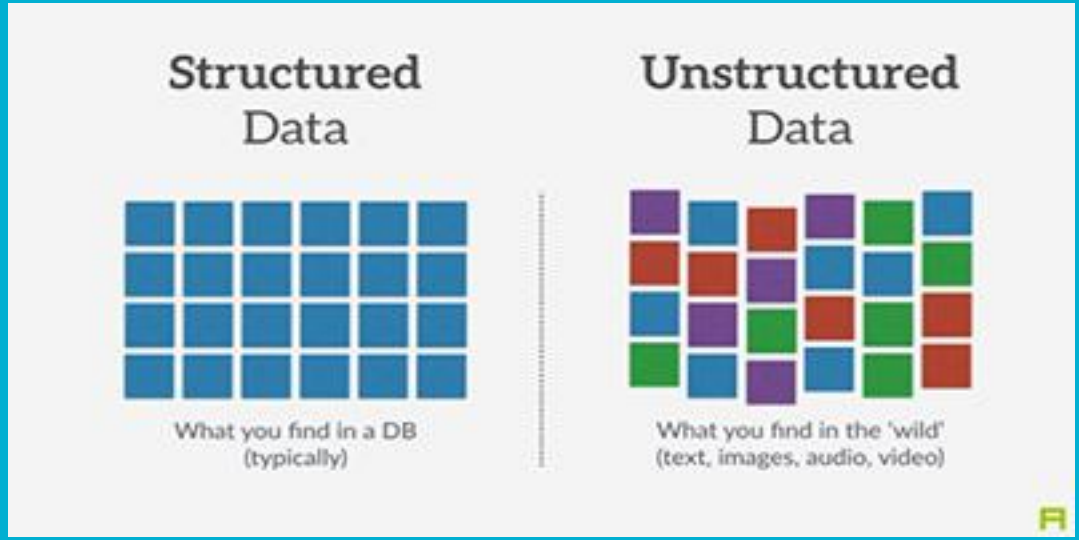
History

- 1966-1980
  - Powerful computers
- CHATBOT!!!!
- SHRDLU
  - Program dev. @ MIT 1968-1970
    - Lisp
  - Accurately understanding lang. & retain memory + understand word in other contexts
  - Real word applications



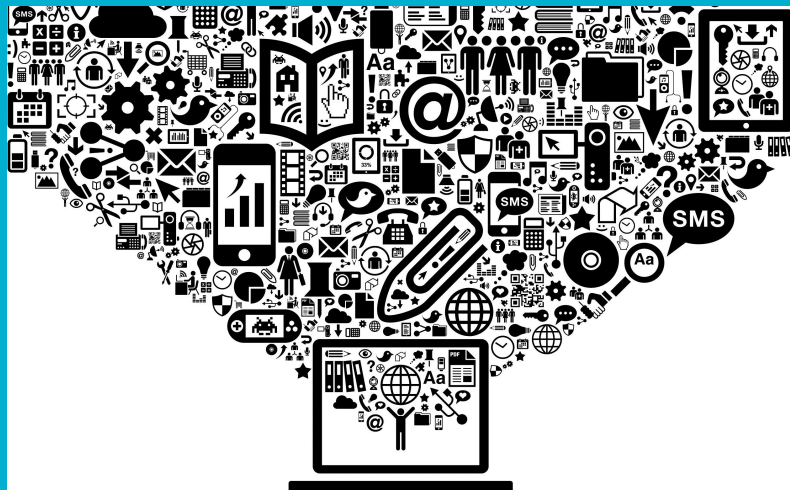
Structured vs. Unstructured Data

- Structured
  - Table
  - Charts
  - Data base
  - Very organized
- Unstructured
  - Long winding text
  - Difficult to organize

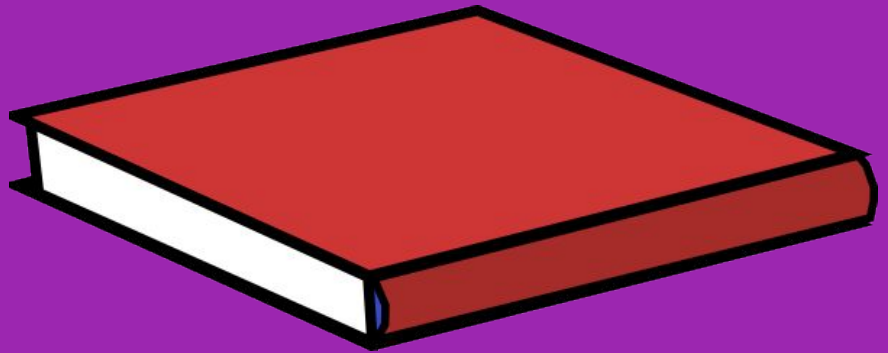


Importance/Impact

- Industry advantage
  - Med : record info in natural manner
- Natural way of communication
- 2020
  - Free form text
  - Available data : mostly text



Core Concepts



Stemming and Lemmatization

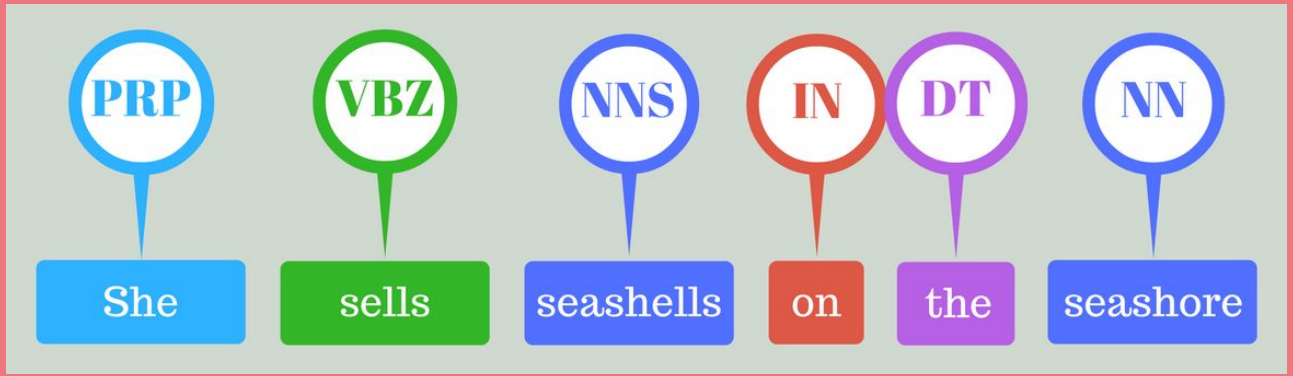
- Reducing word to base form
  - Lemma : base of word
- Stemming : “chopping off” word
  - Ex : amusing, amusement, amused : stem = amus
- Examples
  - I am a student → I be a student
  - My computer’s touch screen is not working → My computer touch screen be not working

Differences between Stemming & Lemmatization

- Differences
  - Stemmer : single word, no context
  - Lemma : appropriate base for diff. Scenarios
    - Ex : meeting
      - We are meeting tomorrow, In our meeting yesterday
- I saw the most awesome NLP bot!
  - Stem : I s the most awesome NLP bot!
  - Lemmatizing : I see the most NLP bot!
- Stemming
  - Flies → fli
  - Bananas → banana
- Lemmatization
  - Flies → fly
- Lemmatization more applied

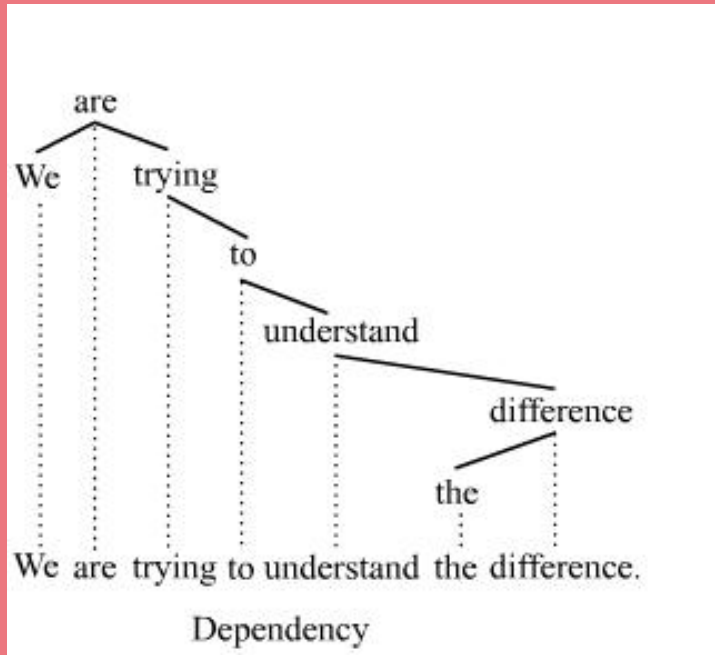
Part of Speech (POS) Tagging

- Objective : categorize & assign POS to each word
- I want to learn NLP
  - I (Preposition)
  - Want (Verb)
  - Learn (Verb)
  - NLP (Noun)



Dependency Parsing

- Understand dependencies and relationships of words
- Dependency tree
- Examples & Terms
  - Subject
  - Object
  - Complement
  - Modifier



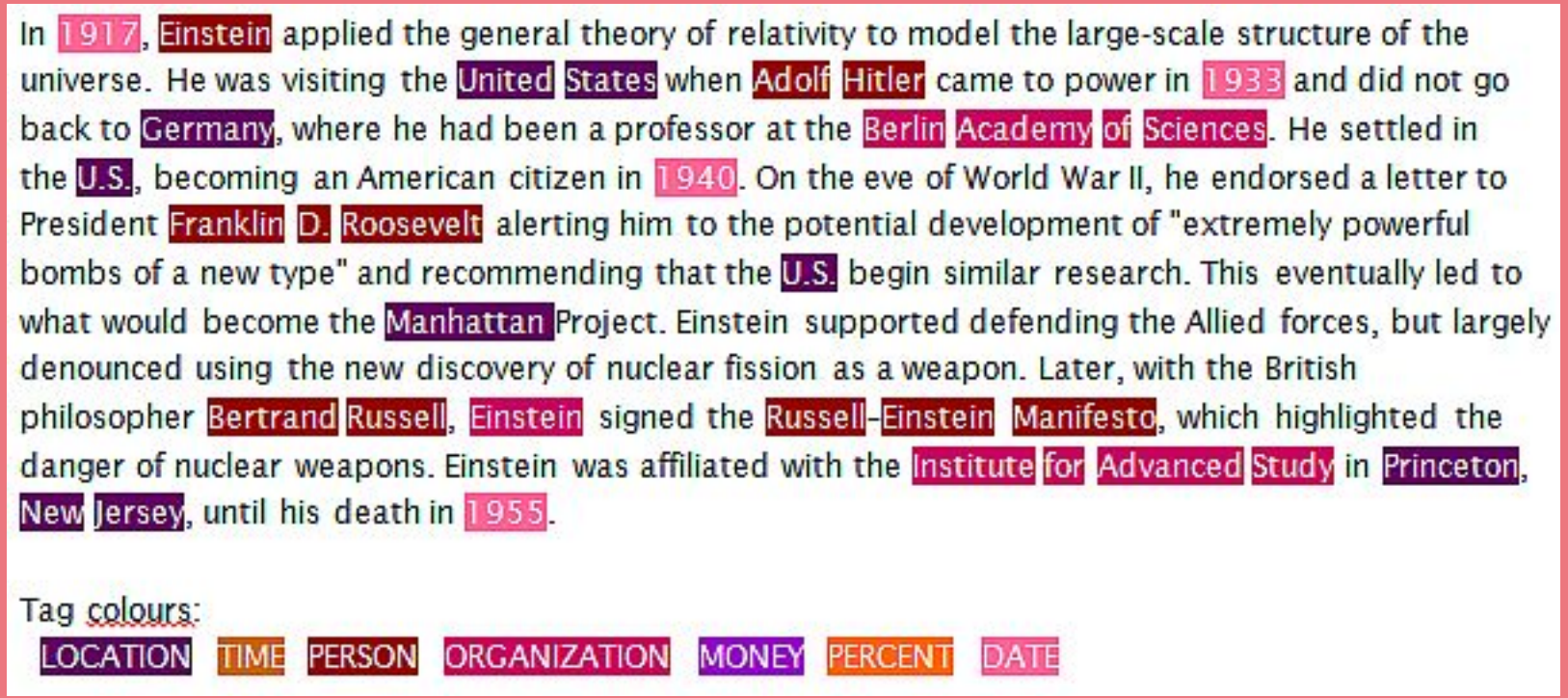
Dependency

Named Entity Recognition (NER)

- Associates entities (categories) with different words
- Examples
  - The Jungle Book : FILM
  - Plainsboro Public Library : LOCATION
  - Taylor Swift : SINGER/SONGWRITER

Figure 1: An example of NER application on an example text

NER Example

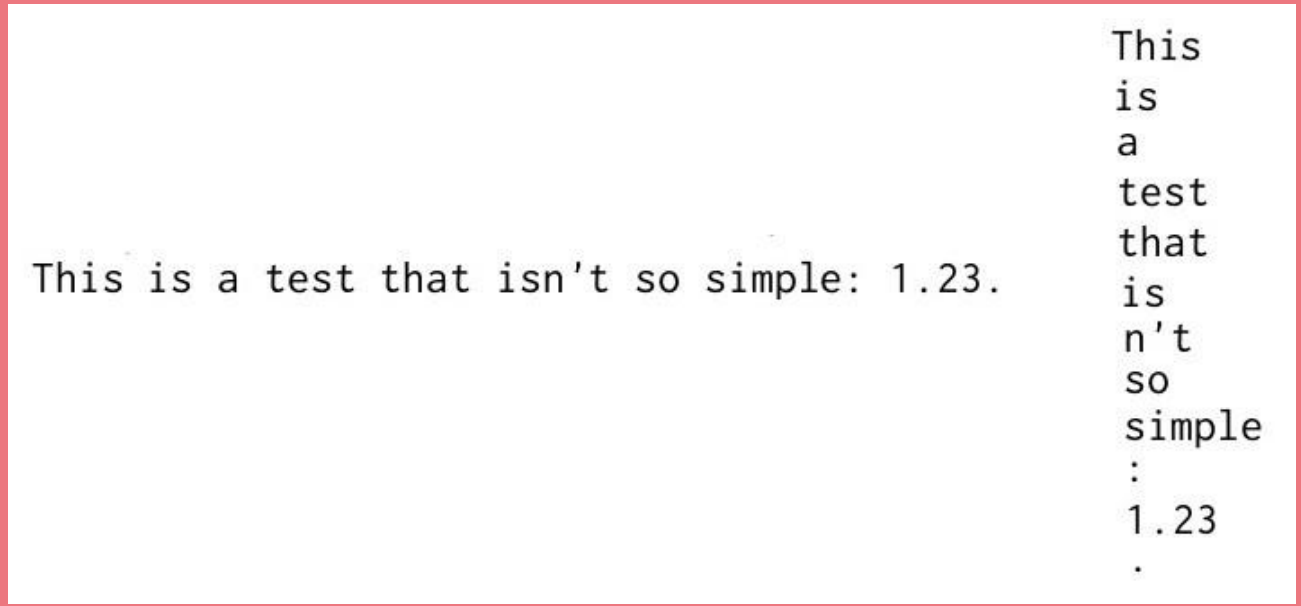

In 1917, Einstein applied the general theory of relativity to model the large-scale structure of the universe. He was visiting the United States when Adolf Hitler came to power in 1933 and did not go back to Germany, where he had been a professor at the Berlin Academy of Sciences. He settled in the U.S., becoming an American citizen in 1940. On the eve of World War II, he endorsed a letter to President Franklin D. Roosevelt alerting him to the potential development of "extremely powerful bombs of a new type" and recommending that the U.S. begin similar research. This eventually led to what would become the Manhattan Project. Einstein supported defending the Allied forces, but largely denounced using the new discovery of nuclear fission as a weapon. Later, with the British philosopher Bertrand Russell, Einstein signed the Russell-Einstein Manifesto, which highlighted the danger of nuclear weapons. Einstein was affiliated with the Institute for Advanced Study in Princeton, New Jersey, until his death in 1955.
Tag colours:
LOCATION
TIME
PERSON
ORGANIZATION
MONEY
PERCENT
DATE

Automatic Summarization

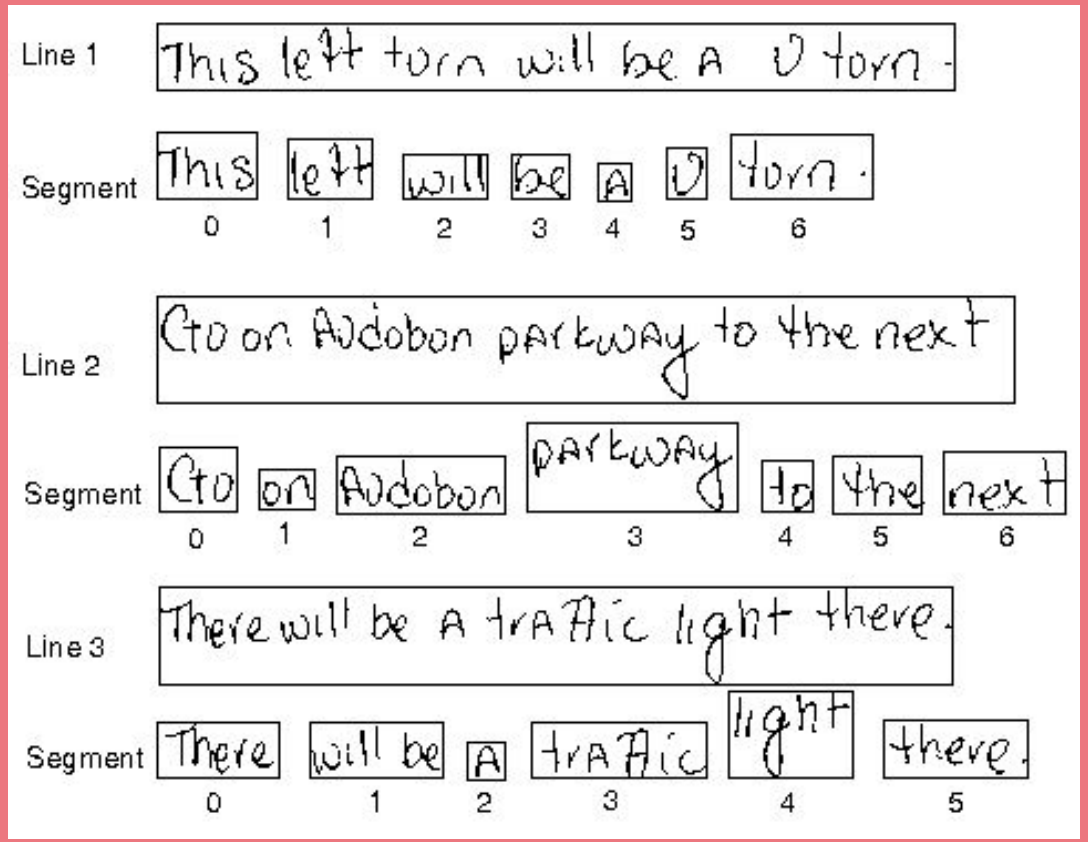
- Shortening, or summarizing long document
  - Accuracy
  - Major points
- Examples:
  - IBM Watson : ability to write good headline
    - Feed large amt. Of data & Watson has ability to summarize text

Tokenization

- Separate individual words & punctuation marks
  - Breaking up text by identifying positions of spaces & punctuation



Text Segmentation Example



Removal Of Stop Words

- Stop words
  - Don’t contain significance in search queries
  - Stop words are filtered out of queries (google search)
- Examples of Stop Words
  - A, about, along, behind, etc.
  - **http://xpo6.com/list-of-english-stop-words/**
- Improves efficiency
- Reduces file size

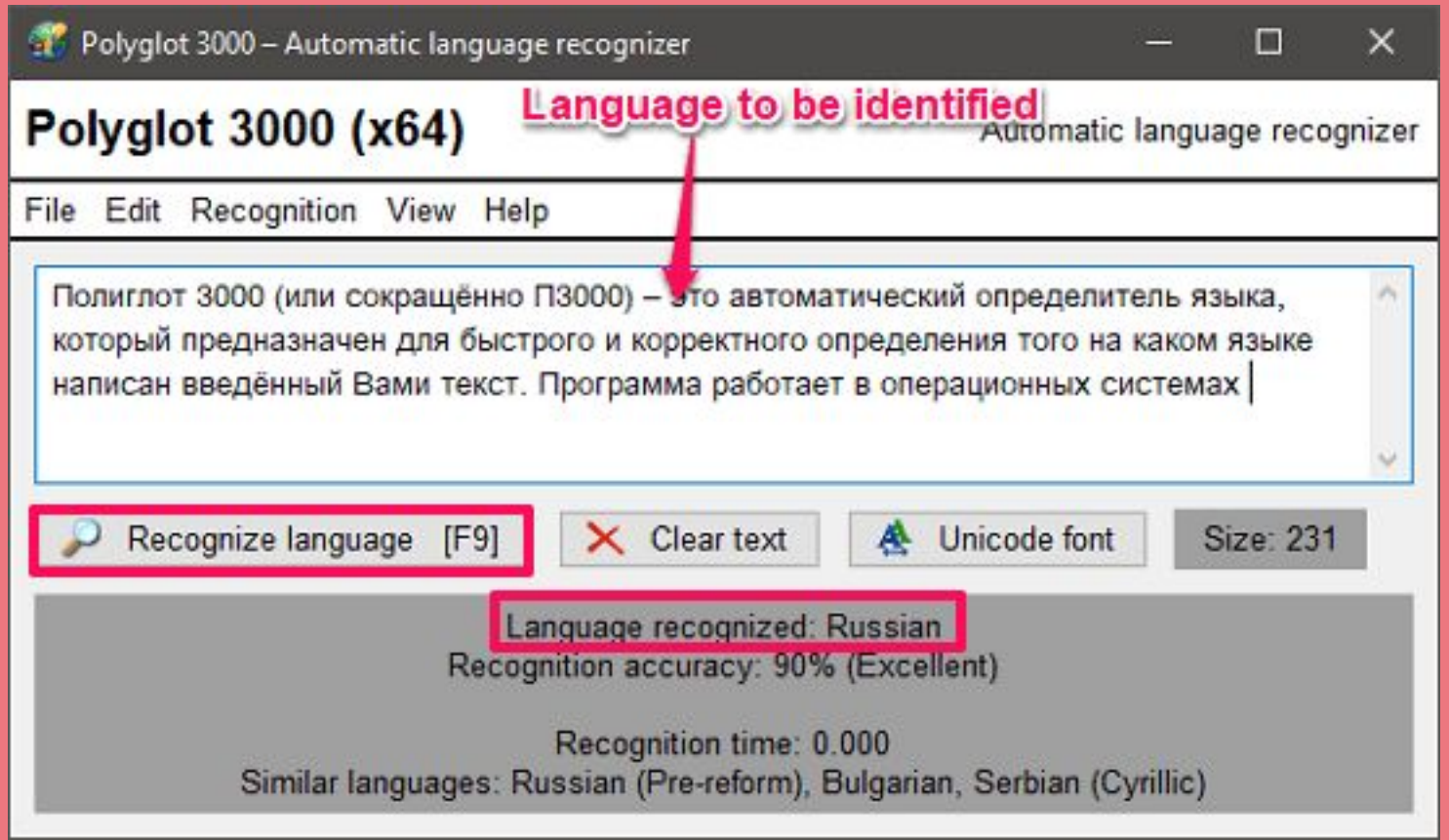
Table 3: Example of stop word removal.

A swimmer likes swimming, thus he swims.					
↓					
swimmer	likes	swimming	,	swims	.

Language Identification

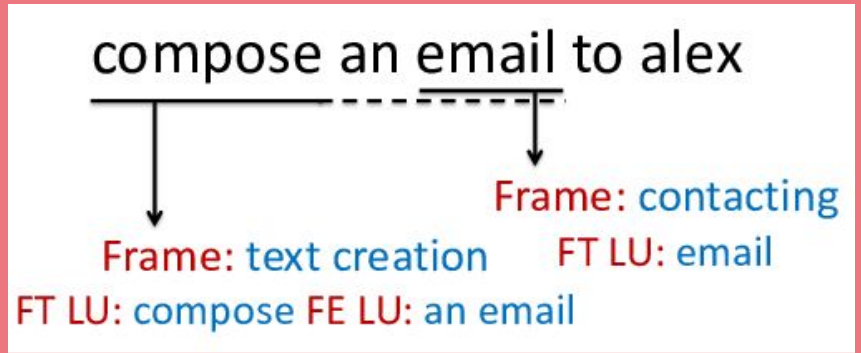
- Determines language of content/doc.
- After identifying language
  - Applies best dictionaries for each file
- Google Translate

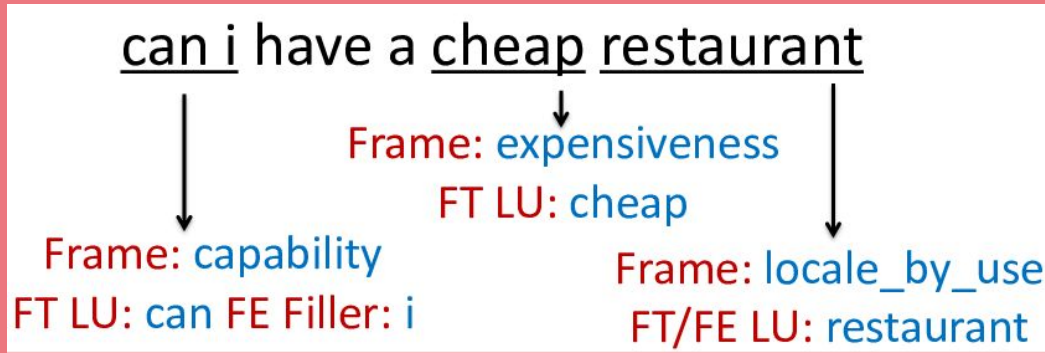
Language Identification Example



Semantic Parsing

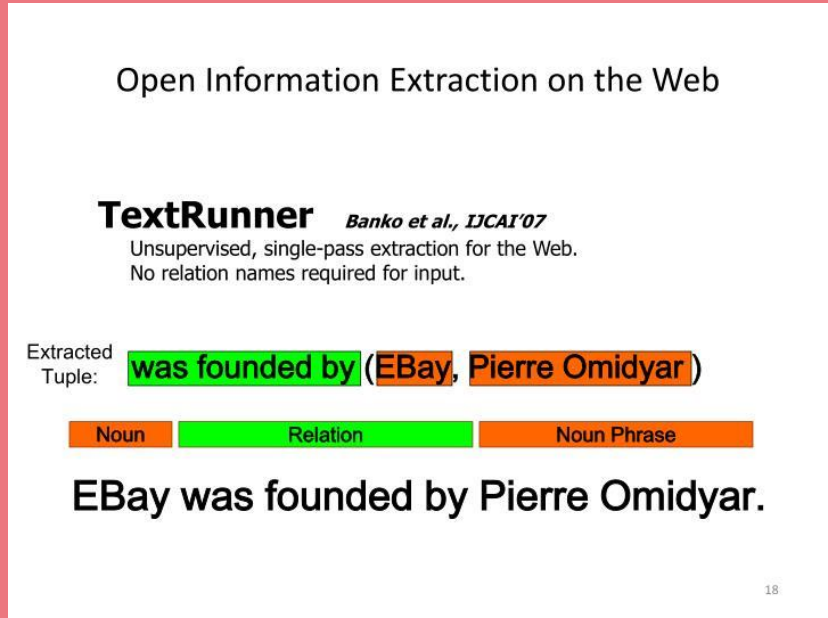
- Mapping sentence into a formal rep.
- Ability to identify roles
  - Student, location, food, etc.





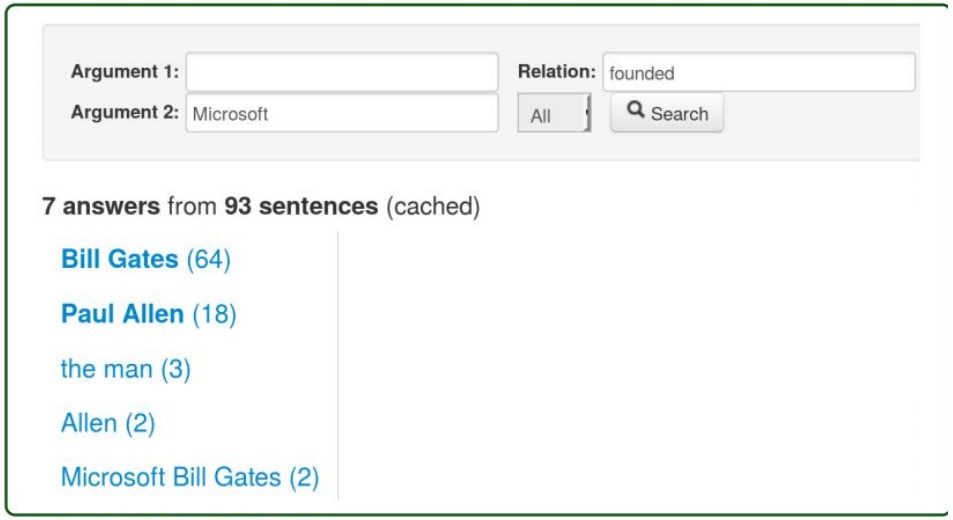
Open Information Extraction (OIE)

- Generates machine rep. Of text
  - Uses POS tagger to establish relationships
- Objective
  - Find relationships expressed in huge
  - How to find accurate relationships

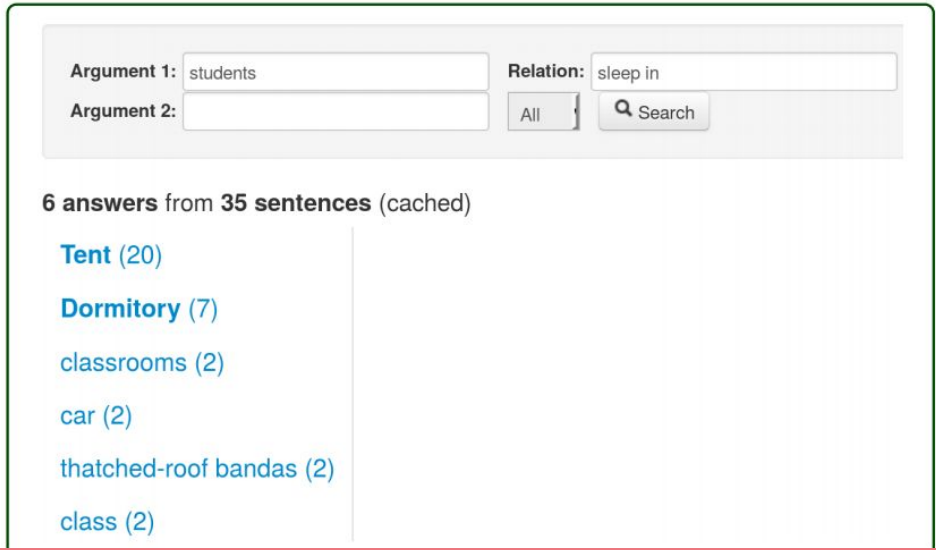


OIE Examples

Open IE: Motivation (5)



Open IE: Motivation (8)



Huck, Matthias, and Alexander Fraser. “Open Information Extraction.” *Http://Www.cis.uni-Muenchen.de/~Fraser/information_extraction_2016_lecture/12_open_IE.Pdf*, 18 Jan. 2017.

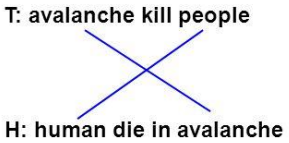
Textual Entailment

➢ Gives direct relation between text fragment & hypothesis

The Textual Entailment Task: More complex example

T: An avalanche has struck a popular skiing resort in Australia, killing at least 11 people.
H: Humans died in an avalanche.

- Again, need to match semantic roles:



- Again need for "frame semantics"
  - (syntax) X kill Y → (semantics) *cause* kill *victim*
  - (syntax) X died in Y → (semantics) *protagonist* died in *cause*
- ALSO:
  - *progagonist* isa *victim*, Killing → Death

BOEING

Textual Entailment

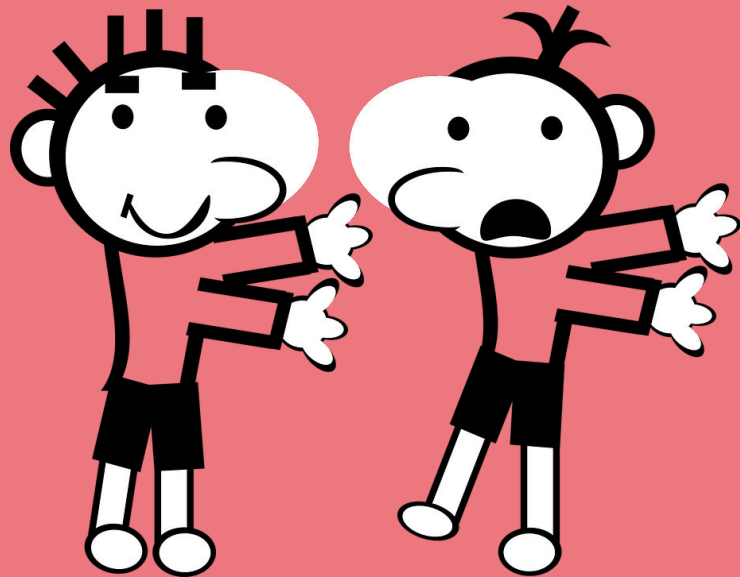
- Example 1 (TRUE)

Text:
His family has steadfastily denied the charges.

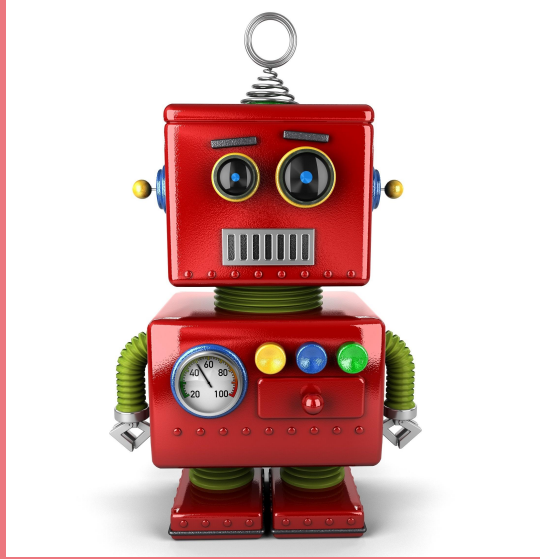
Hypothesis:
The charges were denied by his family.

Argumentation Mining

- Detection of arguments & relationships
  - Understanding POV (Point of View) of user
  - Point vs. counterpoint
- Argumentation model components
  - Claim
  - Data
  - Qualifier
  - Warrant
  - Rebuttal
  - Backing



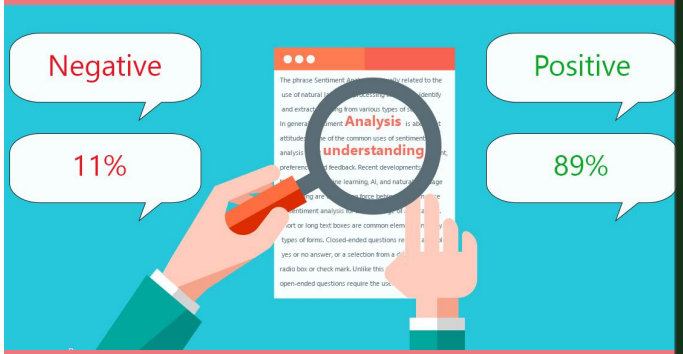
Linguistic Empathy



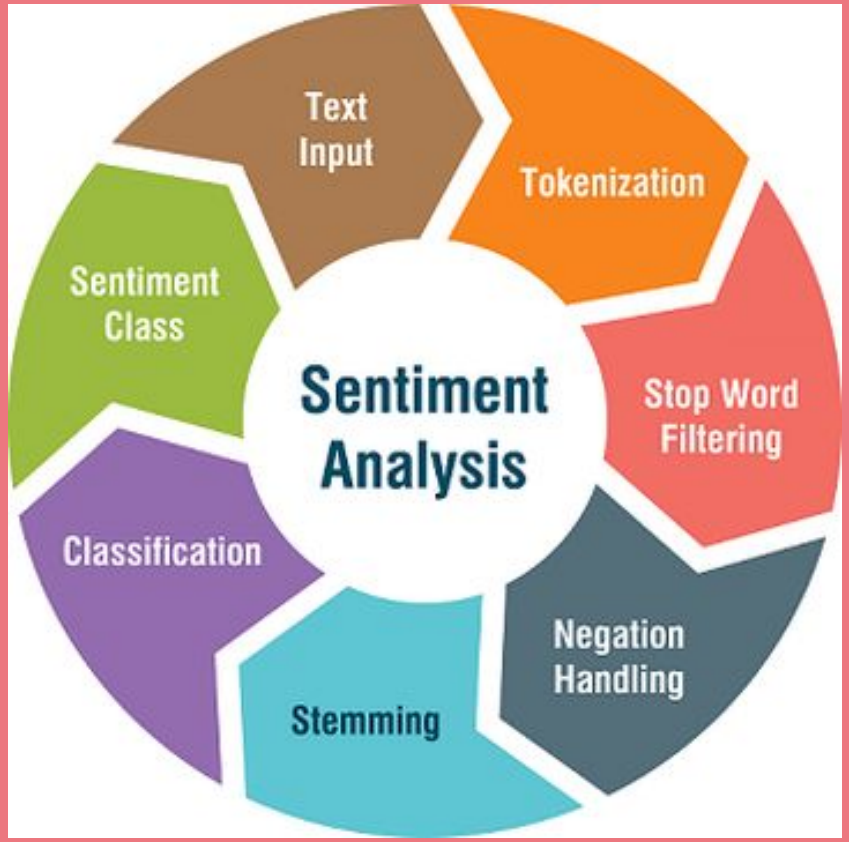
- Ability for machine to understand POV
  - Point of view of user
  - Empathize
- Empathic chatbots
  - Building feelings in robots

Sentiment Analysis

- Ability for machine to understand human emotion
- Categorizing + classifying opinions
  - Positive, negative, or neutral



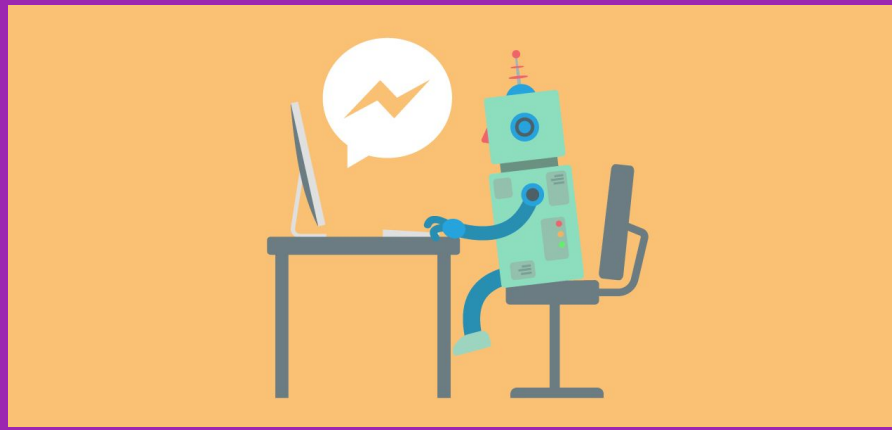
Steps to Sentiment Analysis



Sentiment Analysis Activity

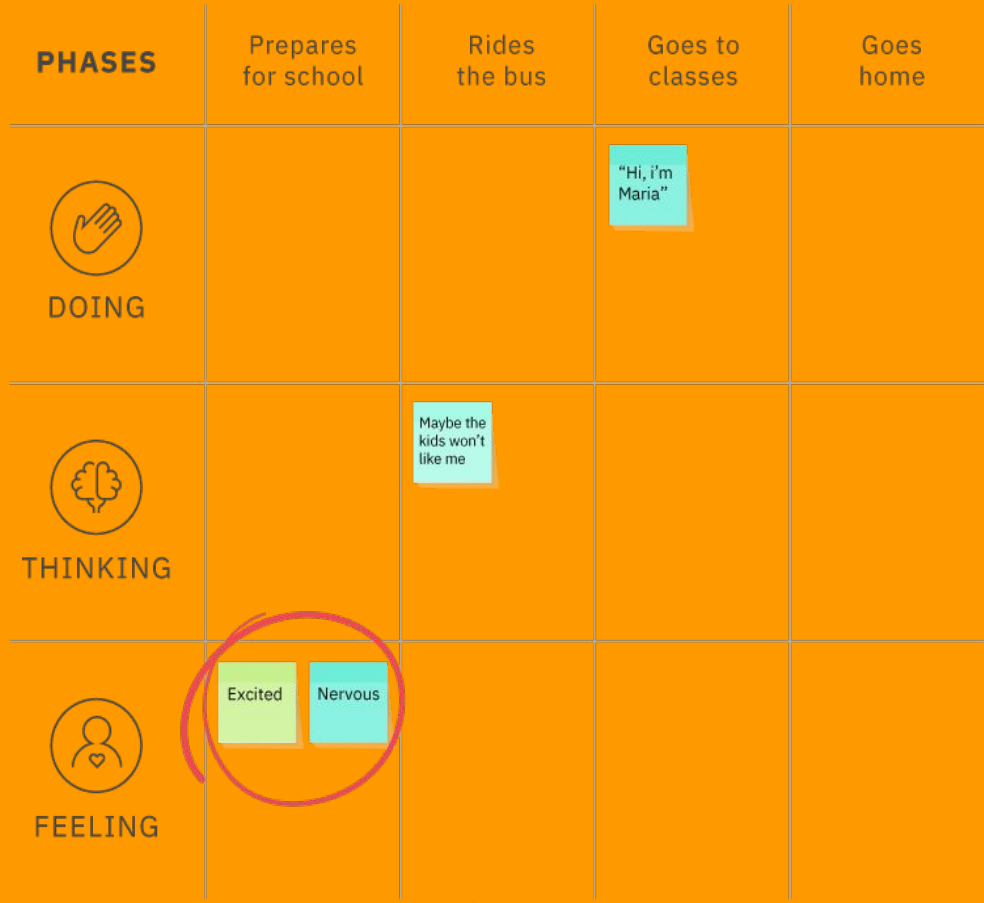
1. Gather materials
    a. Markers
    b. Paper
2. Focus on one sentiment
    a. Examples : angry, sad, nervous, etc.
3. Create a web
    a. In the middle : have sentiment, “anger”, or “sadness”
    b. Branch off words indicating feeling/sentiment(Noun)
        i. Ex : anger → frustrated, infuriated (Adjectives)
    c. Branch off poss. Sentences indicating feeling (Sentence, include verb)
4. Share & try with Siri/Galaxy/Cortana.

Chatbot Activity (IBM Chatbots for Good)



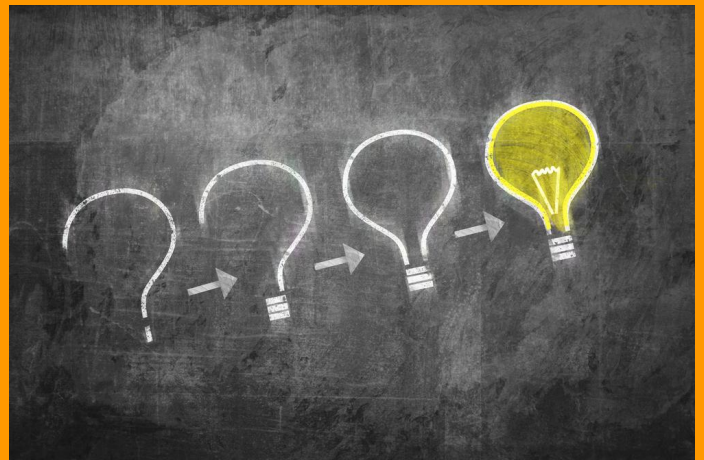
Step 1 : Scenario Map

- Draw 4 rows
- Phases row
  - Write down actions/phases
  - Ex : new student at school
    - Prepares for school, etc.
  - 4-6 phases
- Doing row
  - Actions person is performing
- Thinking row
  - Person's internal thoughts
- Feelings row
- Use marker to circle most intense moments (terrified, nervous)



Step 2 : Creating Solutions

- Choose one of the circled items
- Brainstorm possible solution ideas
  - Different POVs
    - Ex : As a teacher, I can guide a new student through the school.
    - Ex : As a student, I can help the new student by...



Step 3 : Code - Outside of Workshop @ Home

- Code and step by step process to create simple chatbot @ home
- Resources and info. @ end of presentation
  - Resources will be posted on Classroom

```
for i in people.data.users:
    response = client.api.statuses.user_timeline.get(screen_name=i.scre
    print 'Got', len(response.data), 'tweets from', i.screen_name
    if len(response.data) != 0:
        ltdate = response.data[0]['created_at']
        ltdate2 = datetime.strptime(ltdate,'%a %b %d %H:%M:%S +0000 %Y'
        today = datetime.now()
        howlong = (today-ltdate2).days
        if howlong < daywindow:
            print i.screen_name, 'has tweeted in the past' , daywindow,
            totaltweets += len(response.data)
            for j in response.data:
                if j.entities.urls:
                    for k in j.entities.urls:
                        newurl = k['expanded_url']
                        urlset.add((newurl, j.user.screen_name))
        else:
            print i.screen_name, 'has not tweeted in the past', daywind
```

What are Bots?

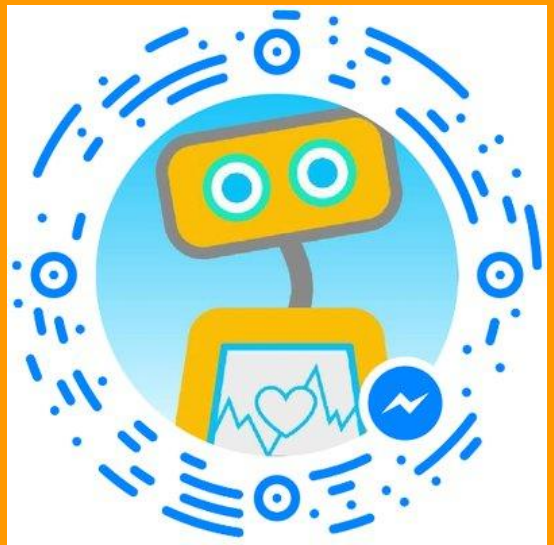
---

https://www.youtube.com/watch?time_continue=60&v=fEbzk4vTHsQ

https://www.youtube.com/watch?v=uE_WJTnqUwA

Examples of Sentiment Analysis Driven Chat Bots

- Woebot : https://woebot.io/
- Siri, Alexa, Cortana, Echo Dot, Watson, etc.
- https://www.youtube.com/watch?v=lmKqmPoU2ec



NLP With Java & Python

Resources

- Java
  - Eclipse - IDE
    - Basics on classes, etc.
    - If/else statements
    - Import Stanford Core NLP APIs or Apache OpenNLP
- Python
  - JetBrains PyCharms or IDLE
    - Basics on Python
    - Logic if/else statements
    - Import **Natural Language Toolkit (NLTK)** libraries
      - **NLTK** performs all the NLP concepts we discussed

Steps with Java/Python

---

1. Download & Install __
2. Go through tutorial posted on classroom (posted by tomorrow)
    a. Code I wrote, step by step procedure with meaning
    b. After understanding code, copy + run in IDE
3. Run and see the result!

How Watson Really Works

https://www.youtube.com/watch?v=_Xcmh1LQB9I

Cognitive Computing : Machine uses AI concepts to simulate human thinking by self learning through recognizing patterns and developing models