

# Introduction to Natural Language Processing

Maggie, Wenjie Li

Department of Computing

The Hong Kong Polytechnic University

# Natural Language Processing (NLP)



What is Natural Language Processing (NLP)?

# Natural Language Processing (NLP)



Can Hear, Can Read

Natural Language  
Understanding

Can Speak, Can Write

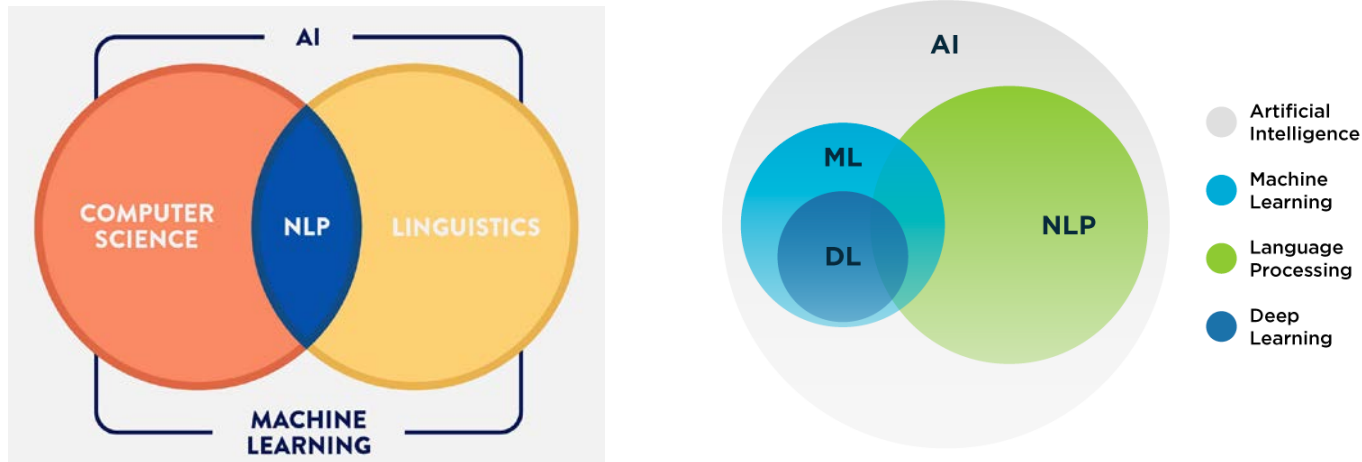
Natural Language  
Generation

Natural Language Processing (NLP) is a interdisciplinary field of linguistics, computer science and artificial intelligence, concerned with the **interactions between computers and human language**.

[Wikipedia]

# Natural Language Processing (NLP)

Natural Language Processing (NLP) is a subfield of **Artificial Intelligence** which has close ties with **Machine Learning**.

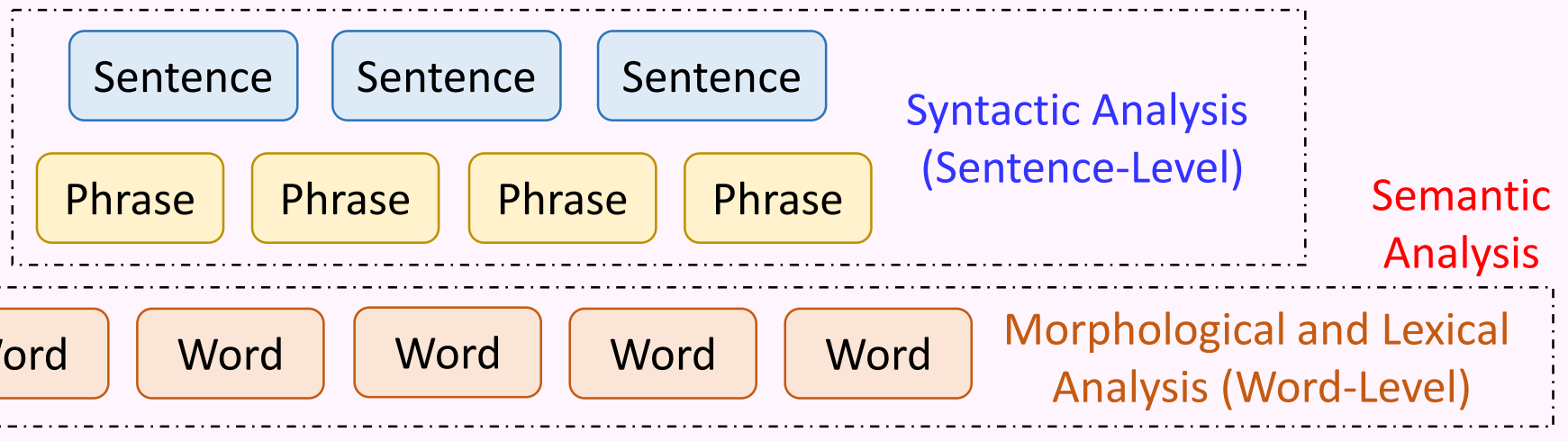
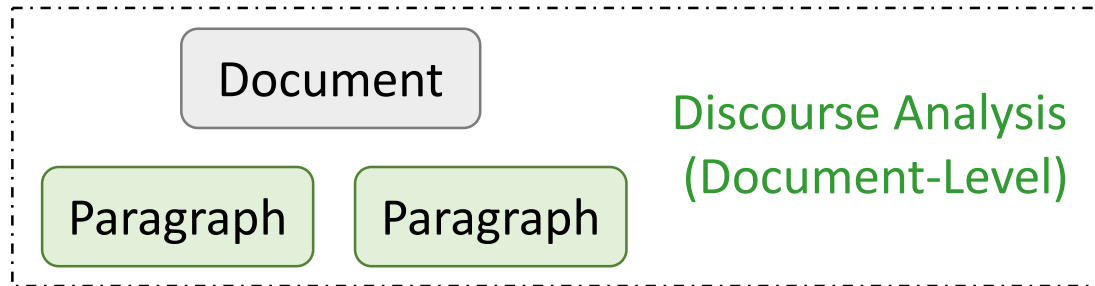


Natural Language Processing (NLP) is a interdisciplinary field of linguistics, computer science and artificial intelligence, concerned with the **interactions between computers and human language**.

[Wikipedia]

# Fundamental NLP Problems

- From Linguistics Point of View



# Fundamental NLP Problems

- Morphological and Lexical Analysis
  - Tokenization, Stemming/Lemmatization

Friends, Romans, Countrymen, lend me your ears.

Friends	Romans	Countrymen	lend	me	your	ears
Friend	Roman	Countryman	lend	me	your	ear

- Chinese Word Segmentation

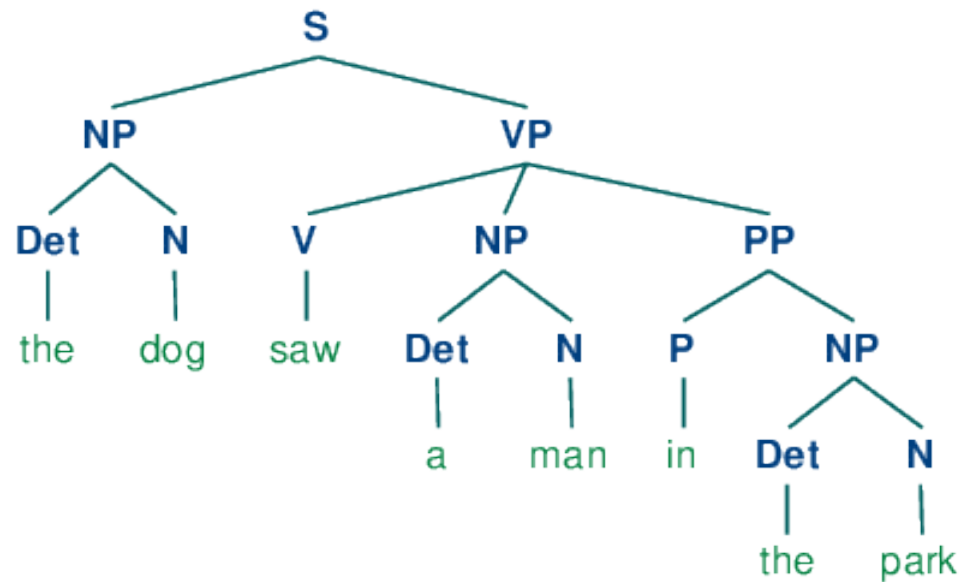
□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	4	□	9	□	□	□
□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	18	□	□	□	□							

□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	4	□	9			
□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	18	□	□	□	□	

# Fundamental NLP Problems

- Syntactic Analysis
  - Part-of-Speech Tagging and Syntactic Parsing



# Fundamental NLP Problems

- Semantic Analysis
  - Ambiguity in Word Meaning



...a **bank** can hold the investments  
in a custodial account ...



...as agriculture burgeons on  
the east **bank**, the river ...



.... a quiet animal like a **mouse**.

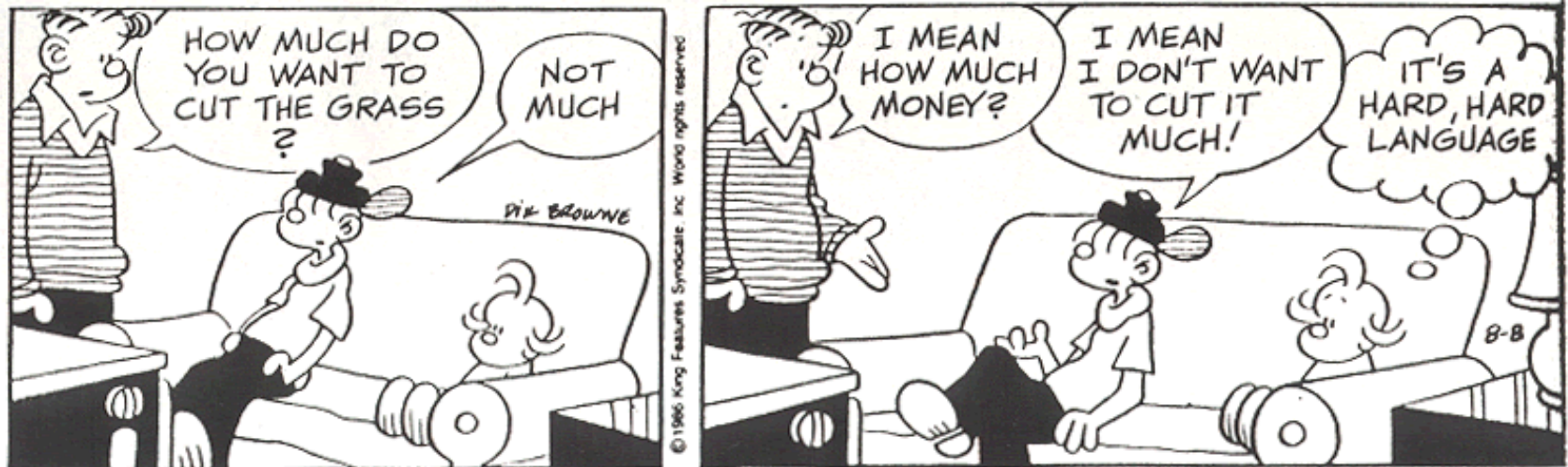


.... a **mouse** controlling a  
computer system in 1968.



# Fundamental NLP Problems

- Semantic Analysis
  - Ambiguity in Sentence Meaning

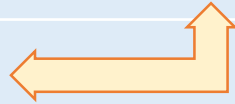


# Fundamental NLP Problems

- Discourse Analysis
  - Co-reference Resolution

I used to have the **key**.

But I lost **it**.



- Discourse Coherence

John took a train from Paris to Istanbul.

He likes spinach.

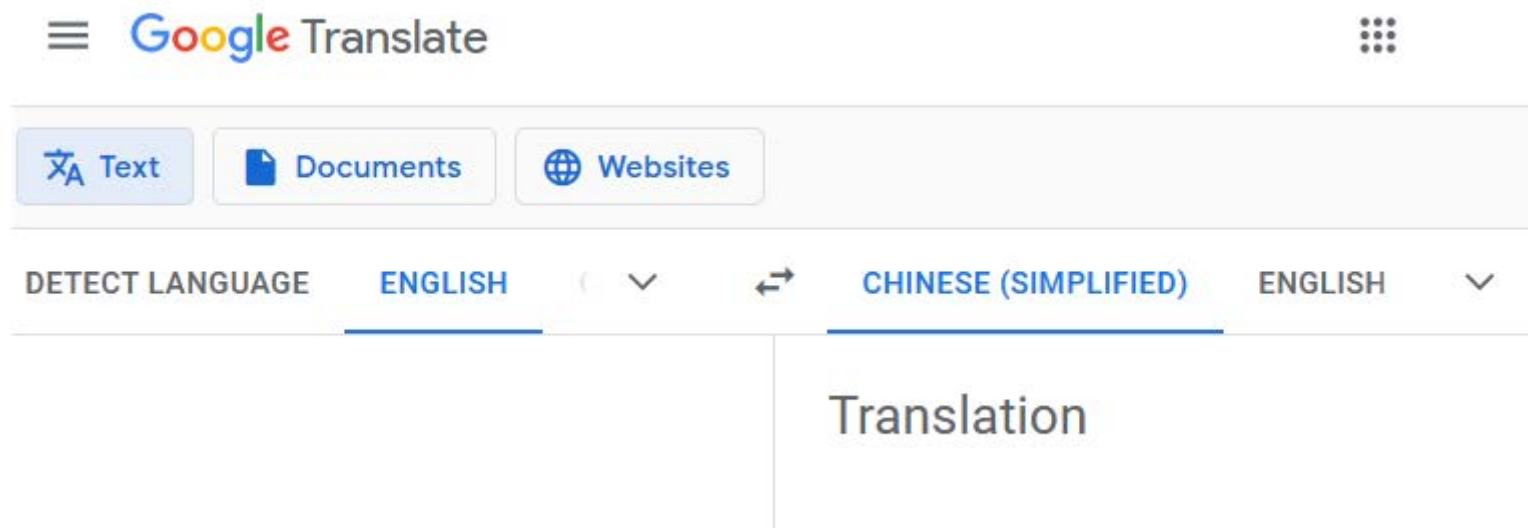
Jane took a train from Paris to Istanbul.

She had to attend a conference.

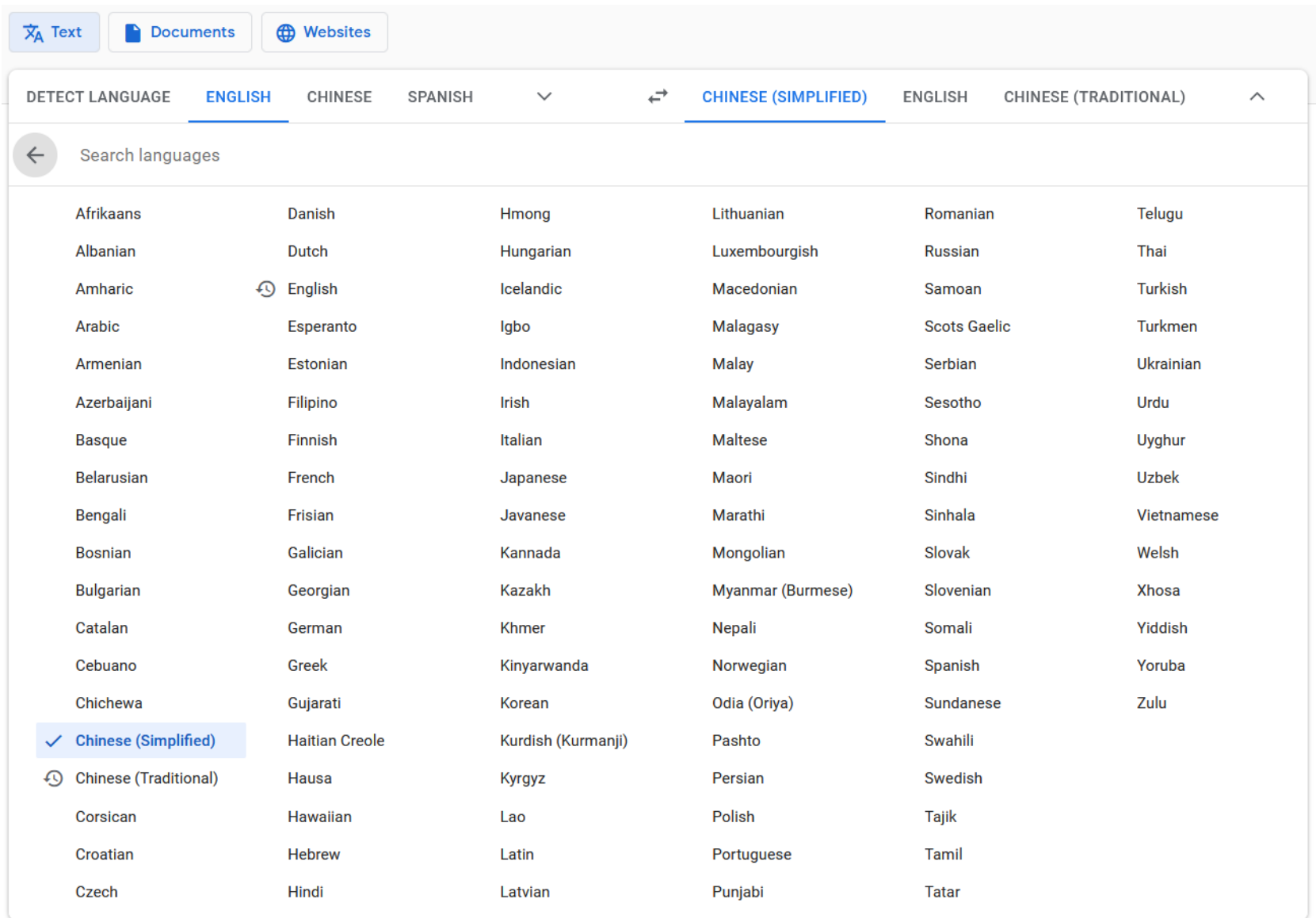


# Wide Range of NLP Applications

- Machine Translation



- It requires morphological, syntactic and semantic analysis/understanding and text **generation** techniques.



Google Translate supports more than 100 languages.

# Wide Range of NLP Applications

- Sentiment and Emotion Analysis



My experience  
so far has been  
fantastic!

POSITIVE



The product is  
ok I guess

NEUTRAL



Your support team  
is useless

NEGATIVE



- It requires linguistic/text feature engineering (or representation learning) and **text classification** techniques.

# Wide Range of NLP Applications

- Information Retrieval (IR)

The Google logo, consisting of the word "Google" in its characteristic multi-colored font.A white, rounded rectangular search bar with a magnifying glass icon on the left side.The Baidu logo, featuring the word "Baidu" in red and blue, with a blue paw print icon between the characters "ai" and "du".A white, rounded rectangular search bar with a camera icon on the right side and a blue button labeled "百度一下" (Baidu一下) next to it.

- It requires morphological (e.g., tokenization and stemming) analysis in NLP and **text ranking** techniques.

# Wide Range of NLP Applications



information retrieval



All

Images

Videos

News

Books

More

Tools

About 1,860,000,000 results (0.48 seconds)

<https://nlp.stanford.edu> › IR-book › information-retriev...

## Introduction to Information Retrieval

The book aims to provide a modern approach to **information retrieval** from a computer science perspective. It is based on a course we have been teaching in ...

You've visited this page many times. Last visit: 2/12/22

<https://nlp.stanford.edu> › html › htmledition › irbook

## Introduction to Information Retrieval - Stanford NLP Group

Introduction to **Information Retrieval**. By Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze. Website: <http://informationretrieval.org/>.

<https://www.geeksforgeeks.org> › what-is-information-re...

## What is Information Retrieval? - GeeksforGeeks

25 Mar 2022 — **Information Retrieval** is the activity of obtaining material that can usually be documented on an unstructured nature i.e. usually text which ...

<https://www.tutorialspoint.com> › natural\_language\_pro...

## NLP - Information Retrieval - Tutorialspoint

**Information retrieval** (IR) may be defined as a software program that deals with the organization, storage, retrieval and evaluation of information from ...

# Wide Range of NLP Applications

- Text Summarization

I just need  
the main ideas



- It requires **text ranking** (keyword word/phrase/sentence identification), **sentence retrieval/generation** techniques.



# Wide Range of NLP Applications

- Information Extraction (IE)



Person p Loc l Org o Event e Date d Other z

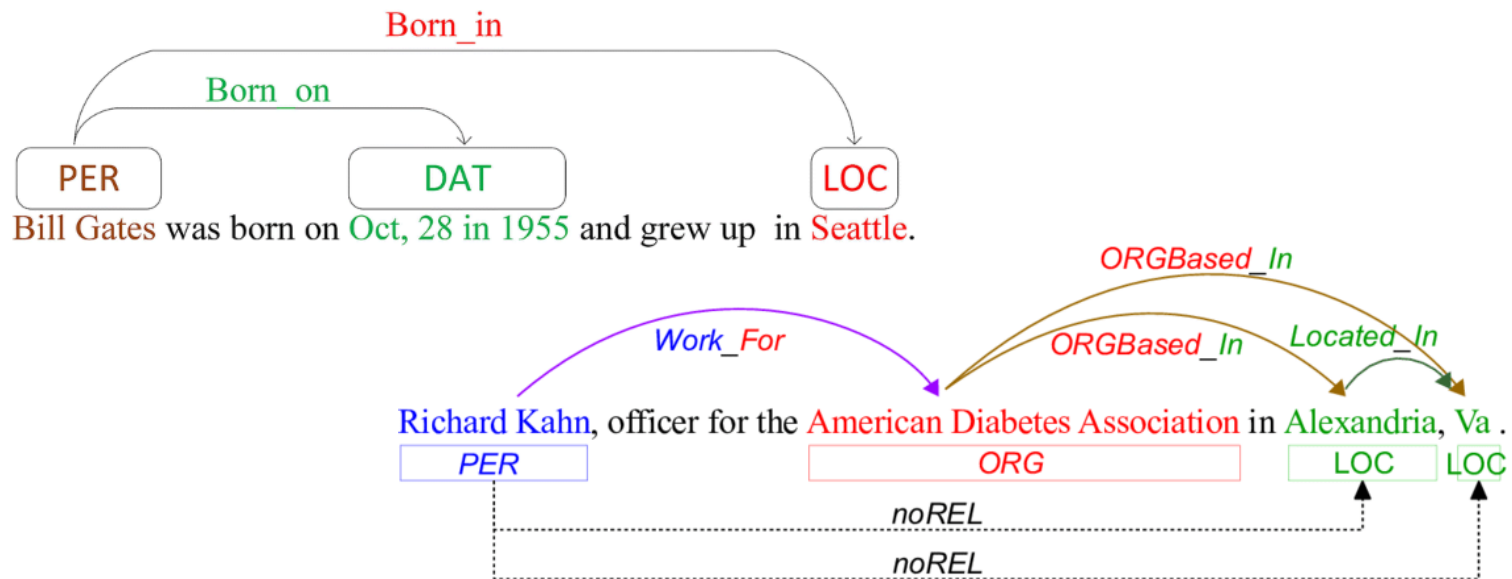
Barack Hussein Obama II (born August 4, 1961) is an American attorney and politician who served as the 44th President of the United States from January 20, 2009, to January 20, 2017. A member of the Democratic Party, he was the first African American to serve as president. He was previously a United States Senator from Illinois and a member of the Illinois State Senate.

## Named Entity Recognition

- It requires text feature engineering (or representation learning) and **text segment extraction** techniques.

# Wide Range of NLP Applications

- Information Extraction (IE)



## Entity Relationship Recognition

- It requires text feature engineering (or representation learning) and **text classification** techniques.

- Information Extraction (IE)



# Wide Range of NLP Applications

- Question Answering

## Passage 1

Bacteria are extremely small living things. While we measure our own sizes in inches or centimeters, bacterial size is measured in microns. One micron is a thousandth of a millimeter a pinhead is about a millimeter across. Rod shaped bacteria are usually from two to four microns long, while rounded ones are generally one micron in diameter. Thus if you enlarged a founded bacterium a thousand times, it would be just about the size of a pinhead. An adult human magnified by the same amount would be over a mile (1.6 kilometers).

Even with an ordinary microscope, you must look closely to see. At a magnification of 100 times, one finds that bacteria are barely visible and cannot make out anything of their structure. Using special stains, one can see that bacteria have attached to them wavy - looking "hairs" called flagella. The flagella rotate, pushing the bacteria through the water and cannot move about by their own power while others can glide by means of a little understood mechanism.

From the bacterial point of view, the world is a very different place. To a bacterium water is as thick as molasses is to us. Bacteria are influenced by the movements of the chemical molecules around them. Under a microscope, even those with no flagella, often bounce about in the water and collide with the water molecules and are pushed this way and that. That within a tenth of a second the molecules around a bacterium have changed. Even bacteria without flagella are thus constantly exposed to

1. Which of the following is the main topic of the passage?  
(A) The characteristics of bacteria  
(B) How bacteria reproduce  
(C) The various functions of bacteria  
(D) How bacteria contribute to disease
2. Bacteria are measured in  
(A) inches  
(B) centimeters  
(C) microns  
(D) millimeters
3. Which of the following is the smallest?  
(A) A pinhead  
(B) A rounded bacterium  
(C) A microscope  
(D) A rod-shaped bacterium
4. According to the passage, someone who examines bacteria using only a microscope that magnifies 100 times would see  
(A) tiny dots  
(B) small "hairs"  
(C) large rods  
(D) detailed structures
5. The relationship between a bacterium and its flagella is most nearly analogous to which of the following?  
(A) A rider jumping on a horse's back  
(B) A ball being hit by a bat  
(C) A boat powered by a motor  
(D) A door closed by a gust of wind
6. In line 16, the author compares water to molasses, in order to introduce which of the following topics?  
(A) The bacterial content of different liquids  
(B) What happens when bacteria are added to molasses  
(C) The molecular structures of different chemicals  
(D) How difficult it is for bacteria to move through water

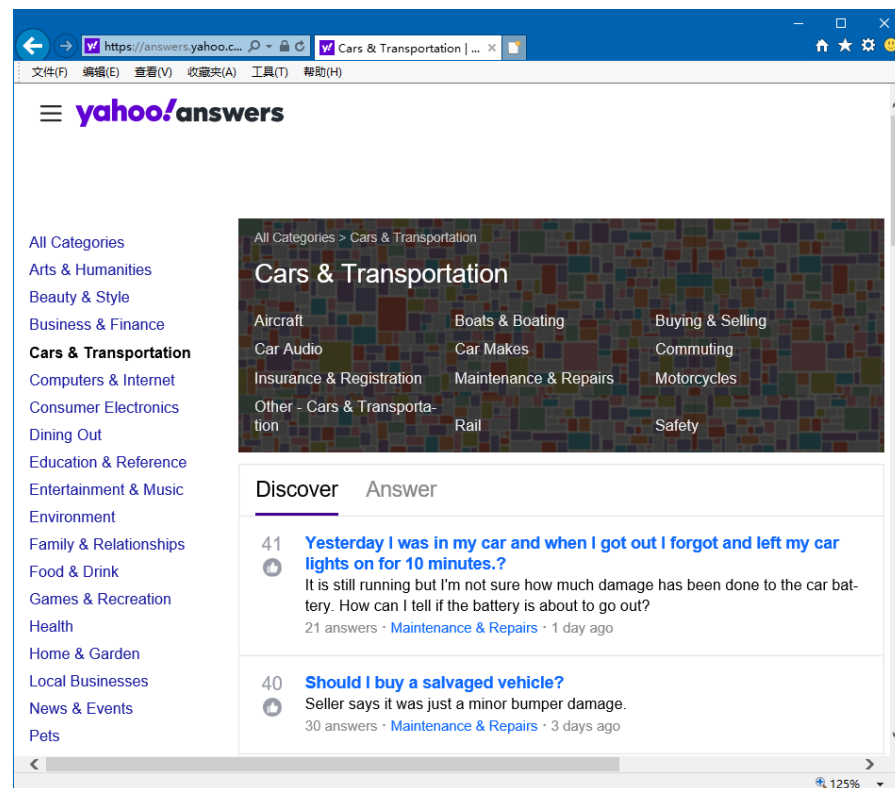
## Multiple Choice QA

# Wide Range of NLP Applications

- Question Answering



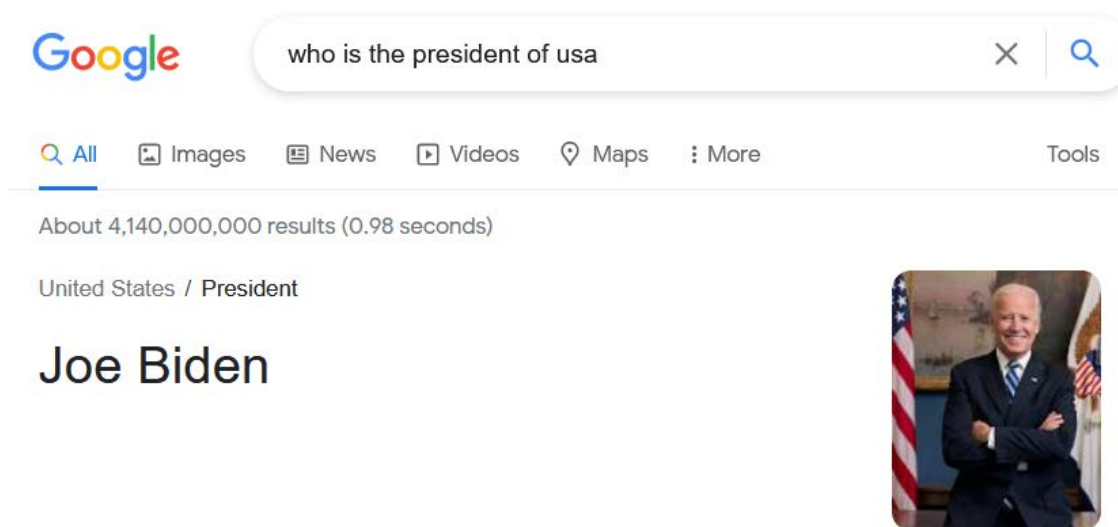
Document-based QA (DBQA)



Community QA

# Wide Range of NLP Applications

- Question Answering

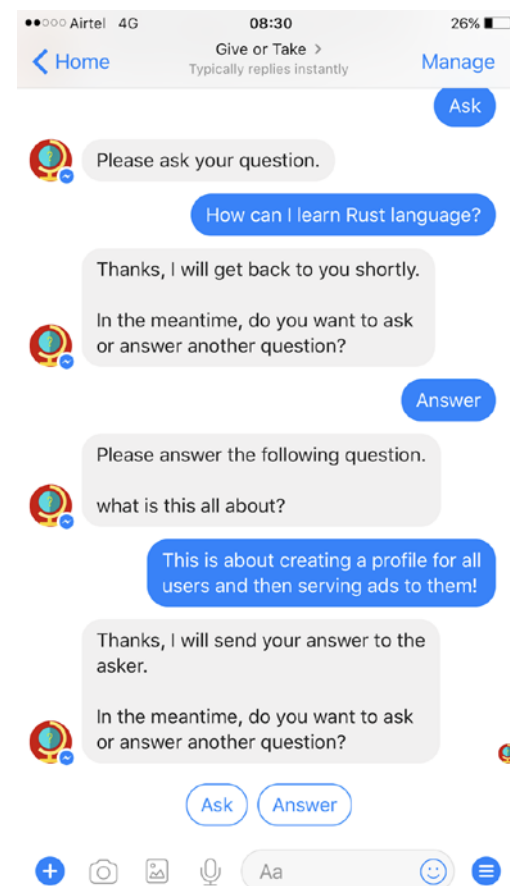
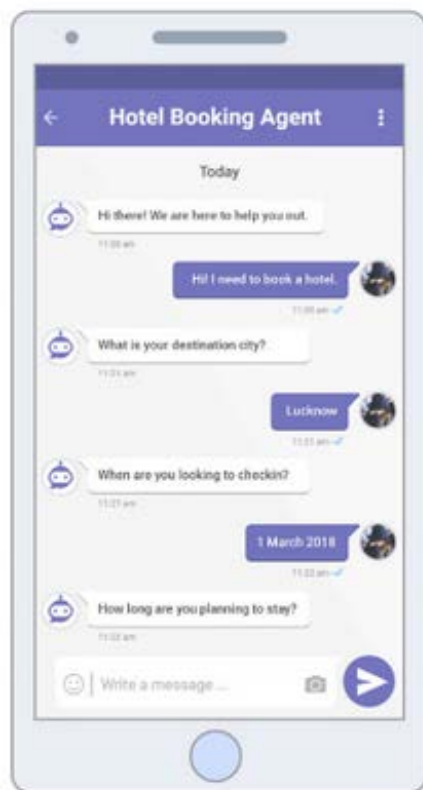


Open Domain QA

- It requires text **classification**, **information retrieval**, and **information extraction** techniques.

# Wide Range of NLP Applications

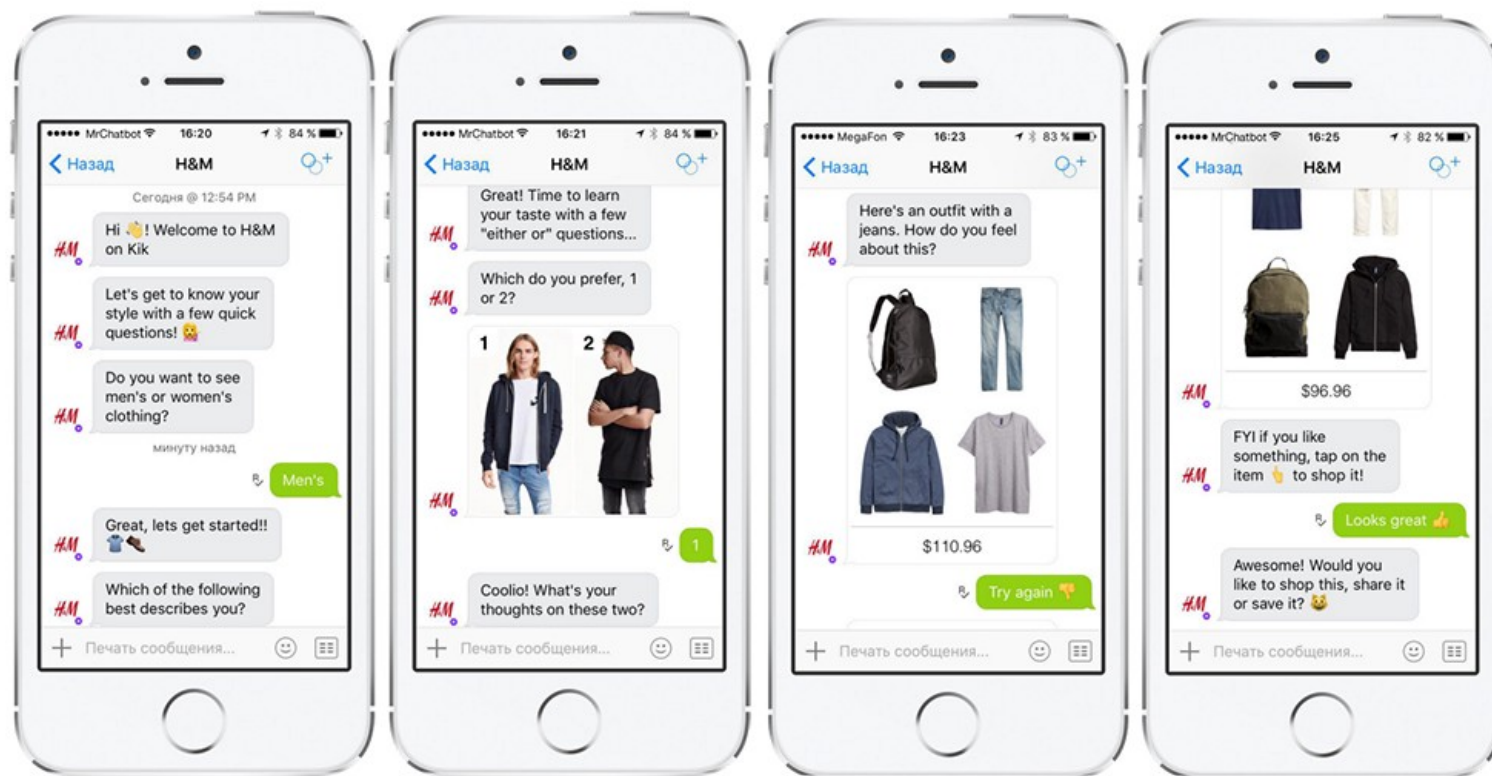
- Dialogue Systems (Chatbots)
  - Task-Oriented





# Wide Range of NLP Applications

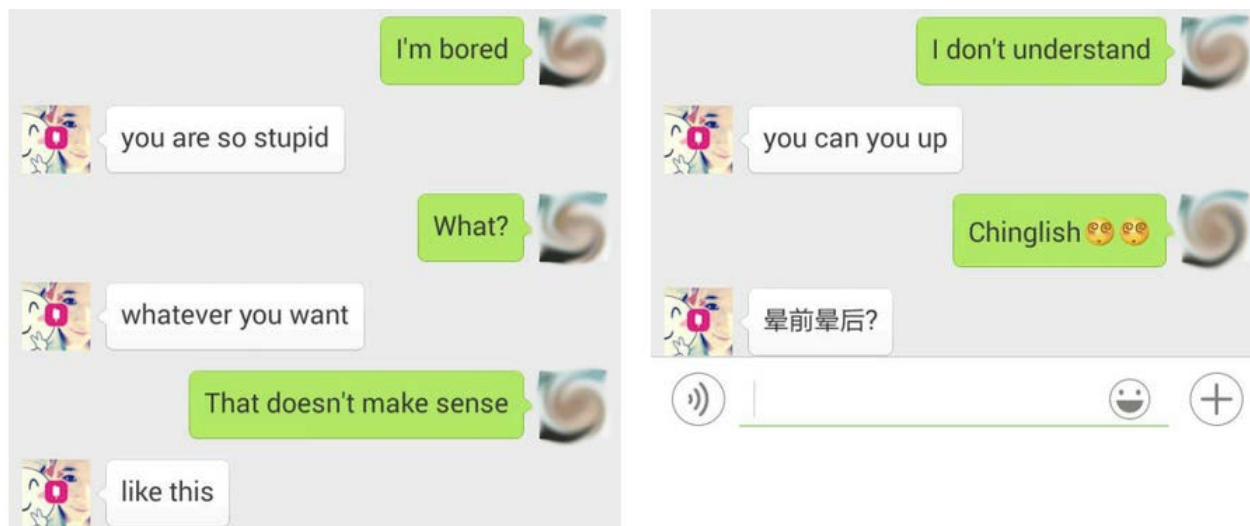
- Dialogue Systems (Chatbots)
  - Task-Oriented





# Wide Range of NLP Applications

- Dialogue Systems
  - Social-Oriented (Chatbots)



- It requires language **understanding**, language **generation** and human-machine **communication** (dialogue management) techniques.

# Wide Range of NLP Applications

- Summary

