

# Natural Language Processing

Introduction: Class 1-3

# Syllabus

- Introduction
- NLP-Hard
- Finite State Automata
- HMM
- Phonetics & Phonology
- Syntax and Parsing
- Semantic Parsing
- Discourse
- Applications
- Regular Expressions
- N-Gram
- Morphology
- Part of Speech tagging
- Named Entity Recognition
- Pragmatics

# Marks BreakUp

- Grading

- Project Task 1 (17%)

- Presentation 1: Project Proposal/Literature Review (7%) **(Mid Sept before Sept 15)**

- Presentation 2: Midterm report (10%) **(Oct 25-28)**

- Project Task 2 (28%)

- Project 2 implementation and presentation with term paper(min 8 page double column) **(Nov 20-25)**

- Assignment/Quiz (15%)

- Mid 1 (10%)

- Mid 2 (10%)

- End Sem (20%)

# Introduction

- Natural Language Processing (NLP)
  - It is the study of the computational treatment of natural (human) language.
  - In other words, teaching computers how to understand (and generate) human language.

# Applications

- Search engines
  - Google, Yahoo, Bing, Baidu
- Question answering
  - IBM's Watson
- Natural language assistants
  - Apple's Siri, MS Cortana
- Translation systems
  - Google Translate
- News digest
  - Yahoo!
- Automated earthquake reports
  - LA Times
- Automated stock market reports
  - Narrative Science

# Introduction: Watson



<http://www.geekwire.com/2013/ibm-takes-watson-cloud/>

# Introduction: Watson

- In February 2011, the world was introduced to Watson (after Thomas J. Watson, the founder of IBM), IBM's cognitive computing system that defeated **Ken Jennings** and **Brad Rutter** at the American TV quiz show Jeopardy!, and received the 1<sup>st</sup> place prize of \$1 million.
- The program, which has been broadcast in the United States for more than 25 years, pits 3 human contestants against one another to answer rich natural language questions over a board.
- Watson was developed in IBM's DeepQA project by a research team led by principal investigator David Ferrucci.

# Introduction: Watson

- Watson is a Question-Answering (QA) system that applies advanced natural language processing, information retrieval, knowledge representation, automated reasoning, and machine learning technologies to the field of open domain question answering.
- Watson had access to **200 million pages** of structured and unstructured content consuming **4 terabytes** of disk storage including dictionaries, references, and the full text of Wikipedia, but was not connected to the Internet during the game.





# Introduction: Watson



How Watson Works

# Introduction: Watson

- ❖ What does “divergent” mean?
- ❖ What year was Abraham Lincoln born?
- ❖ How many states were in the United States that year?
- ❖ How much Chinese silk was exported to England in the end of the 18th century?
- ❖ What do scientists think about the ethics of human cloning?

# **Introduction: Watson**

IBM Watson Demo

# Introduction: Information Extraction

Event: NLP meeting

Date: 16-08-2020

Start: 11:00am

End: 11:30am

Location: IIIT Sri City Auditorium

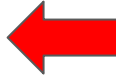
Subject: **NLP meeting** 

Date: August 10, 2020

To: XYZ

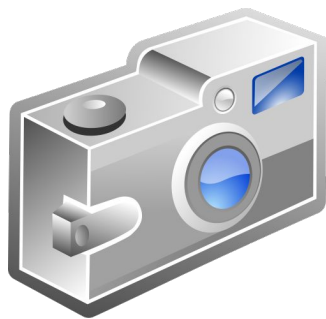
---

Dear XYZ,

Good morning, hope you are doing well. For the upcoming semester a meeting is scheduled on 11-11:30AM, 16th August 2020 at IIIT Sri City Auditorium 

-

# Introduction: Information Extraction



## Attributes:

zoom  
affordability  
size and weight  
flash  
ease of use

## Size and weight

- ✓ nice and compact to carry!
- ✓ since the camera is small and light, around those heavy, bulky profess
- ✗ the camera feels flimsy, is plastic a have to be very delicate in the han

★★★★★ **Definitely worth it.**

Reviewed in the United States on 4 May 2018

**Verified Purchase**

I got a refurbished d5600 from Beach Camera two days ago and I love it. It works great so far. Connected to snapbridge with no problems, pics transfer flawlessly. I bought it body only, so as of right now I use my old 35mm lens and I love the pictures. I just love the sound of d5600 shutter (compared to my d3100 lol), I love how I can turn the lcd screen to take pics high up or down low, love the touchscreen also. The camera is a great size, lighter then what I'm used to with d3100, also the grip fits in my hand so much better, and I am not tired to carry it around in my hand. (D3100 I had to hang on my neck.) Great camera! I am happy with this purchase.



36 people found this helpful

Helpful

Report abuse



Raghul S.

★★★★★ **Great entry level DSLR**

Reviewed in the United States on 15 November 2018

**Verified Purchase**

I recently purchased this entry level DSLR to learn more about photography and to improve my skills and below is my full review.

Pros:

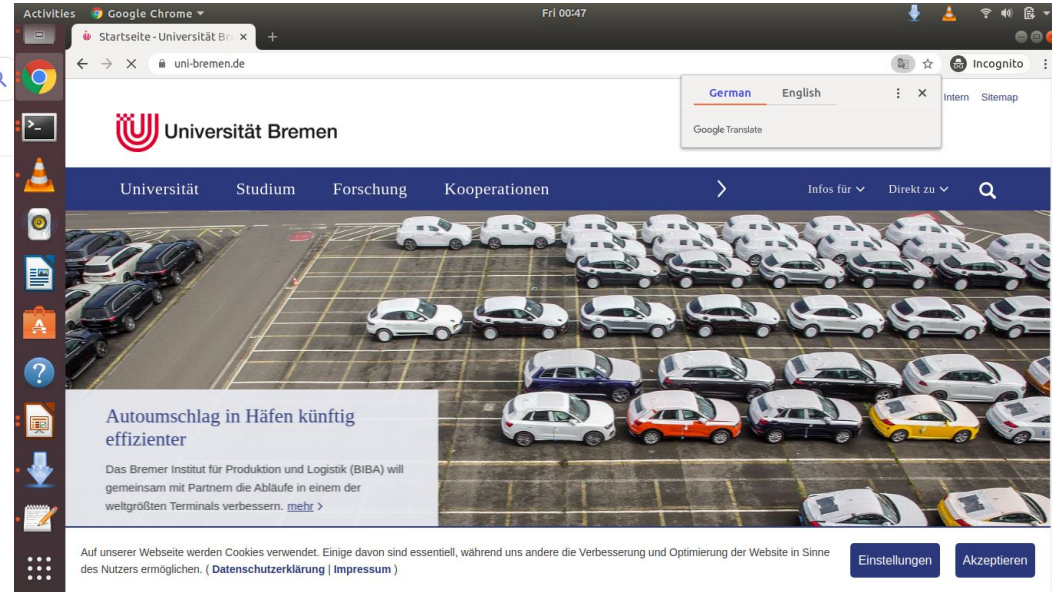
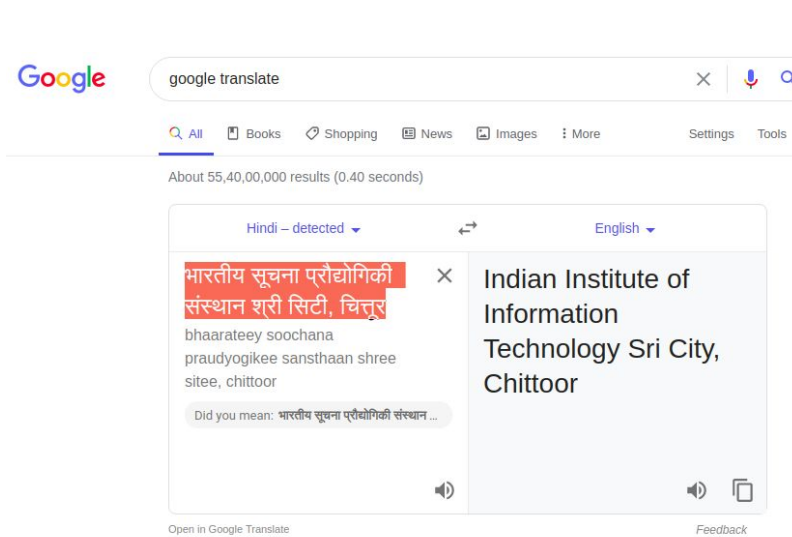
1. Image Quality: The 24.2MP sensor provides sharp and crispier images compared to other entry level models
2. Touch screen Interface: The touch screen interface is a nice to have in a camera and toggling through images is seamless and easy. The large touch screen really helps to adjust Aperture and shutter settings and the flip out feature is cool as well.
3. SnapBridge: Transfer of files from the Camera to the device is seamless through SnapBridge although I haven't tried using Snap Bridge for remote capture.

Cons:

1. Image Stabilization: There is no image stabilization in this camera. There is a possibility for the images to shake a bit here and there. This was a problem for me when I tried to shoot in low light during sunset and sunrise.
2. No Autofocus Motor: This camera doesn't have the autofocus motor in it's body which means that I need an AF-S

[Read more](#)

# Introduction: Machine Translation



# Introduction: Alexa

“Alexa, What is my  
Credit Card Balance “

“Alexa, turn on  
welcome home“

“Alexa, my last 5  
transaction“



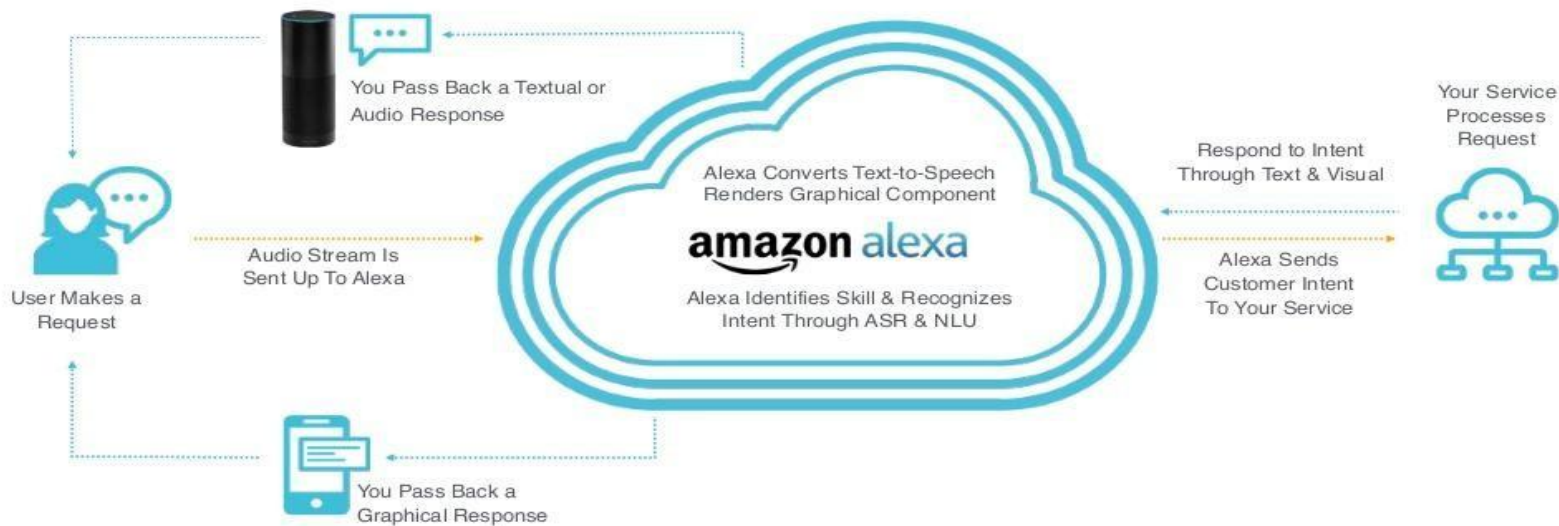
“Alexa, nearest bank  
to me“

“Alexa, turn on  
the TV“

# Introduction: Alexa

## UNDER THE HOOD OF ASK

A closer look at how the Alexa Skills Kit process a request and returns an appropriate response





# Introduction: Text-to-Speech

- **Amazon Polly**

- Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk, and build entirely new categories of speech-enabled products.
- Amazon Polly is a Text-to-Speech service that uses advanced deep learning technologies to synthesize speech that sounds like a human voice.

# Introduction: Text-to-Speech



# **Introduction: Speech Recognition**

Google Speech Recognition

**NLP tools in Google Docs ?**

# Introduction: Interesting tools



[https://www.youtube.com/watch?v=oFUlm\\_o453E](https://www.youtube.com/watch?v=oFUlm_o453E)

# Introduction: ELIZA

Human: Men are all alike.

ELIZA: In what way?

Human: They're always bugging us about something.

ELIZA: Can you think of a special example?

Human: Well, my boyfriend made me come here.

ELIZA: Your boyfriend made you come here?

Human: He says I'm depressed much of the time.

ELIZA: I'm sorry to hear that you are depressed much of the time.

Human: It's true. I'm unhappy.

ELIZA: Do you think coming here will help you not to be unhappy?

Human: I need some help, that much seems certain.

# Introduction:

**Text summarization**

# Introduction: Language Understanding



Where is **A Bug's Life** playing in **Mountain View**?

A Bug's Life is playing at the Century 16 Theater.

When is **it** playing **there**?

It's playing at 2pm, 5pm, and 8pm.

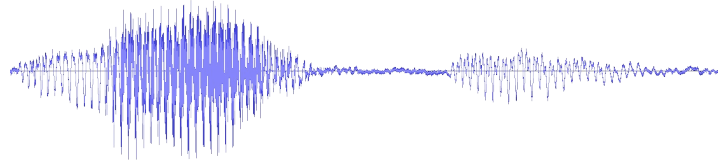
OK. I'd like 1 **adult** and 2 **children** for **the first show**.  
How much would **that** cost?



But we need **domain knowledge**, **discourse knowledge**,  
**world knowledge**, linguistic knowledge.

# Introduction: Language Understanding

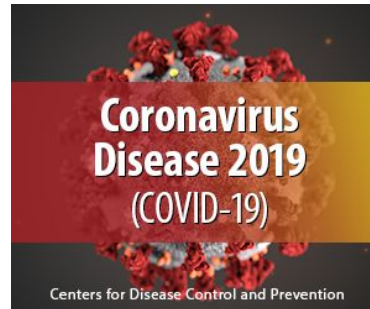
- Sound



- Gesture



- Text





# Introduction: Language Understanding

car



tiger



laptop



doctor



# **Introduction:** Language Understanding

- **Speaker**
  - Intention (goals, shared knowledge and beliefs)
  - Generation (tactical)
  - Synthesis (text or speech)
- **Listener**
  - Perception
  - Interpretation (syntactic, semantic, pragmatic)
  - Incorporation (internalization, understanding)
- **Both**
  - Context (grounding)

A movie poster for 'The Imitation Game'. It features a close-up of Benedict Cumberbatch as Alan Turing on the left, looking directly at the camera with a serious expression. To his right, Keira Knightley as Joan Clarke is also looking forward, her face partially in shadow. The background is a dark, textured wall covered in a grid of circular, red, mechanical-looking components, possibly part of a code-breaking machine. The title 'THE IMITATION GAME' is written in large, bold, yellow capital letters at the bottom right.

# THE IMITATION GAME

# Introduction

- Who?

- Alan Turing

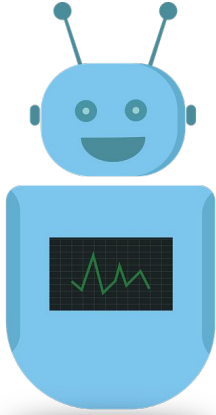
- Why?

- Can A Computer talk like a Human?

# Introduction: The Imitation Game



# Introduction: The Imitation Game



**Respondent 1: Bot**

## Turing Test



**Respondent 2: Human**



**Questioner**

# Introduction: Turing Test

- Turing Test requires three terminals, each of which is physically separated from the other two. One terminal is operated by a computer, while the other two are operated by humans.
- During the test, one of the humans functions as the questioner, while the second human and the computer function as respondents.
- The questioner interrogates the respondents within a specific subject area, using a specified format and context.
- After a preset length of time or number of questions, the questioner is then asked to decide which respondent was human and which was a computer.

# Introduction: ELIZA

Human: Men are all alike.

ELIZA: In what way?

Human: They're always bugging us about something.

ELIZA: Can you think of a special example?

Human: Well, my boyfriend made me come here.

ELIZA: Your boyfriend made you come here?

Human: He says I'm depressed much of the time.

ELIZA: I'm sorry to hear that you are depressed much of the time.

Human: It's true. I'm unhappy.

ELIZA: Do you think coming here will help you not to be unhappy?

Human: I need some help, that much seems certain.



# Introduction: DIALOG SYSTEMS

Microsoft Xiaolce



在日本，仅仅16个多月

# Introduction: Applications

- Machine Translation
- Information Retrieval
- Question Answering
- Dialogue Systems
- Text-to-speech
- Speech Recognition
- Information Extraction
- Sentiment Analysis

.....

# Introduction: Some Links

[WATSON in Hospital](#)