CS565: Intelligent Systems and Interfaces

General Introduction

Semester: Jan – May 2019

Ashish Anand
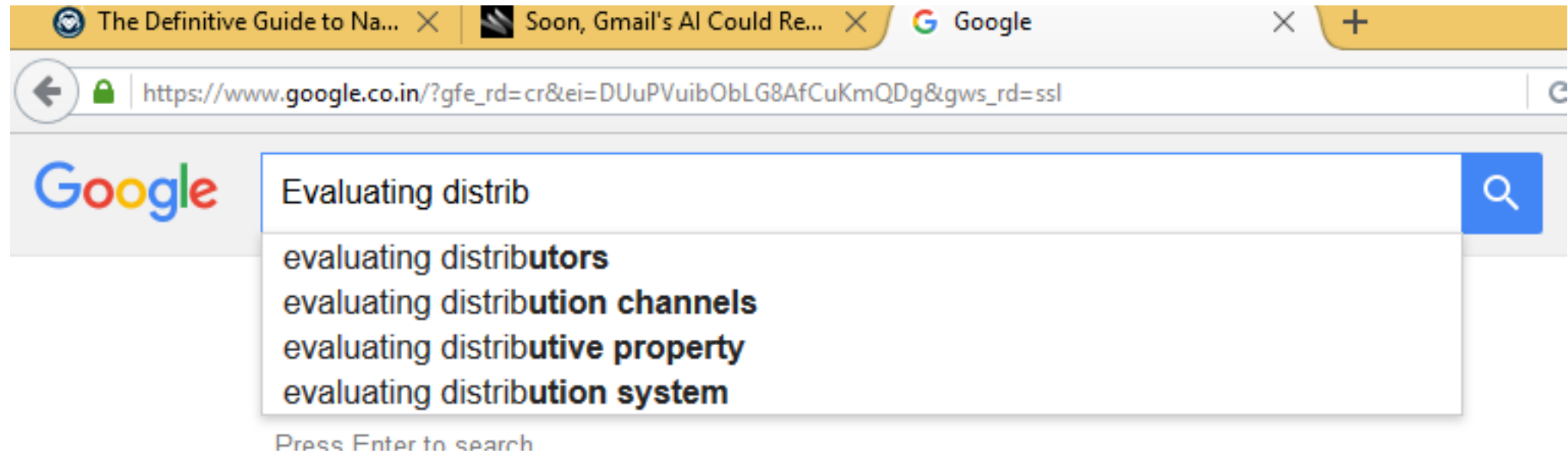
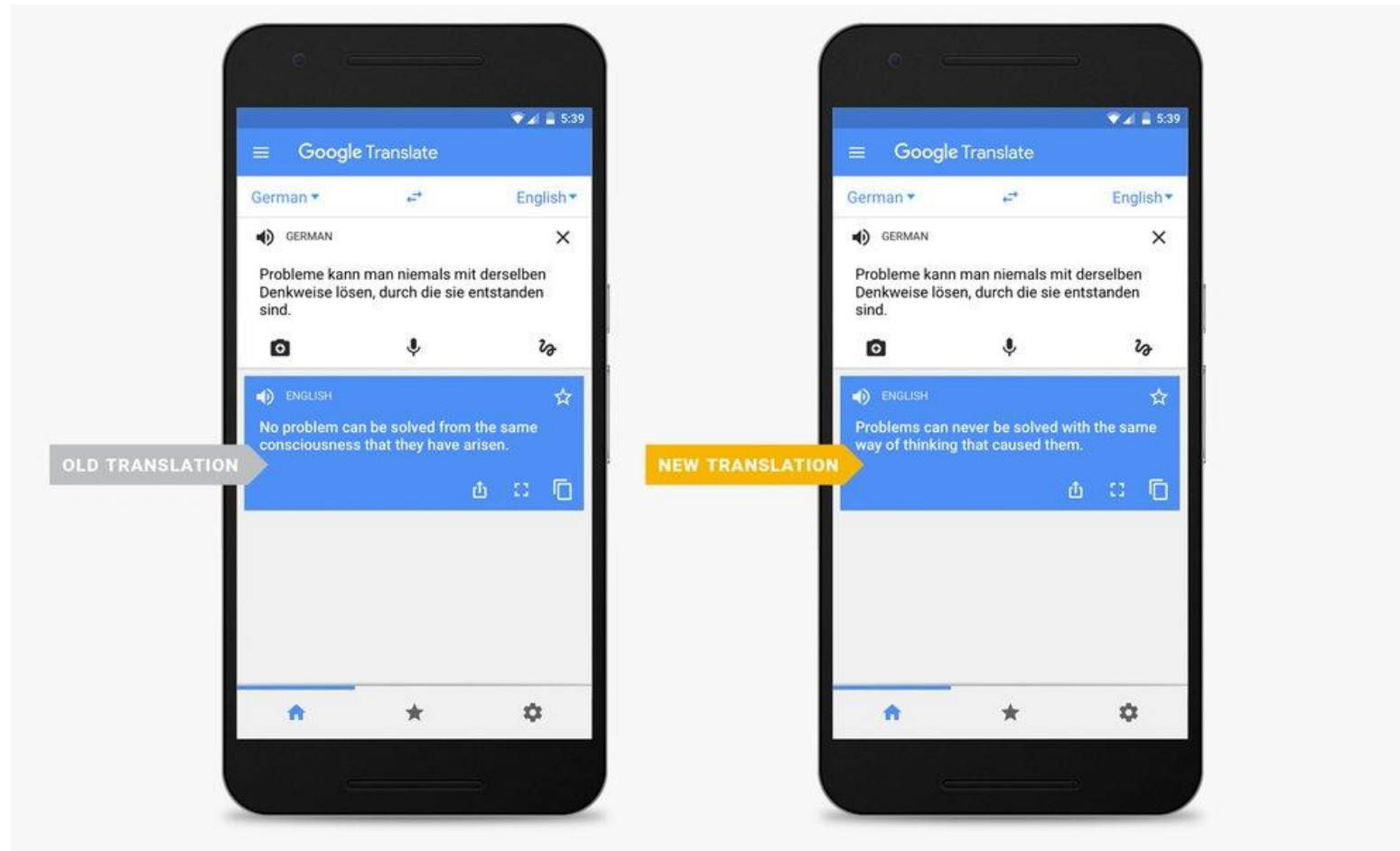Associate Professor, Dept of CSE

IIT Guwahati

What we are going to study in this course?

- Fundamentals and specific NLP Tasks
  - Major NLP Topics: Language Models; Sequence Tagging; Vector Space Models; Parsing; Few Applications – Fine-grained NER, RE;

- Statistical and Machine Learning models to solve them
  - Major Models: Markov Model: HMM; Log-linear Models: MEMM, CRF; Neural Network Models; Representation Learning
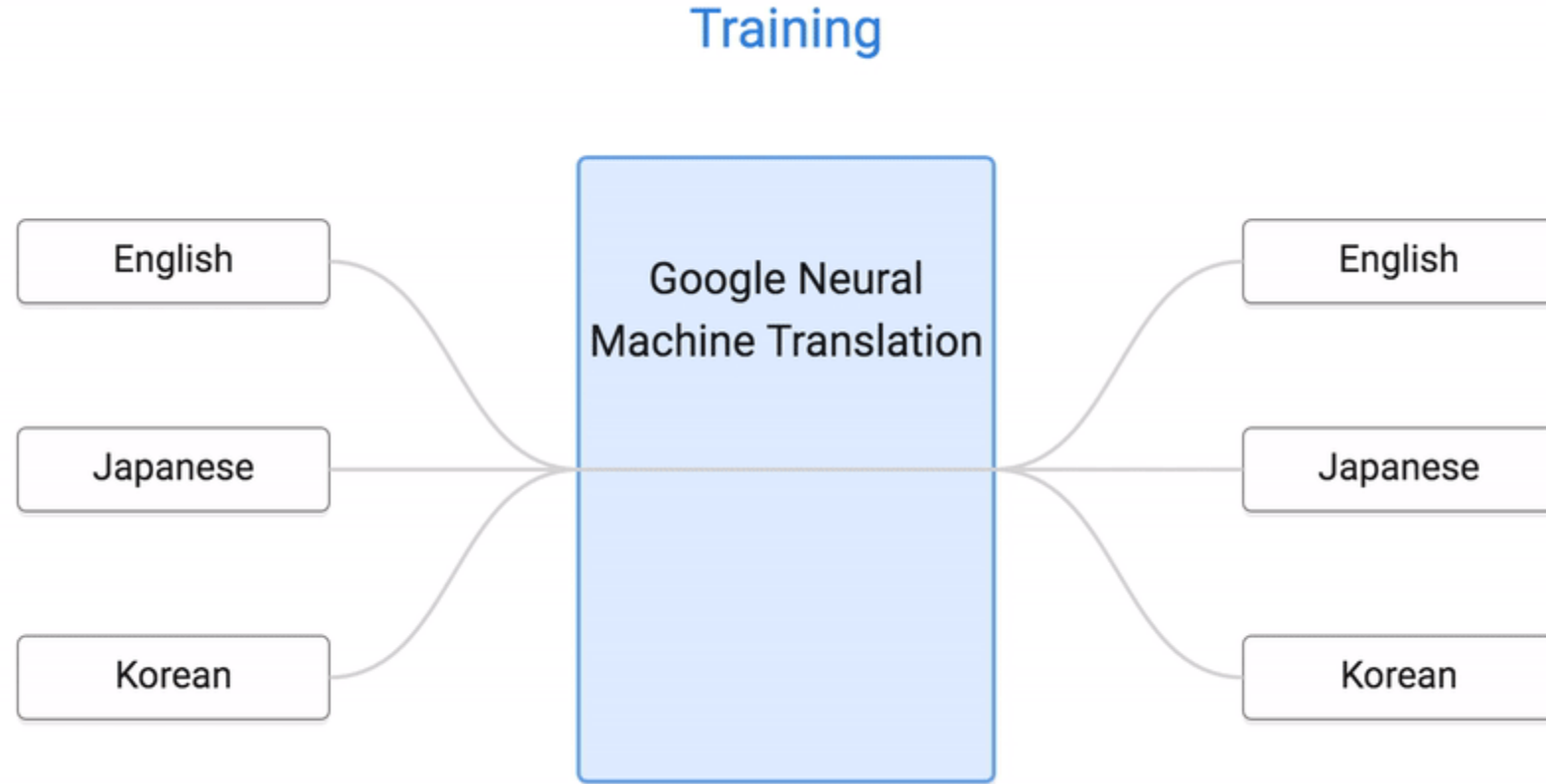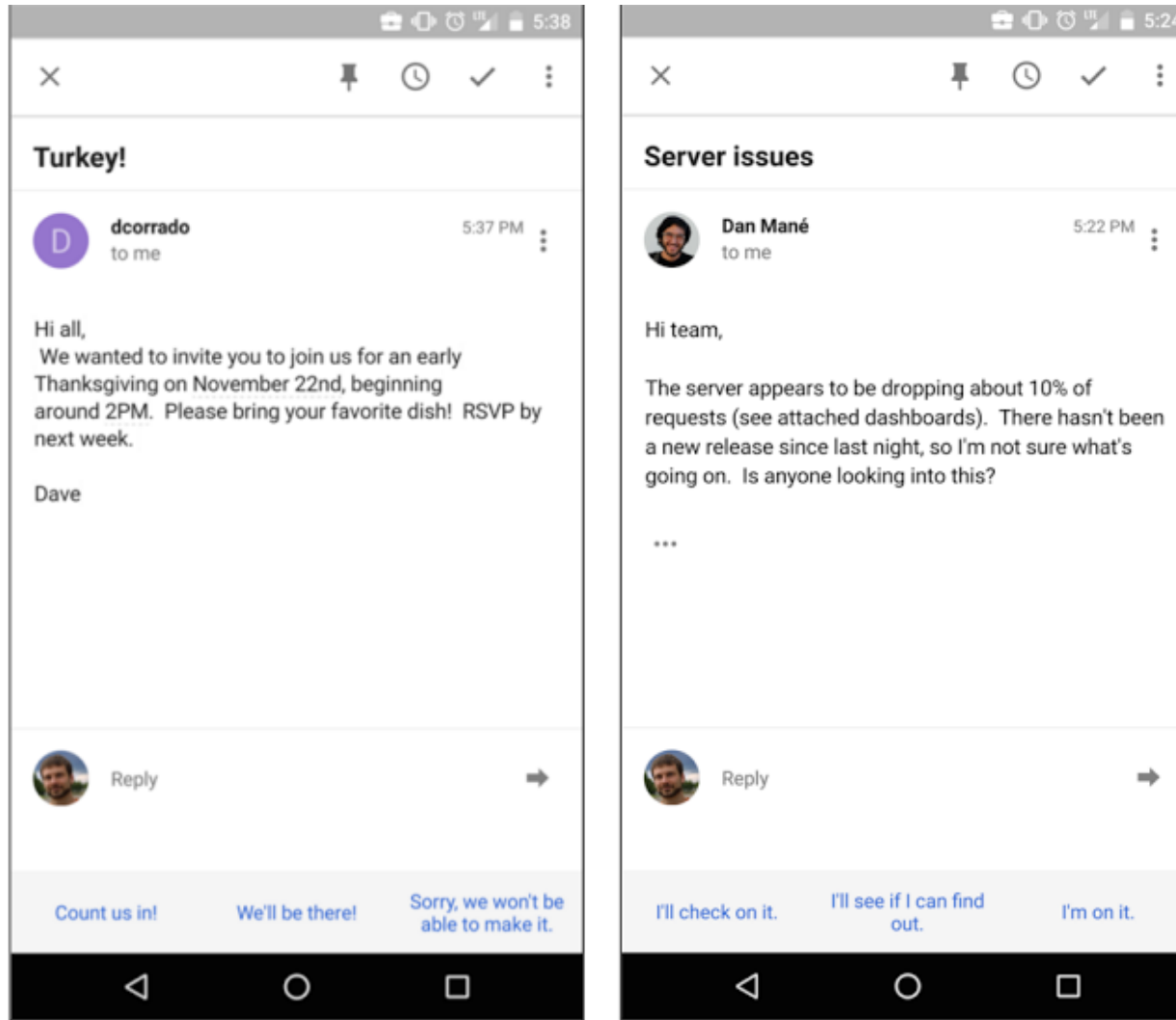
Do we see NLP in our daily life?

Machine Translation

Machine Translation

Source: https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Suggestions for short email response

Some Applications

- Machine Translation
- Automatic Summarization
- Question-Answering System
- Paraphrase detection
- Sentiment Analysis
- Text Classification
- Conversational Agents
- Many many more …..

Motivation to study the course

Lets discuss mine; you show your motivation in Project

Biomedical Text Mining: Semantic graph construction

HISTORY OF PRESENT ILLNESS :

The patient is a 58 year old right hand dominant white male with a long history of `hypertension` , changed his medications from `Aldomet` to `Clonidine` six weeks ago . The patient has a history of `adult onset diabetes mellitus` , `ankylosing spondylitis` , status post myocardial infarction in '96 (?) now with acute onset of left face and arm greater than leg hemiplegia and primary hemisensory loss on the left . His `voice became slurred` and he had a `mild central dull headache` .He was unable to move the left side of his body and felt numb on that side .

He was taken to Wayskemedcalltown Talmi and transferred to Heaonboburg Linpack Grant Medical Center with a `computerized tomography scan` showing a `1x2 thalamic capsular` `hemorrhage` without superficial mass effect .

MEDICATIONS ON ADMISSION :

`Vasotec` `40 mg q.day` , `Soma` `1 tablet q.day` , `Demerolprn` , `Clonidine` .

ALLERGIES :

The patient has no known `drug allergies` .

FAMILY HISTORY : positive for `diabetes mellitus` , positive for `cancer` .
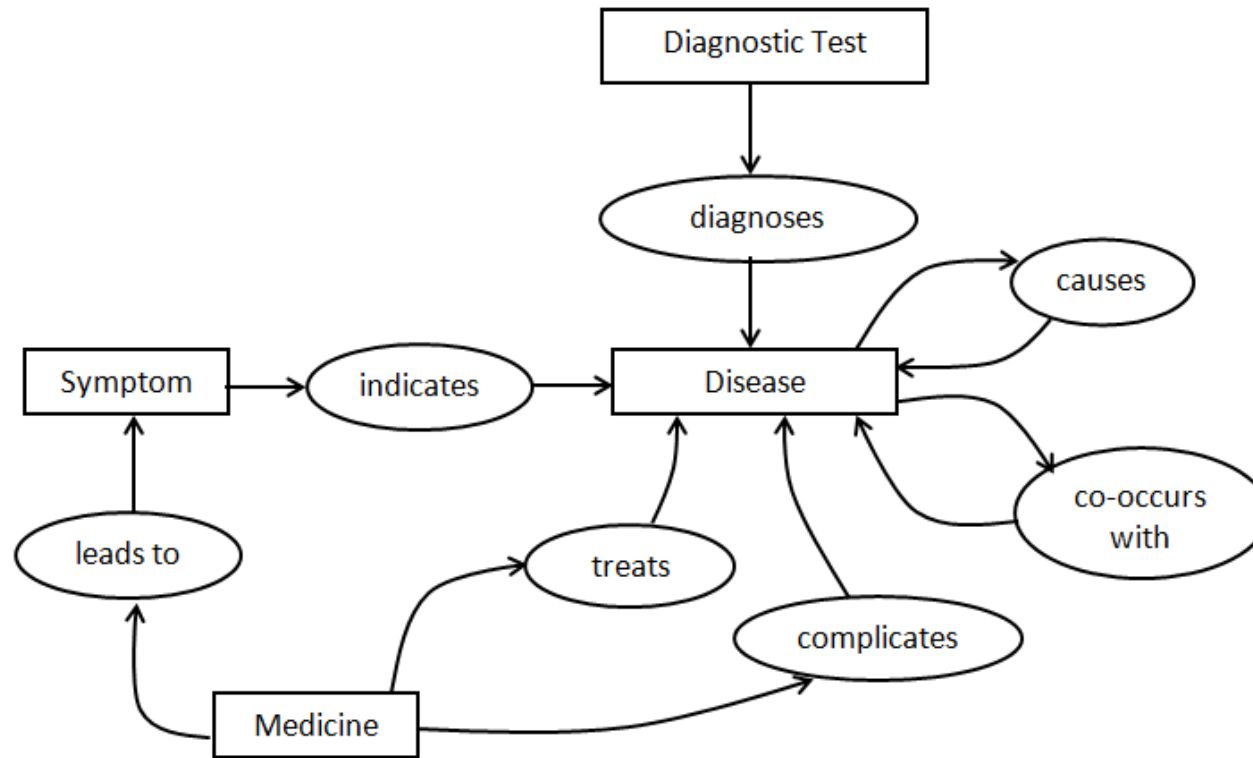
■ **DISEASE**   ■ **MEDICINE & DOSAGE**

■ **DIAGNOSTIC TEST & RESULTS**   ■ **SYMPTOMS**

Semantic graph construction

Semantic graphs in general domain

Structured graph

Unstructured text
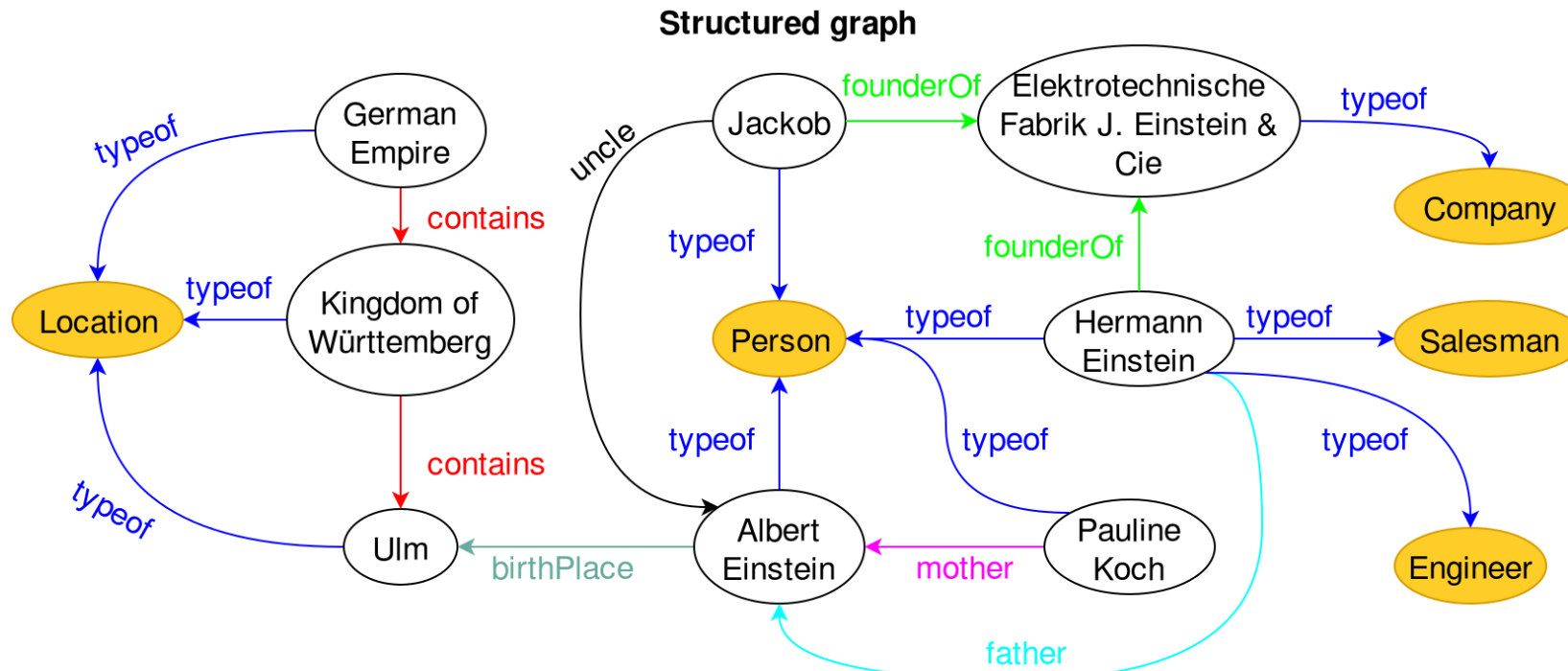
Albert Einstein was born in Ulm, in the Kingdom of Württemberg in the German Empire on 14 March 1879. His parents were Hermann Einstein, a salesman and engineer, and Pauline Koch. In 1880, the family moved to Munich, where Einstein's father and his uncle Jakob founded Elektrotechnische Fabrik J. Einstein & Cie, a company that manufactured electrical equipment based on direct current.

Figure courtesy: Abhishek, Research Scholar, Dept of CSE

Administrative Stuff

Schedule and Marks Distribution

- Class Schedule
  - Mon [5-6pm], Tue [4-5pm], Wed [3-4pm]: A1 Slot
  - Keep your Thursday A1 slot [2-3pm] Free

- Marks Distribution
  - Assignments [Two Assignments expected] [~10%]
  - Scribe [~10%]
  - Project [~40%]
  - Exams [~15+25=40%]

TAs

Abhishek   Aparajita   Akshay   Hema Chandra Kumar   Saptarshi

Assignments and Project

- 2 Assignments within 10th – 15th Feb
- A group of 4 students
- Decide and submit group details by 20th Jan
- Same group for assignment and project
- In each submission of assignment, contribution of individual students need to be explicitly mentioned
- Certain project themes will be posted around 20th – 26th Jan
- Project topic proposal submission deadline [tentative]: 22nd Feb
- Project mid-term evaluation and report submission [tentative]: 15th – 20th Mar

Resources: Books

- No Text Book

- Reference Books
  - Speech and Language Processing: Jurafsky and Martin [https://web.stanford.edu/~jurafsky/slp3/]
  - Foundations of Statistical Natural Language Processing: Manning and Schutze [http://nlp.stanford.edu/fsnlp/]
  - Neural Networks and Learning Machines: Haykin

Resources: Online courses

- http://web.stanford.edu/class/cs224n/
- http://cs224d.stanford.edu/syllabus.html [Deep Learning for NLP]
- http://web.stanford.edu/class/cs224u/ [Natural language Understanding]
- http://www.cs.columbia.edu/~mcollins/notes-spring2013.html
- http://www.cs.columbia.edu/~cs4705/
- http://www.cse.iitd.ac.in/~mausam/courses/col772/spring2018/

Resources: Online relevant materials

- NLTK basic : http://www.nltk.org/book/

- Python Numpy Tutorial : http://cs231n.github.io/python-numpy-tutorial/

- Deep Learning Tools/Library such as PyTorch, Tensor-Flow, Theano etc.

- Many tools: PyText, AllenNLP, CoreNLP, Apache OpenNLP, SpaCy

- Very soon, course website/Canvas/Piazza will be online.