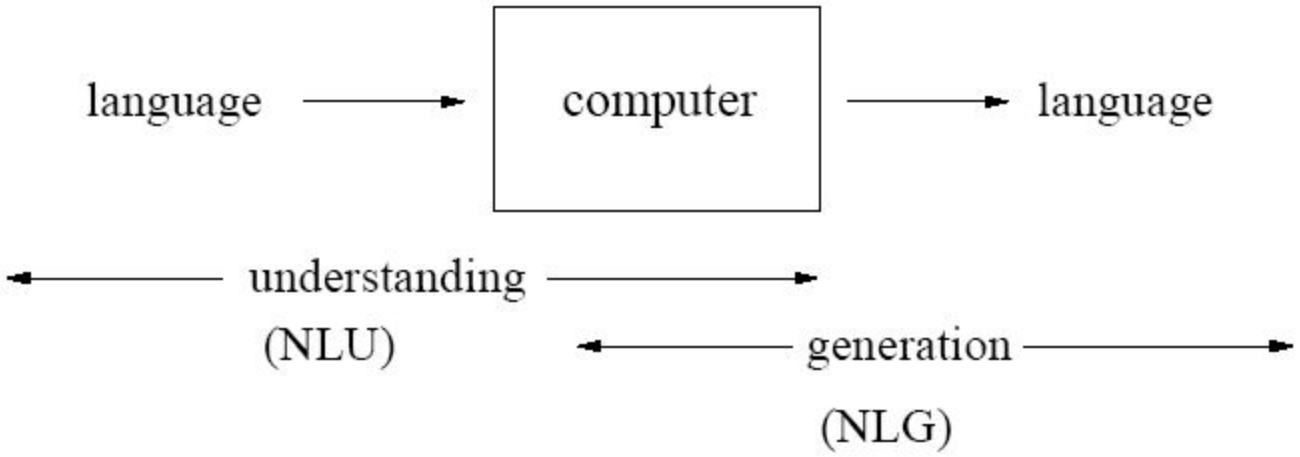
Natural language processing (NLP)

Python for AI

What is NLP?

- Natural Language Processing (NLP) is a field in Artificial Intelligence (AI) devoted to creating computers that use natural language as input and/or output.



Why NLP?

Modern NLP: NLP in the Days of Big Data

Three trends:

1. An **enormous amount of information** is now available in machine readable form as natural language text (newspapers, web pages, medical records, financial filings, product reviews, discussion forums, etc.)
2. Conversational agents are becoming an important form of human-computer **communication**
3. Much of human-human interaction is now mediated by computers via **social media**

NLP Applications

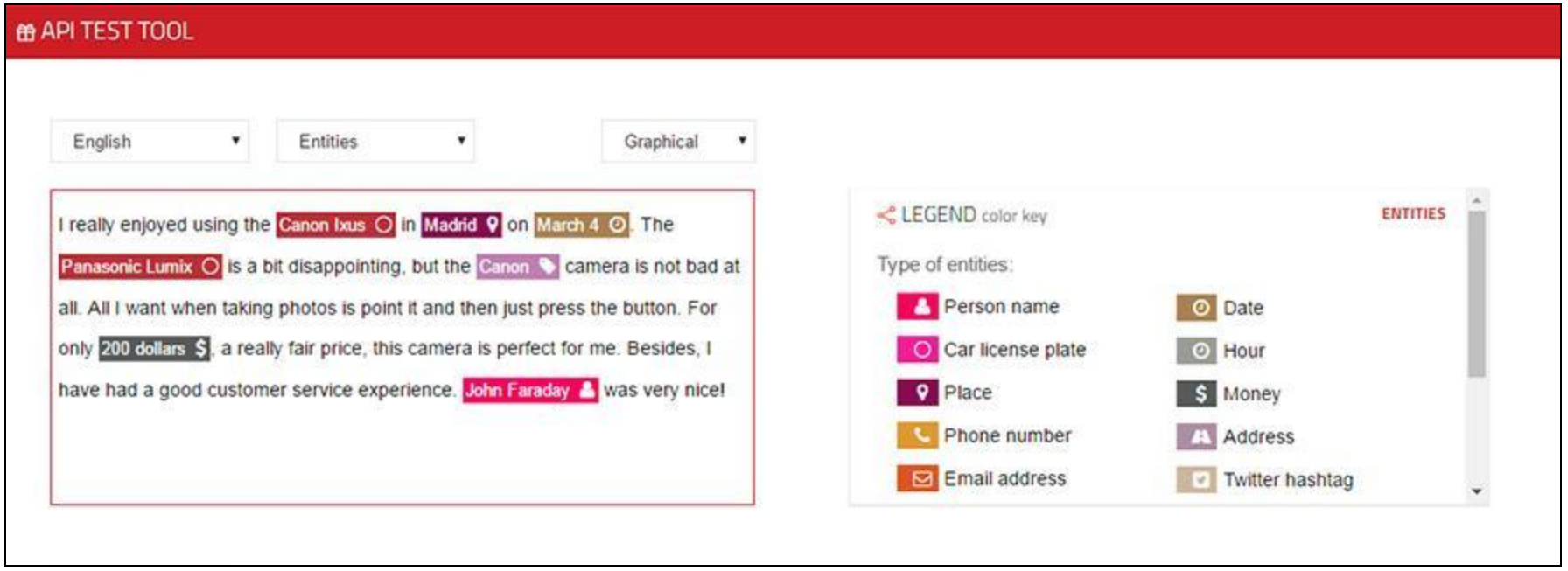
- Three prominent application areas:
  - Text analytics/mining (from *"**unstructured data**")*
    - Sentiment analysis
    - Topic identification
    - Digital Humanities (*"new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing."*)
  - Conversational agents
    - Siri, Cortana, Amazon Alexa, Google Assistant
    - Chatbots
  - Machine translation

Text Analytics

- Data-mining of weblogs, microblogs, discussion forums, user reviews, and other forms of user-generated media.

Text Analytics (cont.)

- Typically this involves the extraction of **limited** kinds of semantic and pragmatic information from texts
  - Entity mentions
  - Concept identification
  - Sentiment

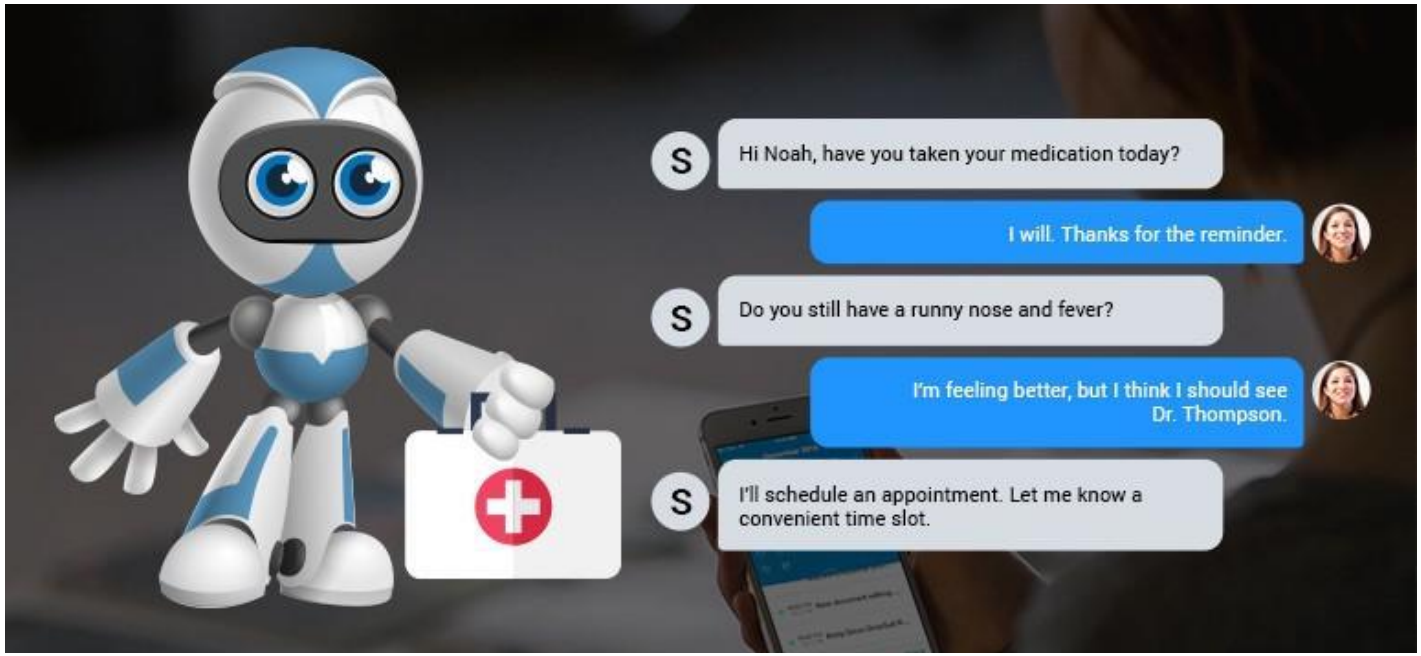


Demo

- Sentiment Analysis with Python NLTK Text Classification
  - http://text-processing.com/demo/sentiment/
- Tweet Sentiment Visualization Tool
  - https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
- Concept Extraction
  - http://aylien.com/concept-extraction/

Conversational Agents

- Combine
  - Speech recognition/synthesis
  - Question answering
    - From the web and from structured information sources (freebase, dbpedia, yago, etc.)
  - Simple agent-like abilities
    - Create/edit calendar entries
    - Reminders
    - Directions
    - Invoking/interacting with other apps


S
Hi Noah, have you taken your medication today?
I will. Thanks for the reminder.
S
Do you still have a runny nose and fever?
I'm feeling better, but I think I should see Dr. Thompson.
S
I'll schedule an appointment. Let me know a convenient time slot.


M
amazon
amazon
amazon
AMAZON ALEXA

Mitsuku

Question Answering

- Traditional *information retrieval* provides documents/resources that provide users with what they need to satisfy their information needs.
- *Question answering* on the other hand directly provides an answer to information needs posed as questions.

IBM Watson

https://www.youtube.com/watch?v=WFR3lOm_xhE

Machine Translation

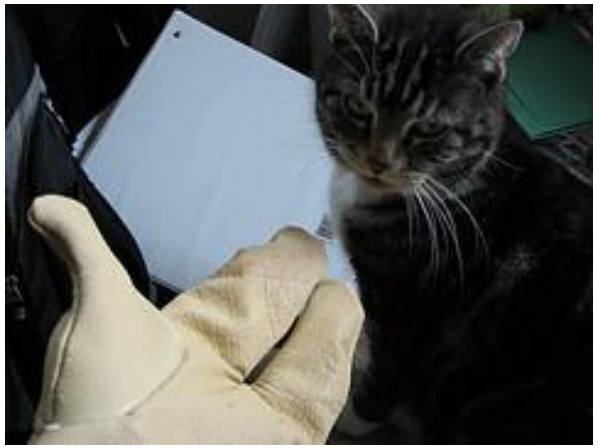
- The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.
- In the past 15 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

But NLP very is hard..

- Understanding natural languages is hard … because of inherent *ambiguity*
- Engineering NLP systems is also hard … because of:
  - Huge amount of data resources needed (e.g. grammar, dictionary, documents to extract statistics from)
  - Computational complexity (intractable) of analyzing a sentence

Ambiguity (1)

“Get the cat with the gloves.”



Source: Marti Hearst, i256, at UC Berkeley

Ambiguity (2)

Find at least 5 meanings of this sentence:

"I made her duck"

1. I cooked waterfowl for her benefit (to eat)
2. I cooked waterfowl belonging to her
3. I created the (plaster?) duck she owns
4. I caused her to quickly lower her head or body
5. I waved my magic wand and turned her into undifferentiated waterfowl

Ambiguity is Pervasive

- **Phonetics**
  - I mate or duck
  - I'm eight or duck
  - Eye maid; her duck
  - Aye mate, her duck
  - I maid her duck
  - I'm aid her duck
  - I mate her duck
  - I'm ate her duck
  - I'm ate or duck
  - I mate or duck

Sound like
"I made her duck"

Source: Jurafsky & Martin "Speech and Language Processing"

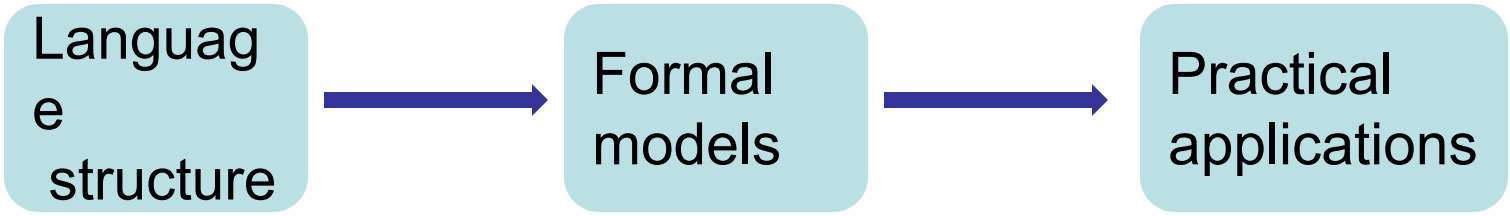
The Bottom Line

- Complete NL Understanding (thus general intelligence) is impossible.
- But we can make incremental progress.
- Also we have made successes in **limited domains**.

The Big Picture Approach

All of these applications operate by **exploiting underlying regularities** in human languages.

Sometimes in complex ways, sometimes in pretty trivial ways.



Topics: Linguistics

- Word-level processing
- Syntactic processing
- Lexical and compositional semantics
- Discourse structure