Introduction to NLP

Lecture-1

Outline

- What is NLP
  - What we do in NLP
  - Goal of NLP
- Applications
- Why NLP is challenging?

NLP

- [Wiki:] **Natural language processing(NLP)** is a field of computer science, artificial intelligence, and computational linguistics **concerned with the interactions between computers and human (natural) languages.**
- Field that covers computer understanding and manipulation of human lan guage
- A normal programming language treats text as a sequence of characters
- By analyzing language for its meaning, NLP systems have long filled useful roles, such as correcting grammar, converting speech to text and automatically translating between languages.”

NLP: Goal

- Ultimate goal: **Natural human-to-computer communication**
  - Very long way to go
- **Engineering goal:** Build sophisticated applications based on the useful linguistic patterns
  - More practical and we are developing in this way
- Today’s NLP models are not perfect
- Still you can find useful applications with acceptable errors
  - Chatbots (e-commerce sites), Google translator

Why NLP is important

- This is the age of data driven application
- Majority of this data is available in the form of Text
  - Social media posts, Wikipedia articles, News, stock market, research articles, clinical reports,...
- Mining these data for useful applications
  - Knowledge of the language is essential
- NLP deals with developing applications to process/understand these textual data automatically

Levels of NLP

- Phonology → deals with sounds
- Morphology → deals with smallest parts of words that carry meaning
- Lexical → deals with formation of words
- Syntactic → grammar and structure of sentences
- Semantic → meaning of sentence
- Discourse → deals with the structure of different kinds of text.
- Pragmatic → deals with the knowledge that comes from the outside world (common sense)

Why NLP is “Hard”?

- Human languages are messy, ambiguous, and ever-changing

Ambiguity in Lexical level

- **Examples**
- Will, will Will will Will Will's will?
- Rose rose to put rose roes on her rows of roses.
- He saw a saw mill

Why NLP is “Hard”?

- Human languages are messy, ambiguous, and ever-changing

Ambiguity in Lexical level

- **Examples**
- Will, will Will will Will Will's will?
  - Will (a person), will (future tense helping verb) Will (a second person) will (bequeath) [to] Will (a third person) Will's (the second person) will (a document)? (Someone asked Will 1 directly if Will 2 plans to bequeath his own will, the document, to Will 3.)
- Rose rose to put rose roes on her rows of roses.
  - Rose [a person] rose [stood] to put rose [pink-colored] roes [fish eggs as fertilizer] on her rows of roses [flower].
- He saw a saw mill

Why NLP is “Hard”?

Ambiguity in syntax level (structural ambiguity)

- The boy saw a man with telescope
- While the man was hunting the deer ran through the forest
- One morning I shot an elephant in my pajamas.

Why NLP is “hard”?

Idioms are challenging

the old girl finally kicked the bucket

×

Why NLP is “Hard”?

Languages are Imprecise and Vague

- It is very *hot* here (hot depends on the location)
- Can the dog swim? (yes/no question)
- Can you please post this letter? (request, not yes or no question)

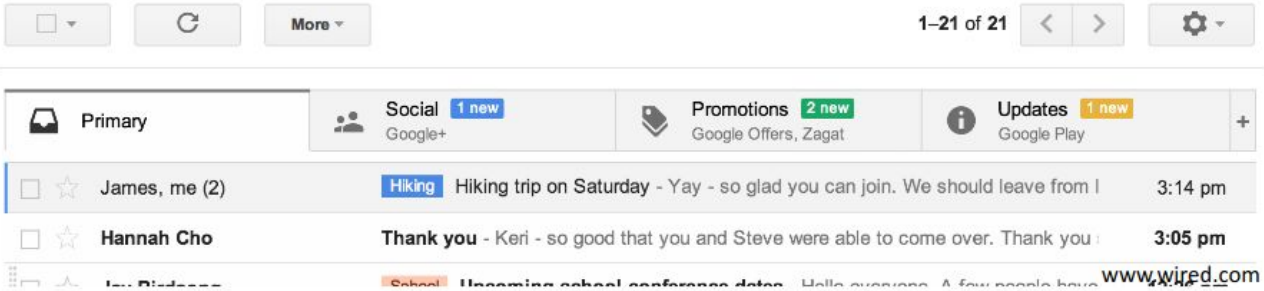
These ambiguity are actually the beauty of language… we cannot compromise for practicability

These challenges make NLP interesting

Many of the problems are open…

Still, we have good applications, that satisfies at least our engineering goal

Applications: Text Classification



www.wired.com

Google Translate BETA

Text and Web | Translated Search | Dictionary | Tools

Translate Text

Original text:

Istotą instytucji wyłączenia organu podatkowego od załatwienia sprawy dotyczącej zobowiązania podatkowego lub innej sprawy normowanej przepisami prawa podatkowego jest utrata właściwości danego organu do

Detect language » Finnish Translate

Translation: Polish (automatically detected) » Finnish

Pelkät vapautusta veron käsittelevälle viranomaiselle tapauksissa, joissa verovelan tai muita aineita, normowanej vero-oikeuden menetys kiinteis kyseisen viranomaisen ratkaist erityinen veronmaksajille.

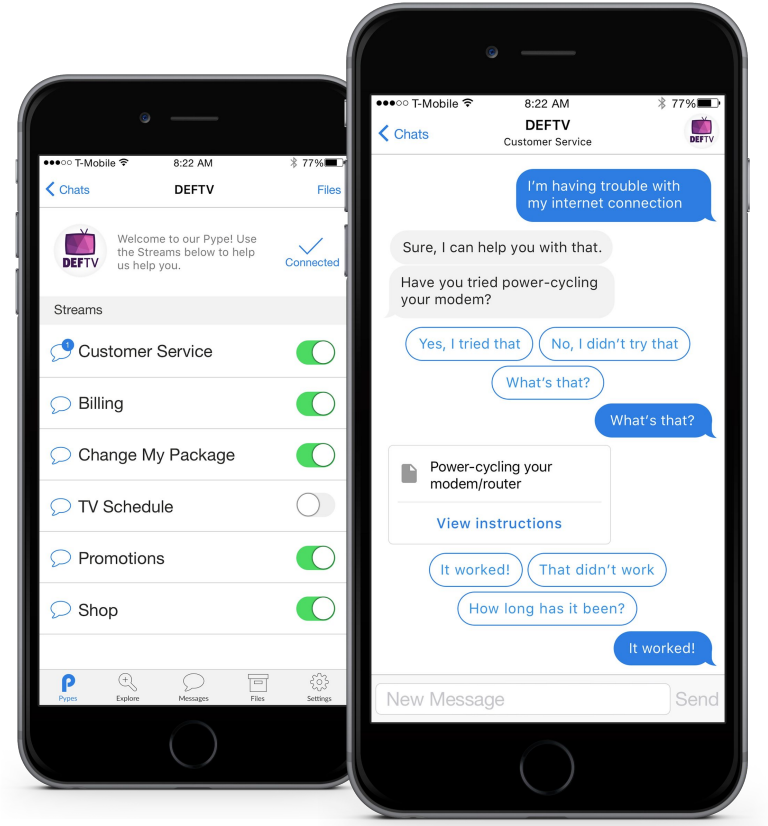
Suggest a better t

Application: Information Extraction

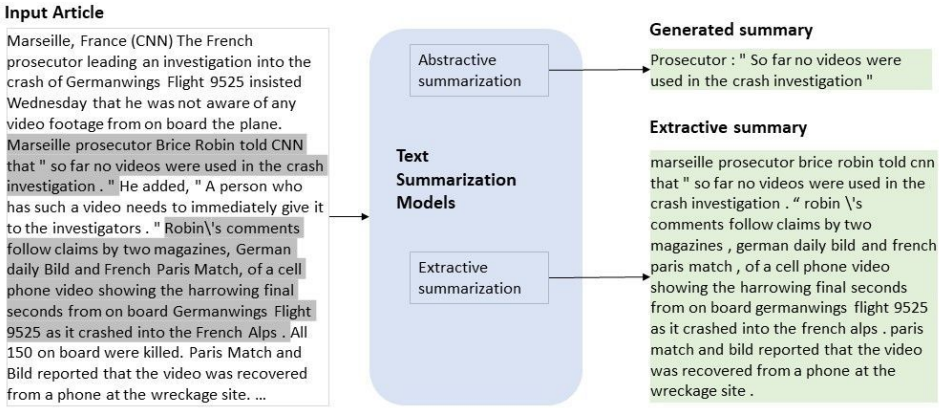
New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

Chatbots



Text summarization



Text Generation

Reading Assignment

- Natural language processing: an introduction
- See this simple introduction to Natural Language Processing (NLP)