Amirkabir University of Technology (Tehran Polytechnic)

Natural Language Processing

Lecture 1: Introduction

Amirkabir University of Technology

Dr Momtazi

Outline

- **Administrative Information**

- Introduction to the Course

- Overview of the Semester

Course Home Page

- SNLP: Statistical Natural Language Processing
  - Administrative information
  - Slides
  - Exercises

Assessment

- Regular attendance in the class

- Exercises (15%)

- Final project or seminar (15%)

- Midterm exam (30%)

- Final exam (40%)

Contact

- Email: momtazi@aut.ac.ir

- Phone: 021-64542737

Textbook

Speech and Language Processing (3rd ed. draft)

An Introduction to Natural Language Processing,
Computational Linguistics, and Speech Recognition

Dan Jurafsky and James H. Martin

Other literature

- Foundations of Statistical Natural Language Processing
  - by Christopher D. Manning, Hinrich Schütze

- Natural Language Processing with Python
  - by Edward Loper, Ewan Klein, and Steven Bird

- Natural Language Understanding
  - by James Allen

Relevance Journal & Conferences

- Journal
  - CL: Computational Linguistics
  - TACL: Transaction of Association for Computational Linguistics

- Conferences
  - ACL: Association for Computational Linguistics
  - NAACL: North American Chapter
  - EACL: European Chapter
  - HLT: Human Language Technology
  - EMNLP: Empirical Methods on Natural Language Processing
  - CoLing: Computational Linguistics
  - LREC: Language Resources and Evaluation

Outline

- Administrative Information
- **Introduction to the Course**
- Overview of the Semester

Different Types of Languages

- Natural languages
  - English
  - Persian
  - French
  - ...

- Formal languages
  - Java
  - Python
  - LaTeX
  - ...

- Descriptive languages
  - Biology: DNA
  - Chemistry: chemical formulas

Natural Language

- A vocabulary consists of a set of words (w_i)

- A text is composed of a sequence of words from the vocabulary

- A language is constructed of a set of all possible texts

Natural Language

- Examples of Vocabularies
  - English
  - the
  - and
  - eat
  - you
  - book
  - …

Natural Language Processing

- Applications
  - Text Technologies
  - Speech Technologies

- Techniques

Outline

- Administrative Information

- **Introduction to the Course**
  - **NLP Applications**
  - NLP Techniques

- Overview of the Semester

Spell and Grammar Checking

- Checking the spelling and the grammar of a text, and suggesting correct alternatives for the errors
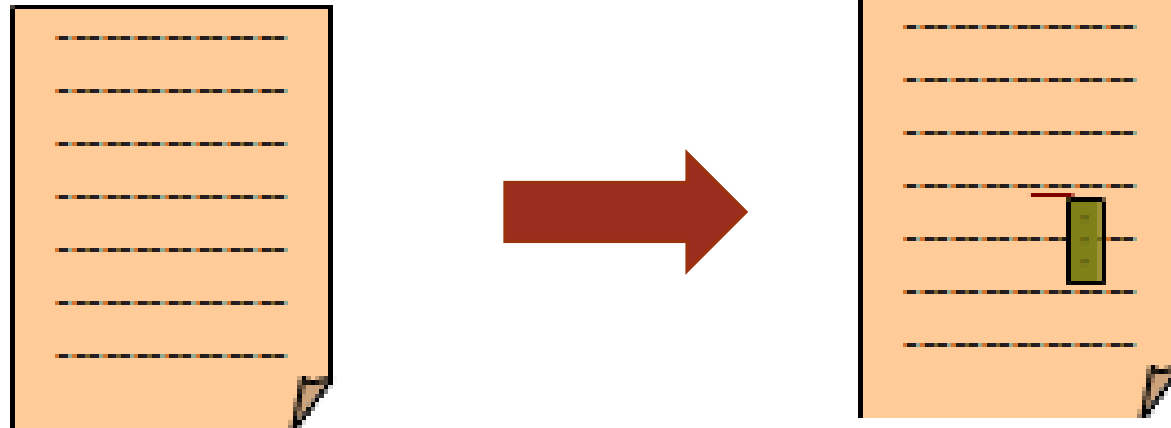
Spell and Grammar Checking
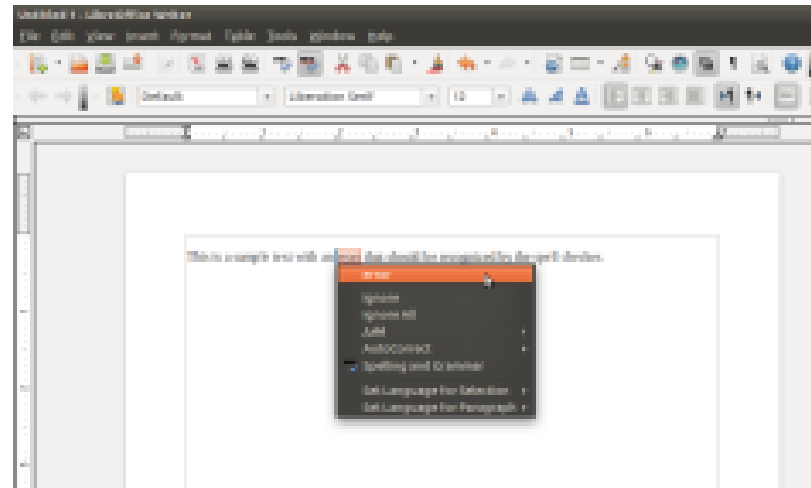
Text Categorization

- Assigning each text to a category
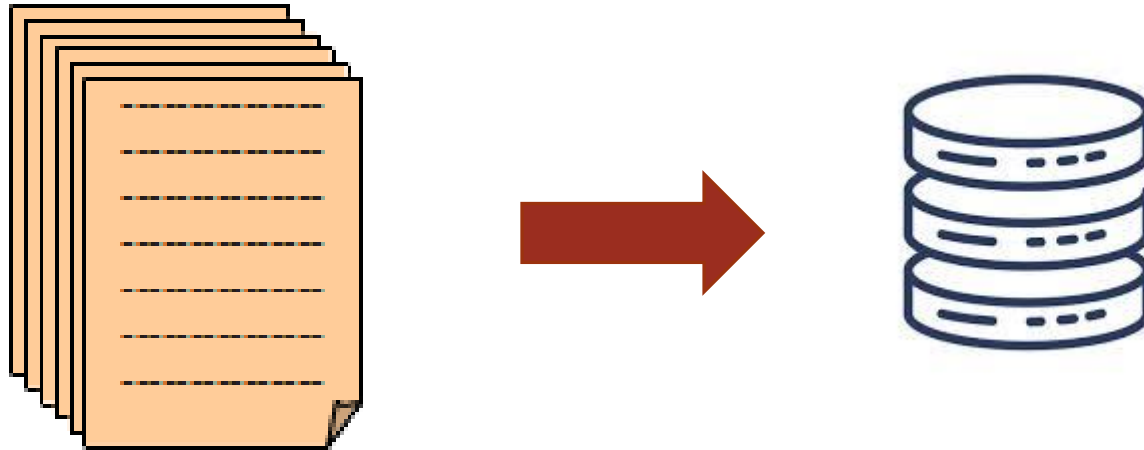
Text Categorization
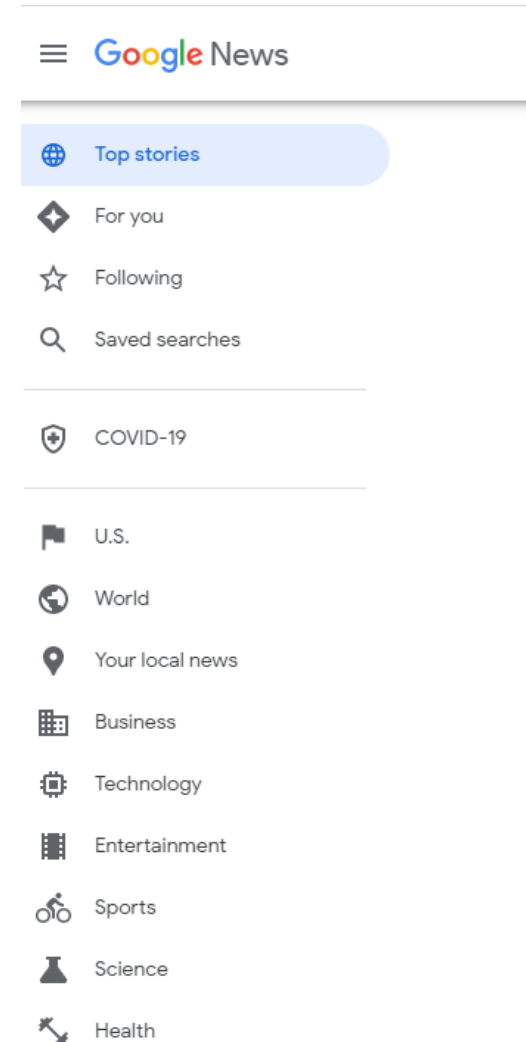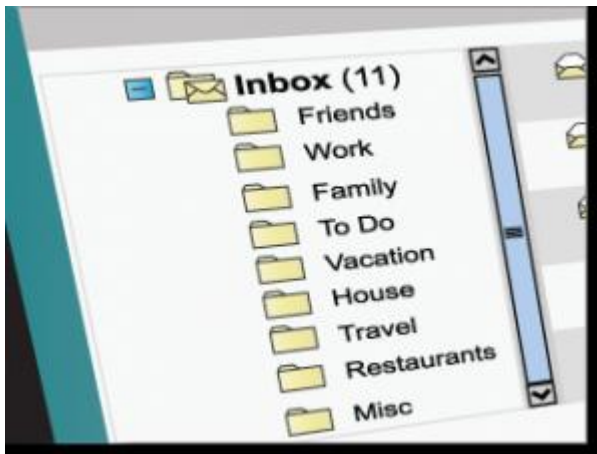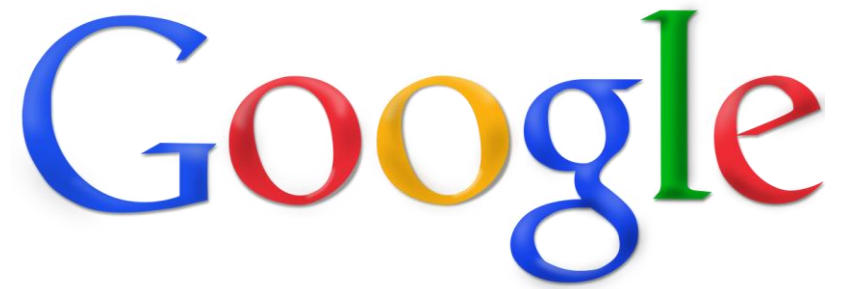
Information Retrieval

- Finding relevant information to the user's query
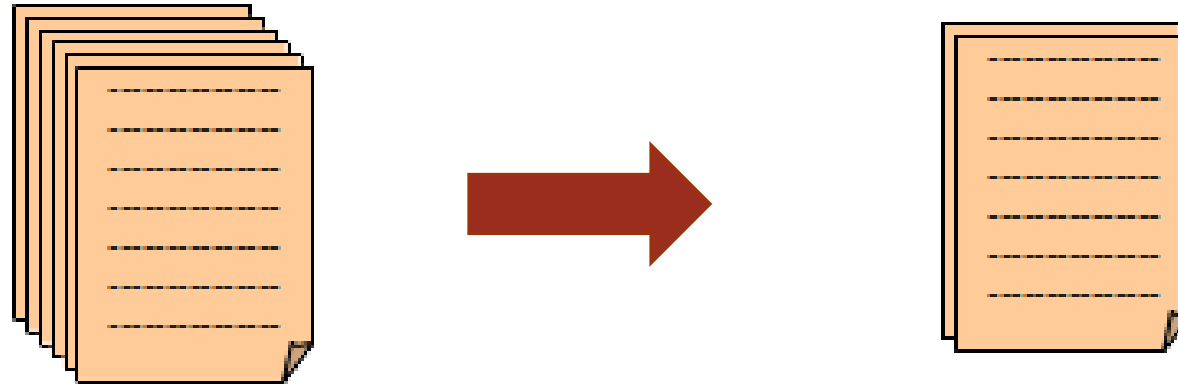
Information Retrieval
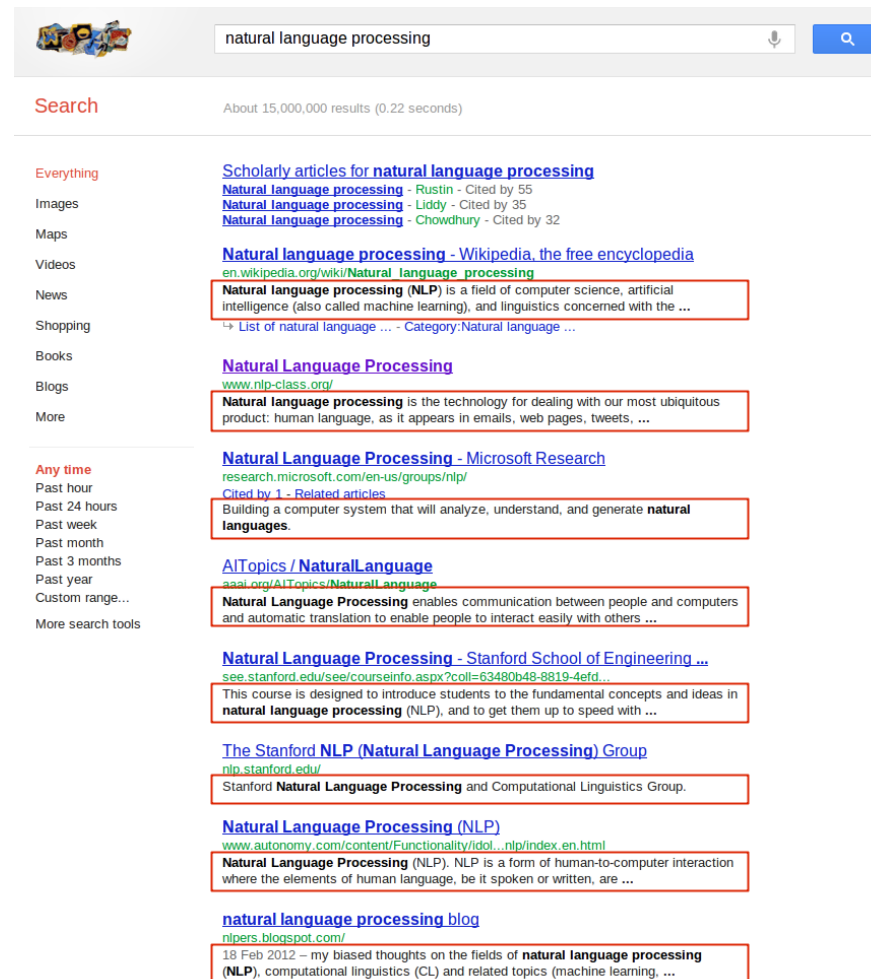
Summarization

- Finding the most relevant part of a document based on the user's information need

Summarization

Information Extraction

- Extracting the important items of a text and structuring them

Information Extraction

Information Extraction

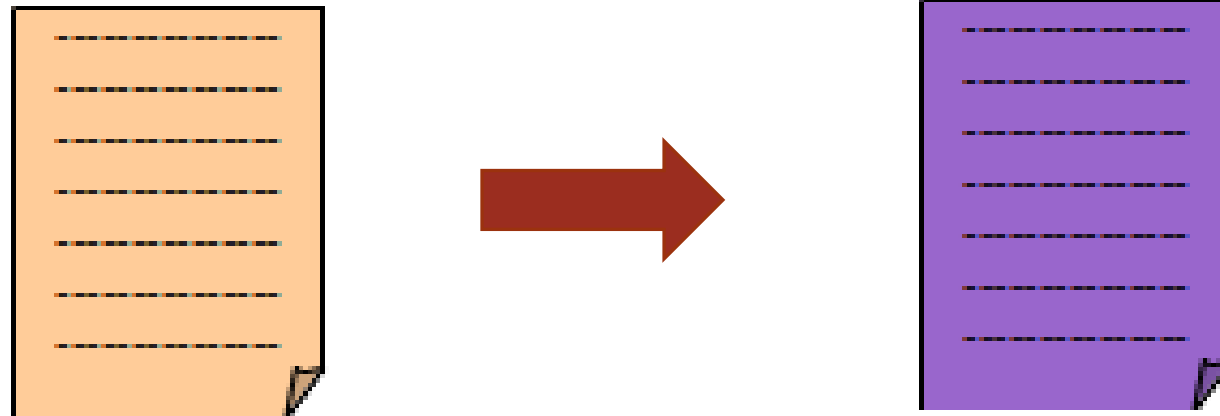# Question Answering

- Answering natural language questions asked by the user
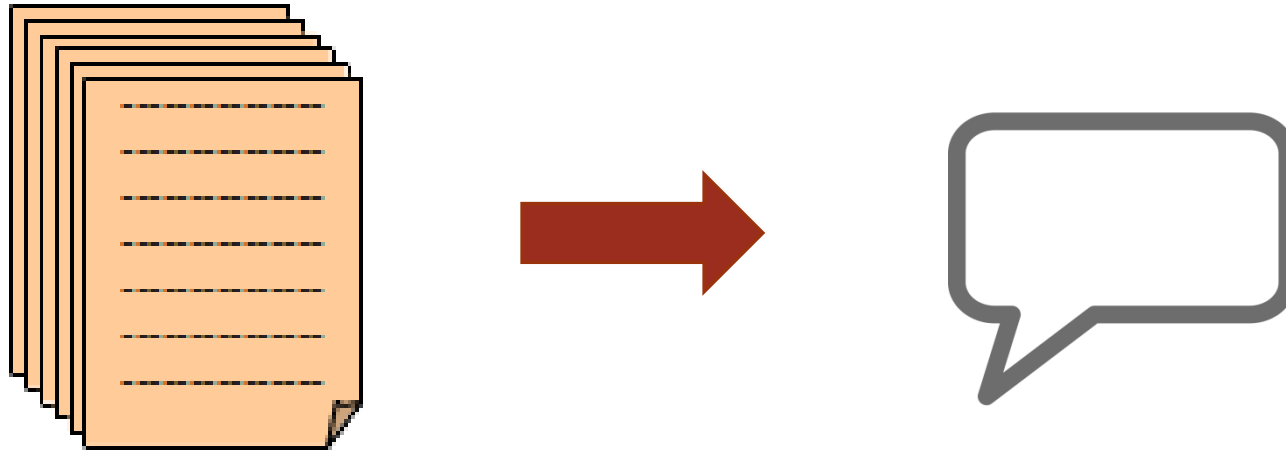
Question Answering

Machine Translation

- Translating a text from one language to another language

Machine Translation

Sentiment Analysis

- Identifying positive and negative opinions stated in a text
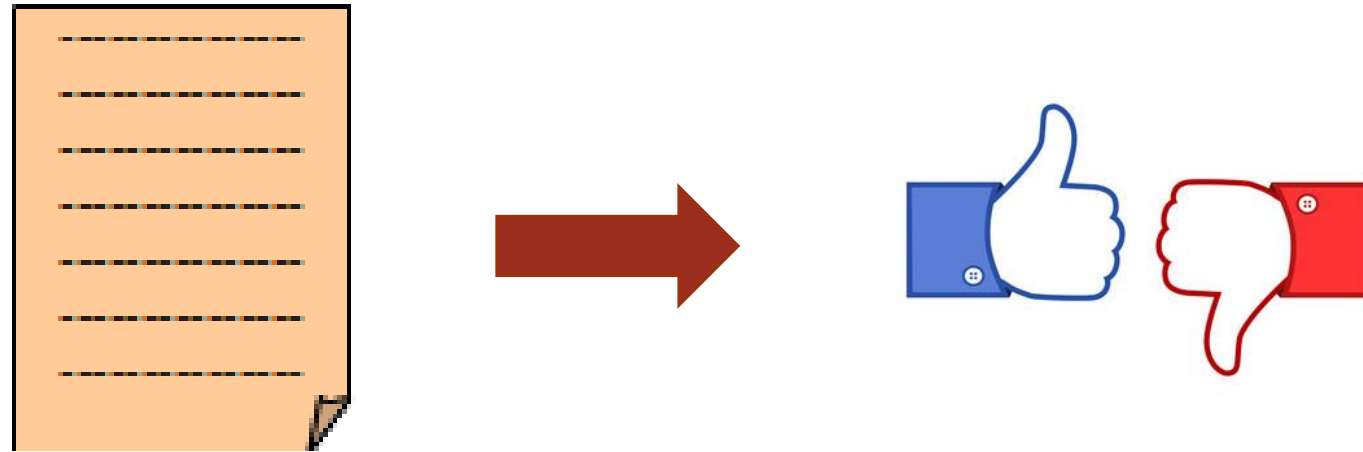
Sentiment Analysis

"Would not go back"
⊙⊙○○○ Reviewed October 19, 2013

We chose this hotel only because during the long weekend when we were visiting all hotels downtown were crazy-priced and this one seemed as a more reasonable price-value combination.
Unfortunately, we err-ed on the value side.
The only plus of this hotel is closeness to the T and cheapness of parking.
The room was small, the bathroom didn't have any counter big enough to put your travel purse on, bed and pillows were uncomfortable.
Sound isolation is terrible - noise from the hall carries all the way in.
The floor in the room is not carpeted in the main area and therefore if someone above you wears hills you hear it very well.

Stayed October 2013, traveled as a couple

⊙⊙○○○ Value   ⊙⊙○○○ Rooms
⊙⊙⊙⊙⊙ Location   ⊙⊙○○○ Cleanliness
⊙⊙○○○ Sleep Quality ⊙⊙⊙⊙○ Service
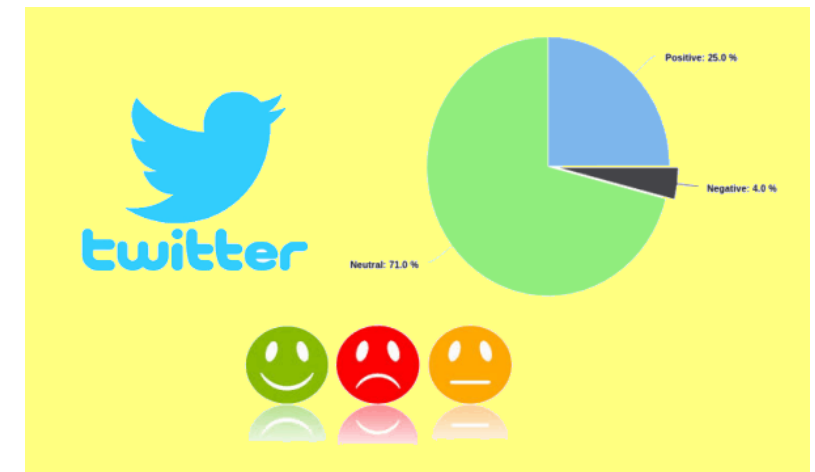
Optical Character Recognition

- Recognizing printed or handwritten texts and converting them to computer-readable texts

Optical Character Recognition
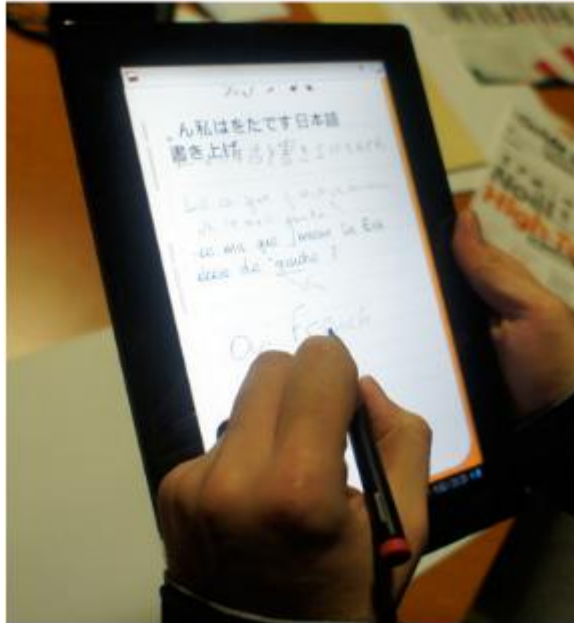
Word Prediction

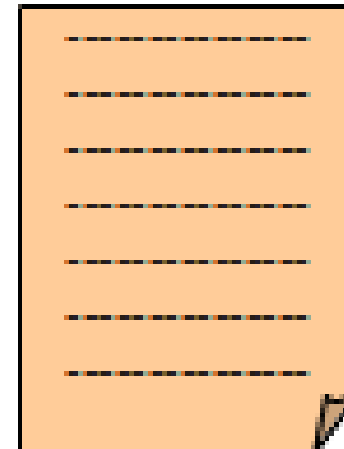- Predicting the next word that is highly probable to be typed by the user

Word Prediction
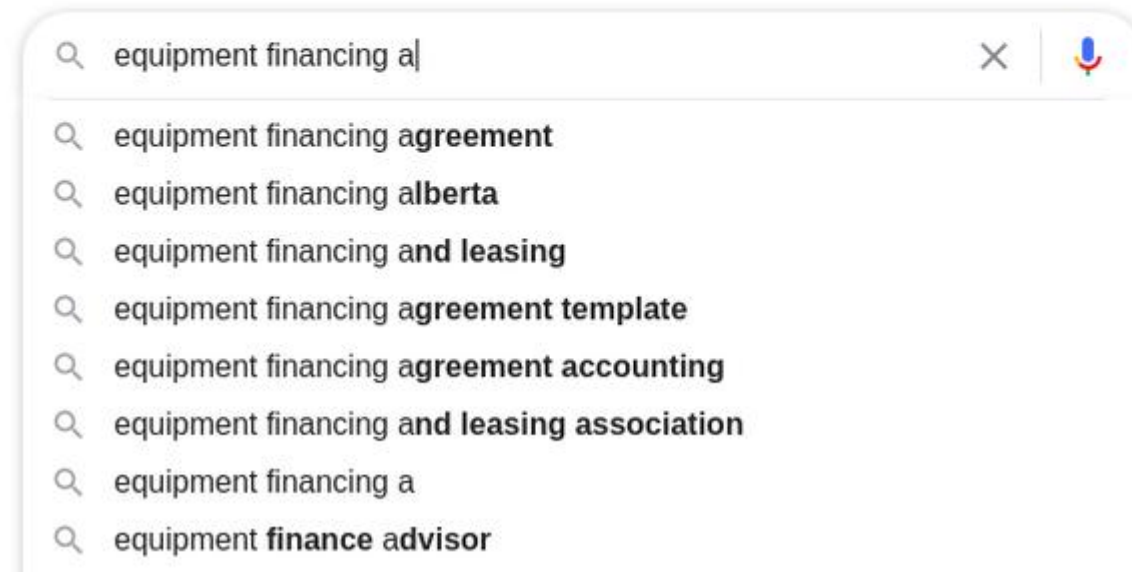
Dialogue Systems
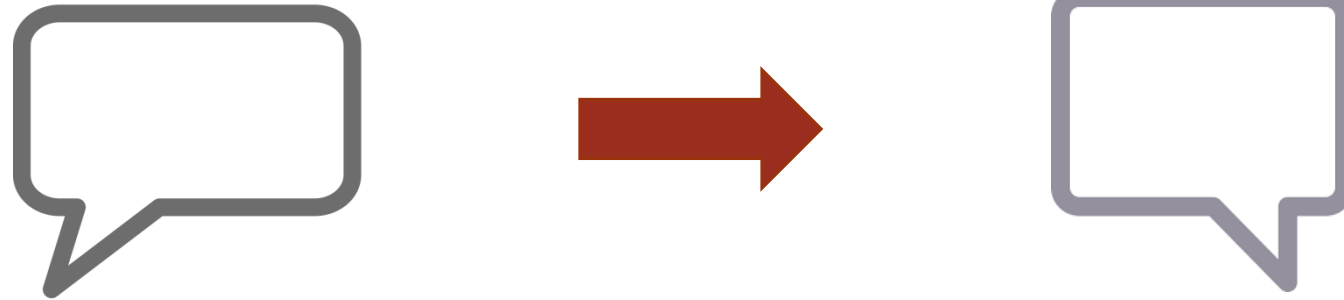
- Running a dialog between the user and the system

Dialogue Systems

Natural Language Processing

- Applications
  - Text Technologies
  - Speech Technologies

- Techniques

Speech Recognition

- Recognizing a spoken language and transforming it into a text
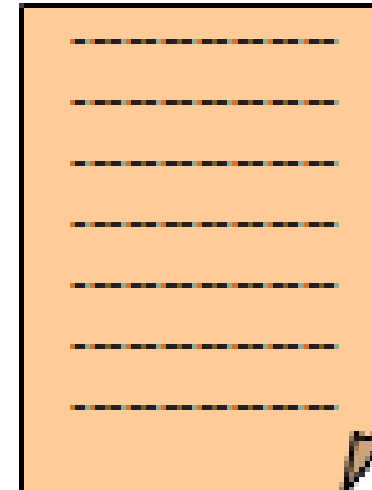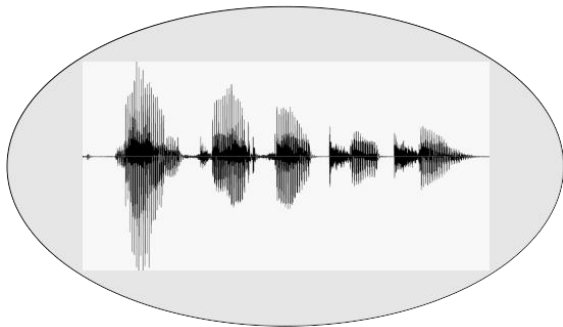
Speech Recognition
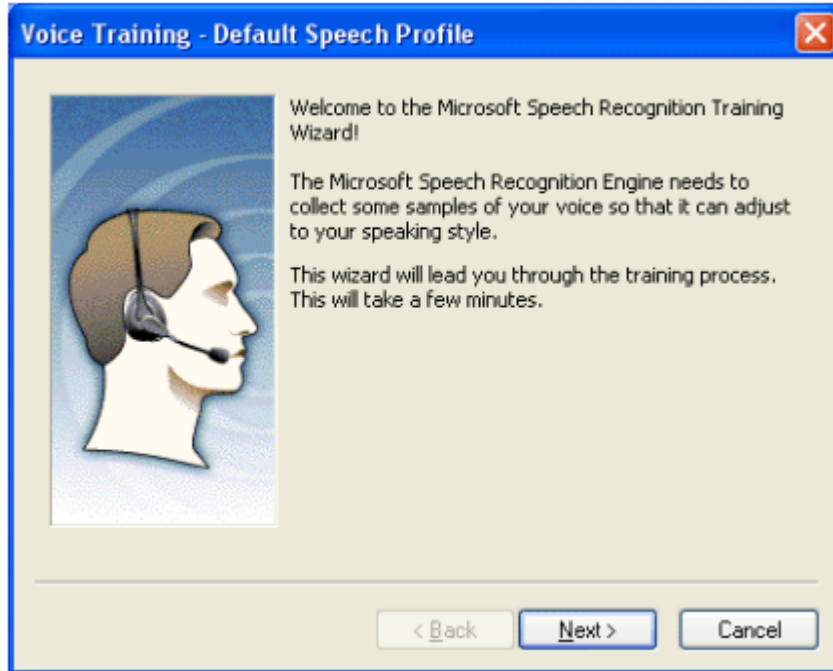
Speech Synthesis

- Producing a spoken language from a text
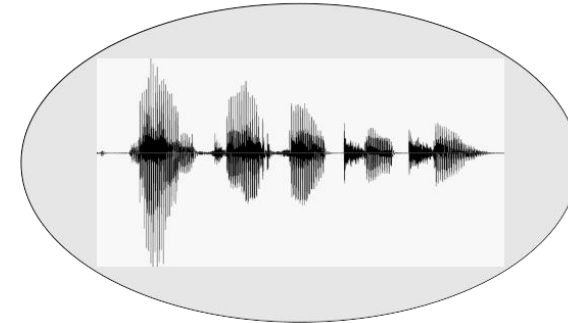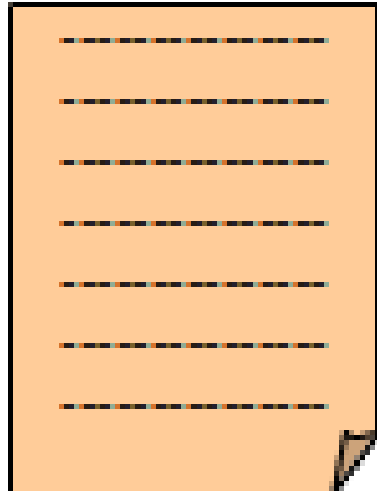
Speech Synthesis
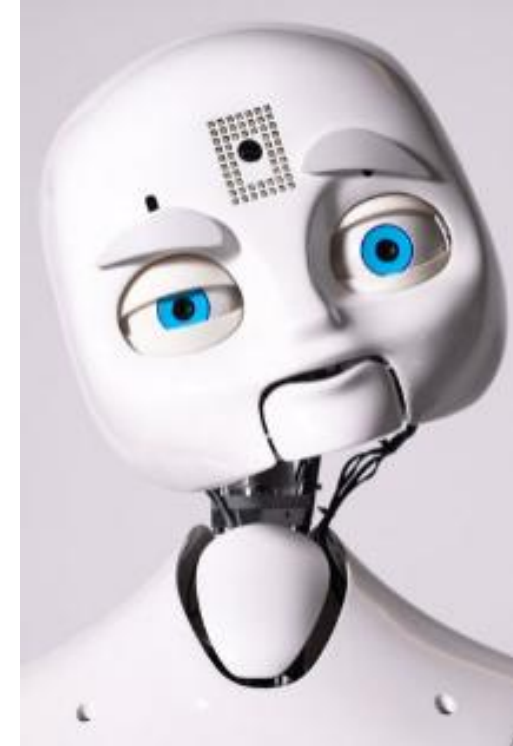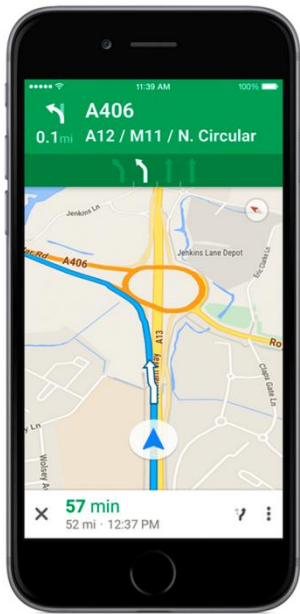
Spoken Dialogue Systems

- Running a dialog between the user and the system
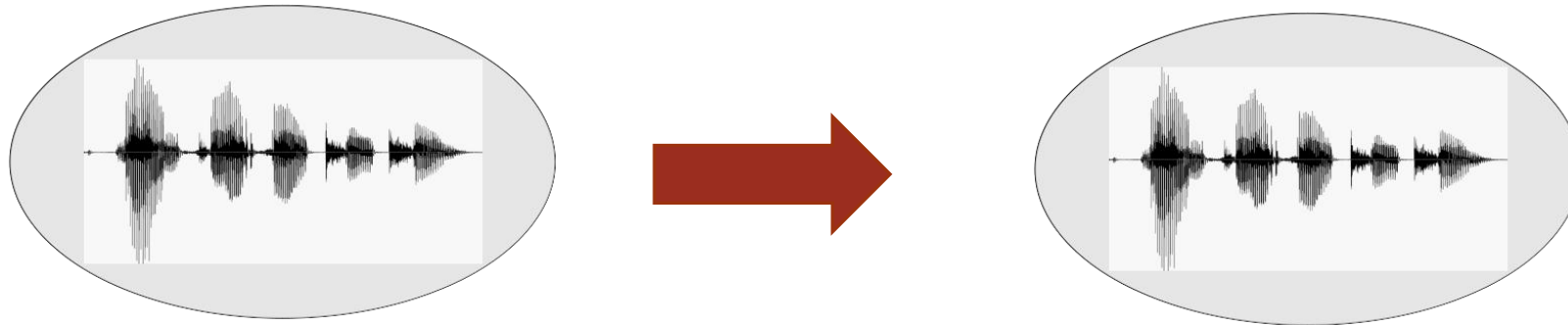
Spoken Dialogue Systems

Applications' Levels

- Easy (mostly solved)
  - Spell and grammar checking
  - Spam detection
  - Word prediction

- Intermediate (good progress)
  - Information retrieval
  - Sentiment analysis
  - Machine translation
  - Information extraction

- Difficult (still hard)
  - Question answering
  - Summarization
  - Dialogue system

Outline

- Administrative Information

- **Introduction to the Course**
  ◦ NLP Applications
  ◦ **NLP Techniques**

- Overview of the Semester

Part Of Speech Tagging

"I saw the man on the roof."

Part Of Speech Tagging

"I saw the man on the roof."

" I$_{[PRON]}$ saw$_{[V]}$ the$_{[DET]}$ man$_{[N]}$ on$_{[PREP]}$ the$_{[DET]}$ roof$_{[N]}$. "

[PRON]	Pronoun
[PREP]	Preposition
[DET]	Determiner
[V]	Verb
[N]	Noun

Parsing

"I saw the man on the roof."

Parsing

"I saw the man on the roof."
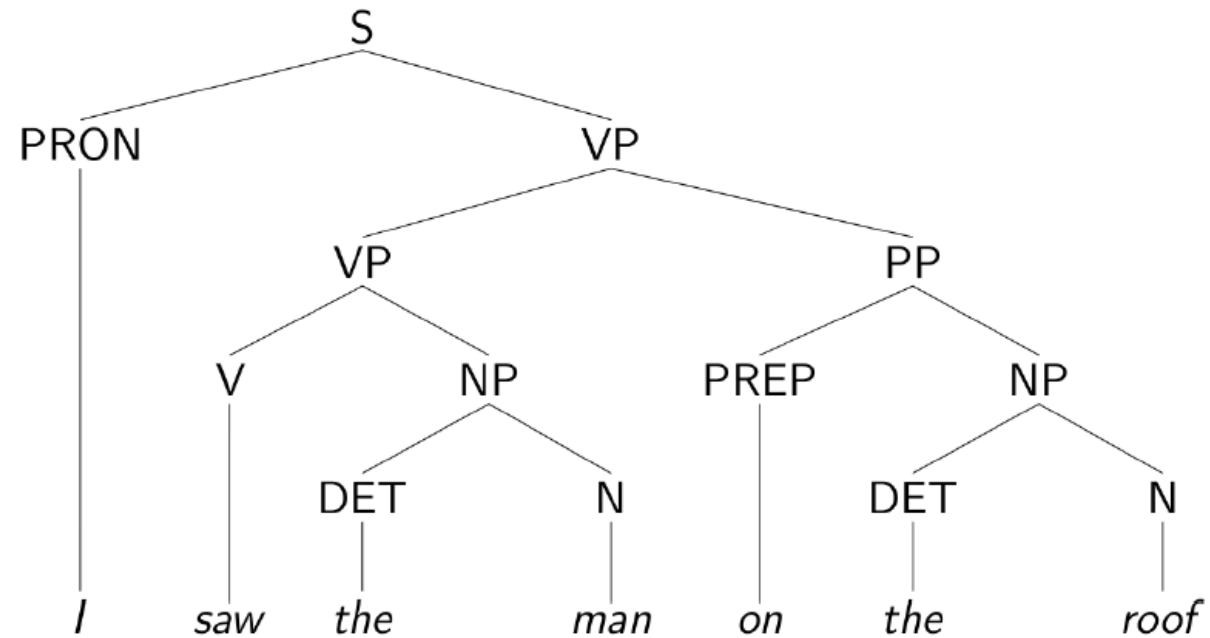
Named Entity Recognition

"Steven Paul Jobs, co-founder of Apple Inc, was born in California."

Named Entity Recognition

"Steven Paul Jobs, co-founder of Apple Inc, was born in California."

" Steven Paul Jobs, co-founder of Apple Inc, was born in California."

Person · Organization · Location

Word Sense Disambiguation

"Jim flew his *plane* to Texas."

"Alice destroys the item with a *plane*."

Word Sense Disambiguation

"Jim flew his plane to Texas."

"Alice destroys the item with a plane."
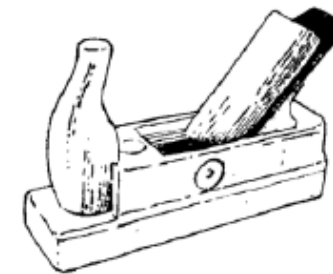
Semantic Role Labeling

"John *grills* a fish on an open fire."

Semantic Role Labeling

"John *grills* a fish on an open fire."

Cook
Food
Heating-Instrument

Semantic Role Labeling

"John *grills* a fish on an open fire."

Cook
Food
Heating-Instrument

"John *grills* a fish on an open fire ."
Cook   Food   Heating−Instrument

Outline

- Administrative Information

- Introduction to the Course

- **Overview of the Semester**

Course Syllabus

- Linguistics Levels and NLP Challenges

- Zipf's law

- Probabilistic Language Model

- Introduction to Neural Processing and FFNN

- Neural LM

- Vector Representation

- Text Classification

- Sequence Processing and Seq2Seq Models

- Part-of-Speech Tagging

- Named Entity Tagging

- Parsing

- Semantics
  - Word Similarity
  - Word Sense Disambiguation

- Text preprocessing and Spell correction

- Persian NLP

- Attention and Transformers

- Contextualized Text Representation

- NLP Applications
  - Sentiment
  - Question Answering
  - Chatbots

Enjoy the Course!