



Statistical Natural Language Processing

Lecture 1: Introduction

Dr. Momtazi

Amirkabir University of Technology

Outline

2

- ① Administrative Information
- ② Introduction to the Course
- ③ Overview of the Semester

Outline

3

- ① Administrative Information
- ② Introduction to the Course
- ③ Overview of the Semester

Course

4

- Credit point: 3
- Lecture
 - Saturdays 13:30 - 15:00
 - Mondays 13:30 - 15:00

Course Home Page

5

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

Assessment

6

- Regular attendance in the class
- Exercises (15%)
- Final project or seminar (15%)
- Midterm exam (30%)
- Final exam (40%)

Contact

7

- Email: momtazi@aut.ac.ir
- Consultations: Sundays 9:15 - 10:45
- Phone: 021-64542729

Text Book

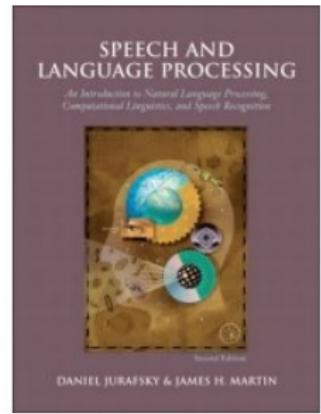
8

SPEECH AND LANGUAGE PROCESSING

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

by Daniel Jurafsky and James H. Martin

Second Edition (2008)
Publisher: McGraw Hill



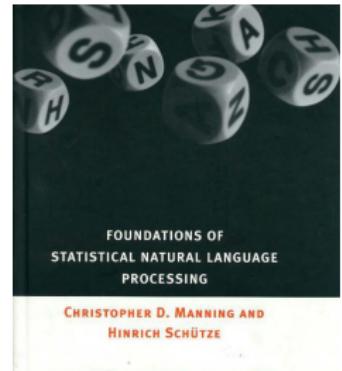
Text Book

9

Foundations of Statistical Natural Language Processing

by Christopher D. Manning, Hinrich Schütze

First Edition (1999)
Publisher: The MIT Press

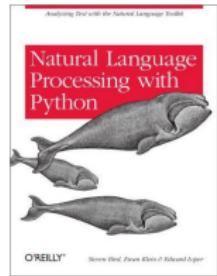


Other literature

10

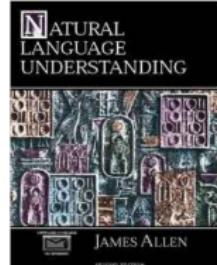
Natural Language Processing with Python

by Edward Loper, Ewan Klein, and Steven Bird



Natural Language Understanding

by James Allen



Relevance Journal & Conferences

11

- 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

- 12
① Administrative Information
- ② Introduction to the Course
- ③ Overview of the Semester

Different Types of Languages

13

- Natural languages
 - English
 - Persian
 - German
 - French
 - ...
- Formal languages
 - Java
 - Python
 - LaTeX
 - ...
- Descriptive languages
 - Biology: DNA
 - Chemistry: chemical formulas
 - ...

Natural Language

14

- 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

15

- Examples of Vocabularies

- English

- the
 - and
 - eat
 - you
 - book
 - ...

Natural Language Processing

16

- Applications
 - Text Technologies
 - Speech Technologies
- Techniques

Outline

17

① Administrative Information

② Introduction to the Course

NLP Applications

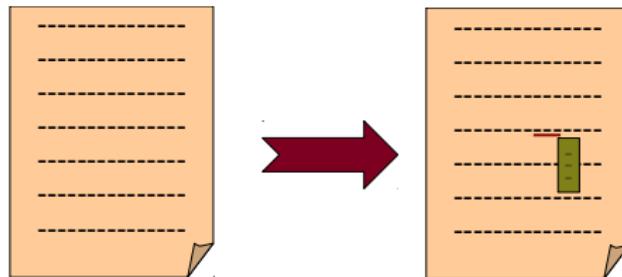
NLP Techniques

③ Overview of the Semester

Spell and Grammar Checking

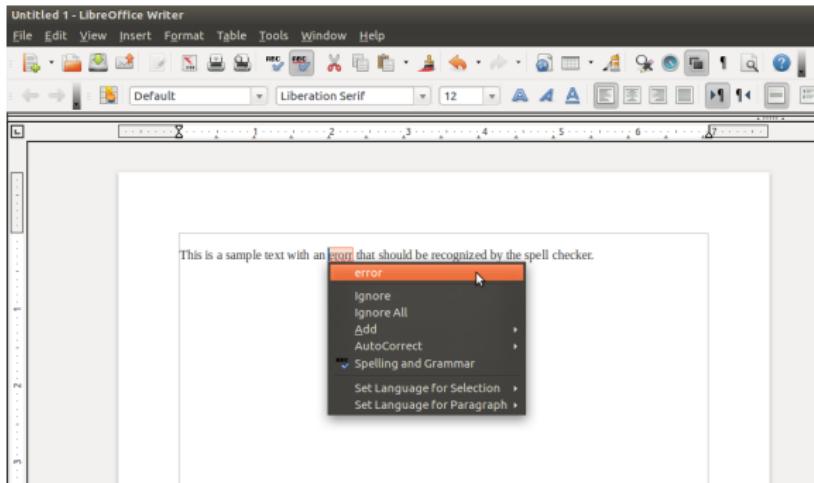
18

- Checking the spelling and the grammar of a text, and suggesting correct alternatives for the errors



Spell and Grammar Checking

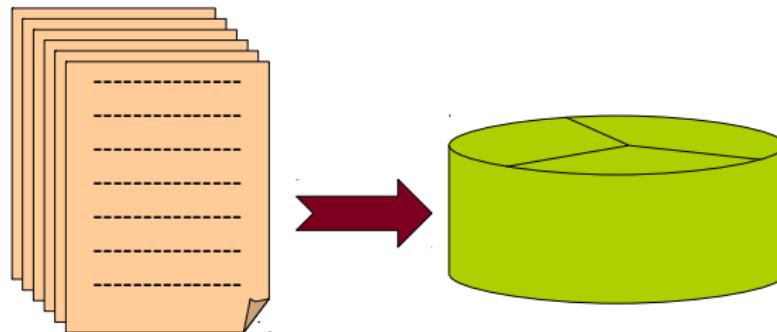
19



Text Categorization

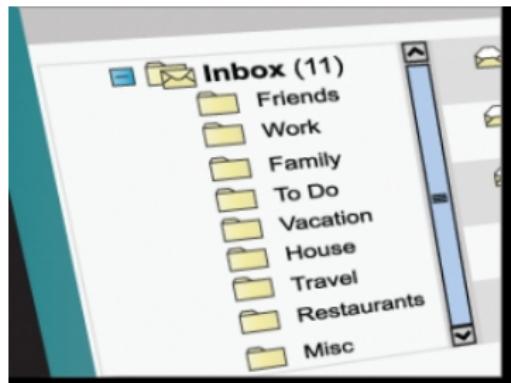
20

- Assigning each text to a category



Text Categorization

21



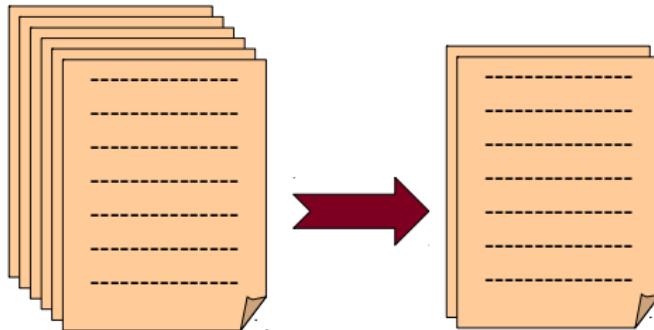
ABOUT DW DW AKADEMIE DW.DE IN 30 LANGUAGES ▾

The screenshot shows the DW website with a navigation bar at the top. The DW logo is on the left, followed by links for 'ABOUT DW', 'DW AKADEMIE', and 'DW.DE IN 30 LANGUAGES'. On the right is a search bar labeled 'Search DW' with a magnifying glass icon. Below the navigation is a blue header bar with links for 'TOP STORIES', 'MEDIA CENTER', 'PROGRAM', and 'LEARN GERMAN'. Underneath is a secondary navigation bar with links for 'GERMANY', 'WORLD', 'BUSINESS', 'SCI-TECH', 'GLOBALIZATION', 'ENVIRONMENT', 'CULTURE', and 'SPORTS'. At the bottom left is a link for 'TOP STORIES'.

Information Retrieval

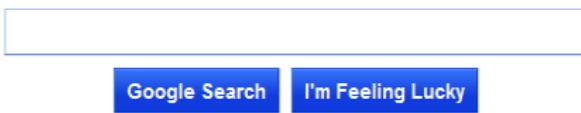
22

- Finding relevant information to the user's query



Information Retrieval

23



Google

Google Search

I'm Feeling Lucky



bing™

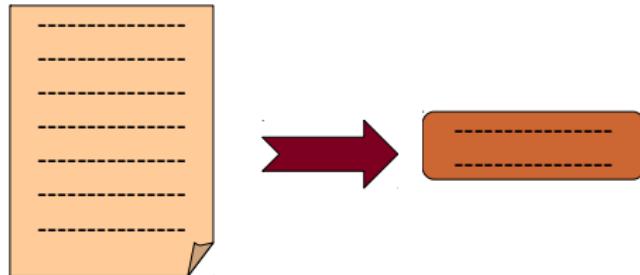


YAHOO!®

Summarization

24

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



Summarization

natural language processing

About 15,000,000 results (0.22 seconds)

Scholarly articles for natural language processing

Natural language processing - Rustin - Cited by 55
 Natural language processing - Liddy - Cited by 35
 Natural language processing - Chowdhury - Cited by 32

Natural language processing - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Natural_language_processing

Natural language processing (NLP) is a field of computer science, artificial intelligence (also called machine learning), and linguistics concerned with the ...

→ List of natural language ... - Category:Natural language ...

Natural Language Processing
www.nlp-class.org/

Natural language processing is the technology for dealing with our most ubiquitous product: human language, as it appears in emails, web pages, tweets, ...

Natural Language Processing - Microsoft Research
research.microsoft.com/en-us/groups/nlp/

Cited by 1 - Related articles

Building a computer system that will analyze, understand, and generate natural languages.

AITopics / NaturalLanguage
aaai.org/AITopics/NaturalLanguage

Natural Language Processing enables communication between people and computers and automatic translation to enable people to interact easily with others ...

Natural Language Processing - Stanford School of Engineering ...
see.stanford.edu/seeconferinfo.aspx?confID=63480&id=8819-4ef0f

This course is designed to introduce students to the fundamental concepts and ideas in natural language processing (NLP), and to get them up to speed with ...

The Stanford NLP (Natural Language Processing) Group
nlp.stanford.edu/

Stanford Natural Language Processing and Computational Linguistics Group.

Natural Language Processing (NLP)
www.algonomy.com/content/Functionality/videt_nlp/index_en.html

Natural Language Processing (NLP). NLP is a form of human-to-computer interaction where the elements of human language, be it spoken or written, are ...

natural language processing blog
moms.blogspot.com/

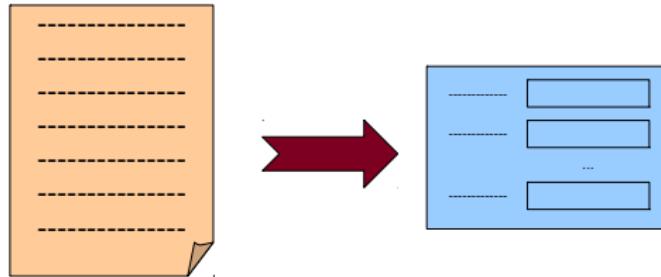
18 Feb 2012 - my biased thoughts on the fields of natural language processing

Momtazi | SNLP

Information Extraction

26

- Extracting the important items of a text and structuring them



Information Extraction



WIKIPEDIA

Article Talk

Albert Einstein

From Wikipedia, the free encyclopedia

"Einstein" redirects here. For other uses, see Einstein (disambiguation).

Albert Einstein (*Albert Bernhard* Einstein; *Albert Leopold* Einstein; 14 March 1879 – 18 April 1955) was a German-born theoretical physicist who developed the theory of relativity, one of the two pillars of modern physics (along with quantum mechanics). His work is also known for its influence on the philosophy of science.

He received the Nobel Prize in Physics in 1921 for his explanation of the photoelectric effect.

He was a leader of the peace movement in the United States during the Cold War, which led him to oppose both World War II and the Marshall Plan.

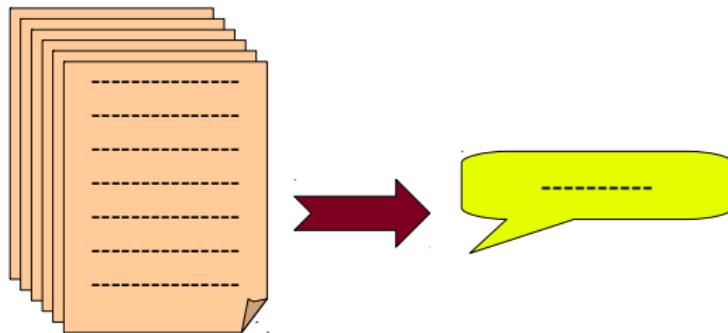
He was awarded the 1951 Nobel Peace Prize "for his services to mankind".

Einstein published more than 300 scientific papers along with over 150 non-scientific works.¹⁷⁷ His great intelligence and originality be-

Question Answering

28

- Answering natural language questions asked by the user



Question Answering

29



Question Answering

30



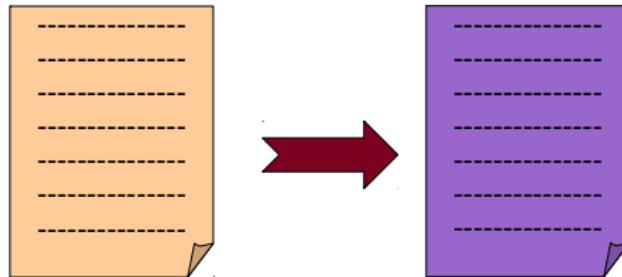
Crete

HARRIET BOYD
HAWES WAS
THE FIRST WOMAN
TO DISCOVER &
EXCAVATE A
MINOAN SETTLEMENT
ON THIS ISLAND

Machine Translation

31

- Translating a text from one language to another language



Machine Translation

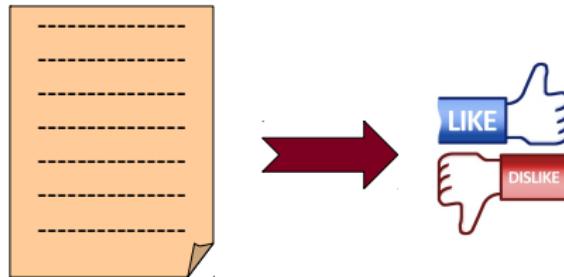
32



Sentiment Analysis

33

- Identifying positive and negative opinions stated in a text



Sentiment Analysis

The Ritz-Carlton, Berlin Hotel

Home Berlin Hotels Flights Vacation Rentals Restaurants Things to Do Best of 2012 More Write a Review

All 662 Berlin hotels ▾

The Ritz-Carlton, Berlin

Potsdamer Platz 1, 10719 Berlin, Germany [Get directions](#)

Ranked #7 of 602 hotels in Berlin
1,047 reviews | [Traveler's Choice® 2012 Winner](#) | [TripAdvisor Traveler Rating](#)

Show the lowest price for this hotel*

Check-in: 4/16/2012 Check-out: 4/18/2012 Adults: 2 ▾

Show Prices

Booking.com Expedia.com Venere.com Hotels.com Hotelde.com

From our partners

Professional photos

282 traveler photos

647 reviews from our community

What travelers say about this hotel

- "Wonderful place!" (5)
- "Wunderbar! great!" (5)
- "Phantastisch!" (5)
- "Very good." (4)
- "Excellent service!" (4)

See which rooms travelers prefer - [22 traveler tips](#)

Only show reviews for [Family \(1\)](#) [Couples \(4\)](#) [Business \(1\)](#)

647 reviews sorted by [Date](#) [Rating](#) [English first](#) ▾

"A really great place to stay in Berlin"

[Globe icon](#) [Reviewed March 21, 2012](#) [Read review](#)

The place is excellent. The staff is very professional and considerate. I like it a lot. It's a major luxury hotel. That means a lot of getting around. It is like the prices you get staying in Chicago versus New York or San Francisco. The hotel itself is very new and nice. The service I got was excellent. They have a family...

[More](#) [Was this review helpful?](#) 

Written with TripAdvisor

"Great location and very comfortable hotel"

[Globe icon](#) [Reviewed March 14, 2012](#) [Read review](#)

1 person found this review helpful!

My family of 5 stayed for 9 nights in end December/beginning January. We had 2

Travelers also viewed...

 [Swissotel Berlin](#) [#1000 in Berlin](#) [Berlin, Germany](#) [More Prices](#)

 [Radisson Blu Hotel](#) [#1000 in Berlin](#) [Berlin, Germany](#) [More Prices](#)

 [Palm Inn Berlin](#) [#1000 in Berlin](#) [Berlin, Germany](#) [More Prices](#)

Browse nearby

Hotels (662) | Restaurants (2,041) | Things to Do (1,000)

 [Interactive map](#) Sponsored by 

Travelers think this accommodation offers a good night's sleep.

Sleep Rating:  [Learn more](#) Sponsored by 

SPG [AccorHotels](#)

Europa ist immer eine Reise wert

Lassen Sie sich von den Angestellten bei [Howard Hotels & Resorts Worldwide](#) beraten!

[EXPEDIA](#)

The Ritz-Carlton, Berlin [Book now](#) [Learn more](#) [Read reviews](#) [Write a review](#)

The Ritz-Carlton, Berlin [See all on hotel.com](#)



Momtazi | SNLP

berlin - including Hotel Adlon - due to its
own dismay that ever since then - as well as
the numerous reviews I read - I really had hoped that the recent
overly particular guests or the occasional
bad best properties. As I booked through
Accorhotels.com, I could get a guarantee that if paid in advance there was no running
out of room.

The best thing about this hotel is the
location - it is the only one close to many personal
activities for - a wonderful selection of fresh
bread, rolls, and even bagels, and
many other choices; there is also an equally
large number of restaurants and bars of all kinds
and styles American style hot spot.
This isn't enough, however, as the price
made me wonder if it extra charge. In
addition, the service was not very good.
I have never been so disappointed.
I usually expect a good service, but it's not
something you can buy. It's something you
have to earn. And I have to say that I
wasn't impressed at all. The service was
extremely friendly and professional and
the staff were always happy and helpful.
The room itself was huge and
decorated in a little later) and the double glazing was
the best when I was staying. So I was very
and they advanced some nice
amenities such as a large bathroom
to shower and get ready to go, the toilet
was not great because I spoke to the
house would resolve the issue with the
room. I am not sure if I will stay again. I
hope the room to be remodeled. I am not
very particular about the room and the room
is very nice and comfortable.

Optical Character Recognition

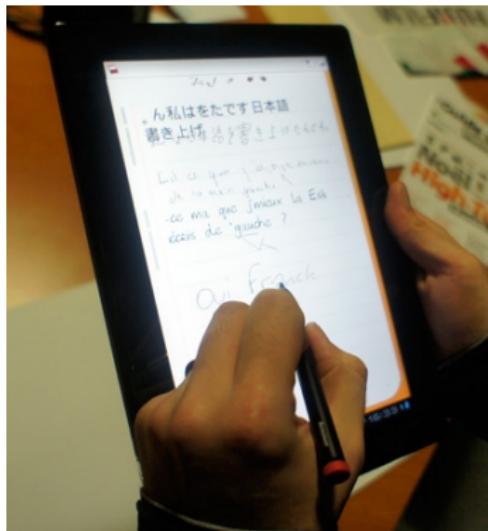
35

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



Optical Character Recognition

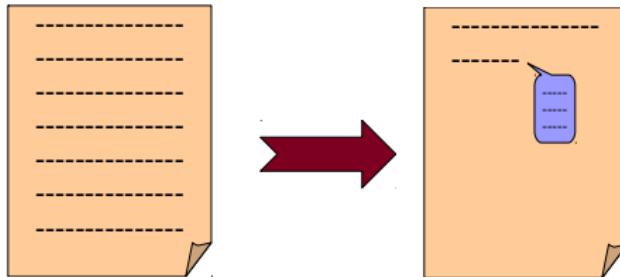
36



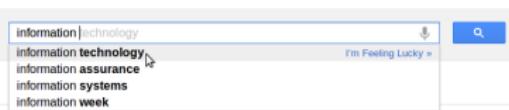
Word Prediction

37

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



Word Prediction



[Information technology - Wikipedia, the free encyclopedia](#)
en.wikipedia.org/w/index.php?title=Information_technology&oldid=910111111

Information Technology (IT) is concerned with technology to treat information. The acquisition, processing, storage and dissemination of vocal, pictorial, textual ...
→ Information systems - Information history - Category:Information technology

[Information Technology – All About Information Technology – Wh...](#)
jobssearchtech.about.com/od/careersintechnology/p/ITDefinition.htm

Information Technology and IT definition: What **Information technology** actually means. How **Information technology** is different from computer science.

[RIT Information Sciences & Technology](#)
www.ist.rit.edu/
offers bachelors and masters degrees in **information technology**, a masters degree in software development and management, and an advanced certificate in ...

[ScienceDaily: Information Technology News](#)
www.sciencedaily.com/news/computers_.../information_technology/
1 day ago – **Information Technology**: Read the latest in IT research from research institutes around the world. Updated daily, full-text, images, free.

[Government of India, Department of Information Technology \(DIT\) ..](#)
www.mit.gov.in/
Developing the **information technology** industry. Includes an organisation chart, subsidiary bodies.

[Information Technology - Everything You Need to Know](#)
informationtechnology.net/
What is **Information Technology**? **Information Technology**, or IT, is the study, design, creation, utilization, support, and management of computer-based ...

[Information Technology](#)
www.ibef.org/industry/informationtechnology.aspx
The Indian **Information technology** (IT) industry has played a key role in putting India on the global map and is now envisioned to become a US\$ 225 billion ...

[Information Technology - WetFeet.com](#)
www.wetfeet.com/careers-industries/careers/information-technology
Information Technology: Overview. E-mail, personal computers, and the Internet: These products of the information age have become common currency among ...

Natural Language Processing

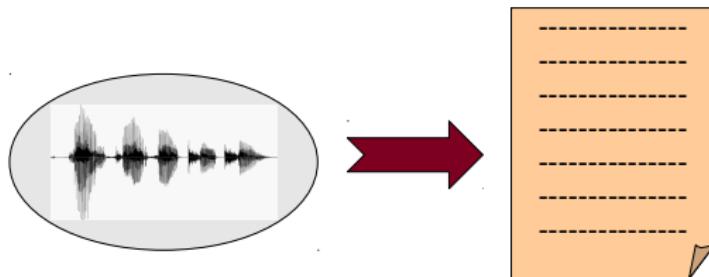
39

- Applications
 - Text Technologies
 - Speech Technologies
- Techniques

Speech Recognition

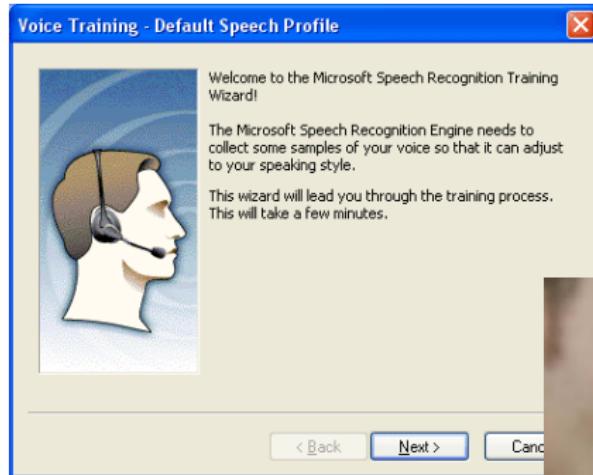
40

- Recognizing a spoken language and transforming it into a text



Speech Recognition

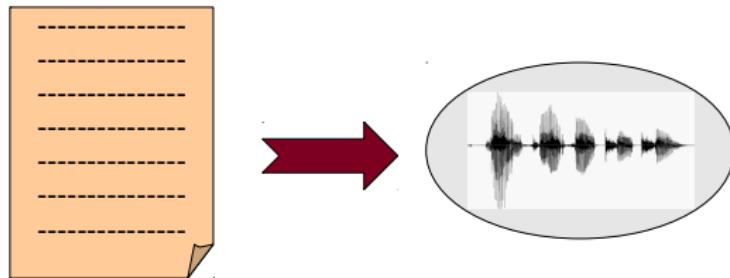
41



Speech Synthesis

42

- Producing a spoken language from a text



Speech Synthesis

43

**Text-to-Speech
TRANSLATOR**

»**oddcast**
CHARACTER DRIVEN COMMUNICATIONS

1. Enter Text: (250 character limit, 50 character limit for Asian language)

I'm your father, Luke

2.

Translate From: English To: Spanish

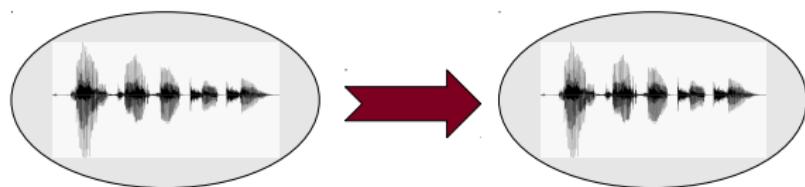
Voice: Alberto (Lat. Am.)



Spoken Dialog Systems

44

- Running a dialog between the user and the system



Spoken Dialog Systems

45



Applications' Levels

46

- 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
 - Dialog system

Outline

47

① Administrative Information

② Introduction to the Course

NLP Applications

NLP Techniques

③ Overview of the Semester

Part Of Speech Tagging

48

"I saw the man on the roof."

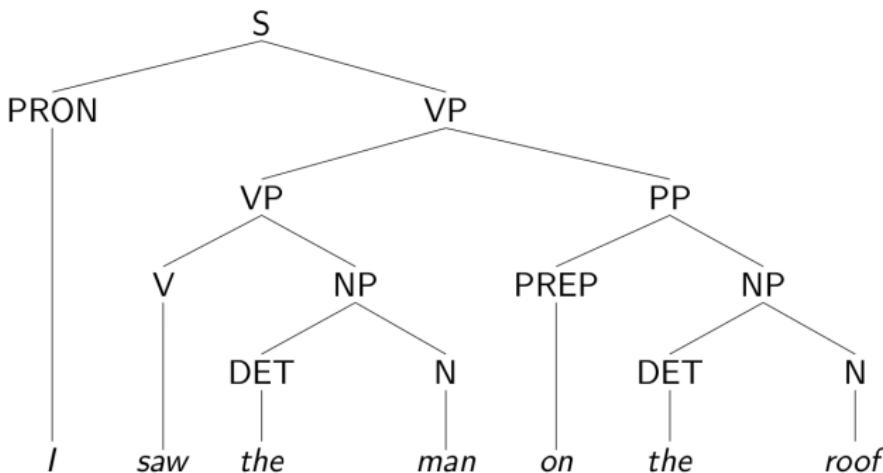
" I[PRON] saw[V] the[DET] man[N] on[PREP] the[DET] roof[N]. "

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

Parsing

49

"I saw the man on the roof."



Named Entity Recognition

50

“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

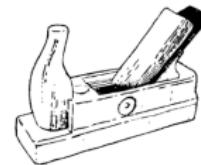
Word Sense Disambiguation

51

*“Jim flew his **plane** to Texas.”*



*“Alice destroys the item with a **plane**.”*



Semantic Role Labeling

52

“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

53

- ① Administrative Information
- ② Introduction to the Course
- ③ Overview of the Semester

Topics

- Introduction
- Linguistics Levels and NLP Challenges
- Mathematical Fundations
- Language Modeling
- Machine Learning Overview
- Part-of-Speech Tagging
- Parsing
- Named Entity Tagging
- Semantics
 - Word Similarity
 - Word Sense Disambiguation
- Semantics Role Labeling
- Text Classification and Clustering
- Topic Detection and Tracking
- NLP Applications

Rules of the Game

55

- In case you don't understand something:
 - ① Ask!!!
 - ② Ask!!!
 - ③ Ask!!!

Question?