Outline

1 Administrative Information
2 Introduction
3 NLP Applications
4 NLP Techniques
5 Linguistic Knowledge
6 Challenges
7 Course Materials
Outline

1 Administrative Information
2 Introduction
3 NLP Applications
4 NLP Techniques
5 Linguistic Knowledge
6 Challenges
7 Course Materials
NLP Course

- Lecture
  - Thursdays 9:15-10:45
  - Location: HS 3
  - Credit Point: 3 LP

- Assessment
  - Regular attendance in the class
  - Doing exercises
  - Final exam

- Contact
  - Saeedeh Momtazi
    - Email: saeedeh.momtazi@hpi.uni-potsdam.de
    - Office: A-1.7
NLP Course

- Course Home Page
  - [http://www.hpi.uni-potsdam.de/naumann/teaching/ss_12/natural_language_processing.html](http://www.hpi.uni-potsdam.de/naumann/teaching/ss_12/natural_language_processing.html)
    - Administrative information
    - Slides
    - Exercises

- Mailing List
  - Will be set later
Outline

1 Administrative Information
2 Introduction
3 NLP Applications
4 NLP Techniques
5 Linguistic Knowledge
6 Challenges
7 Course Materials
Different Types of Languages

- Natural languages
  - English
  - German
  - French
  - Spanish
  - ...

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

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

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
Examples of Vocabularies

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

- German
  - das
  - und
  - essen
  - du
  - buch
  - ...
Outline

1. Administrative Information
2. Introduction
3. NLP Applications
   Text Technologies
   Speech Technologies
4. NLP Techniques
5. Linguistic Knowledge
6. Challenges
7. Course Materials
Spell and Grammar Checking

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

This is a sample text with an **error** that should be recognized by the spell checker.

- Ignore
- Ignore All
- Add
- AutoCorrect
- Spelling and Grammar
- Set Language for Selection
- Set Language for Paragraph
Text Categorization

- Assigning each text to a category
Text Categorization
Information Retrieval

- Finding relevant information to the user’s query
Summarization

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

Everything

Images
Maps
Videos
News
Shopping
Books
Blogs
More

Any time
Past hour
Past 24 hours
Past week
Past month
Past 3 months
Past year
Custom range...
More search tools

Scholarly articles for natural language processing
Natural language processing - Rustin - Cited by 55
Natural language processing - Liddy - Cited by 36
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...

Natural Language Processing
www.nlp-cras.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 answers

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

AITopics / NaturalLanguage
aat.org/ATopics/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/see/coursesinfo.aspx?code=CS5480b&gt;685;94f;9d.

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.autonomy.com/about/products/functionalit.html?pagename=english.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
gapers.blogspot.com/

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

- Extracting the important items of a text and structuring them
Information Extraction

Potsdam

The name "Potsdam" originally refers to the area of the German federal state of Brandenburg, where the city of Potsdam is situated. Potsdam is an important historical site in Germany, home to the Brandenburg Gate, the Kurfürstendamm, and the Palace of Sanssouci.

Geography

The area around Potsdam shows signs of human settlement dating back to the 13th century. The town is located on the Havel River, which flows through the Brandenburg region.

History

Potsdam was founded in 1324 and has played an important role in German history. It was the capital of the state of Brandenburg from 1713 to 1871. In 1990, it became the capital of the federal state of Brandenburg.

Governer of Potsdam

Between 1851 and 1918, the city of Potsdam was the capital of the Province of Potsdam, which was divided into seven historical administrative districts.

Coordinates

The coordinates for Potsdam are 52°24′0″N 13°41′E.

Administration

Potsdam is located in the federal state of Brandenburg, within the district of Potsdam-Babelsberg.

Basic statistics

Area: 187.28 km² (72.31 sq mi)
Elevation: 35-114 m
Population: 156,906 (31 December 2010)
Density: 838 /km² (2,170 /sq mi)

Other information

Time zone: CET/CEST (UTC+1/2)
Postal codes: 14401-14482
Website: www.potsdam.de

Natural attractions in the area include the Brandenburg Gate, the Kurfürstendamm, and the Palace of Sanssouci.

Albert Einstein

Born
14 March 1879
Ulm, Kingdom of Württemberg, German Empire

Died
18 April 1955 (aged 76)
Princeton, New Jersey, United States

Residence
Germany, Italy, Switzerland, Austria, Belgium, United Kingdom, United States

Citizenship
Switzerland/Germany (1879–1901)
Switzerland (1901–1955)
Austria (1912–1921)
Czechoslovakia (1919–1930)
United States (1940–1955)

Fields
Physics

Institutions
Universität Zürich (1900–1909)
Thüringer Universität in Göttingen (1911–1914)
Diplomatic Institute in The Hague and Brussels (1914–1915)
Institut für Physik (1919–1933)
Massachusetts Institute of Technology (1933–1955)
Albert Einstein Institute

Alma mater
ETH Zurich
University of Zurich

Doctoral advisor
Ph.D. (1905)

other academic positions
Diplomatic Institute in The Hague and Brussels (1914–1915)

Notable students
Emil G. Strauss
Nathan Rosen
Leo Szilard
Kazimierz Ajdukiewicz

Known for
General relativity and special relativity
Photovoltaic effect
Mass-energy equivalence
Theory of Brownian Motion
Einstein field equations

Albert Einstein is often regarded as the father of modern physics and one of the most prolific scientists in human history. He is best known for his theory of relativity, which brought him fame and fortune. In 1905, he published a paper on the special theory of relativity, which introduced the concept of time dilation and the equivalence of mass and energy. This work laid the foundation for Einstein's later work on general relativity, which described the gravitational effect of massive objects on spacetime. Einstein's theories have had a profound impact on the development of modern physics and continue to be studied and applied in many fields today.
Question Answering

- Answering natural language questions asked by the user
Question Answering
Question Answering

Crete

Harriet Boyd Hawes was the first woman to discover & excavate a Minoan settlement on this island.
Machine Translation

- Translating a text from one language to another language
This is a sample sentence in English which is translated to German.

Dies ist ein Beispielsatz in englischer Sprache, die auf deutsch übersetzt wird.
Data Fusion

- Combining extracted information from several text files into a database or an ontology
Jean Dujardin

Jean Edmond Dujardin (French pronunciation: [ʒɛ̃ dy.dʒardɛ̃]; born 19 June 1972) is a French actor and comedian. He has worked notably with director Michel Hazanavicius, starring in his OSS 117 spy parodies, the starring role in Hazanavicius' silent movie The Artist, playing actor George Valentin, receiving widespread acclaim. The role won him the Golden Globe Award for Best Actor in a Musical or Comedy, the BAFTA Award for Best Actor in a Leading Role, the Screen Actors Guild Award for Outstanding Performance by a Male Actor in a Leading Role, the Cannes Film Festival Best Actor Award, and the Academy Award for Best Actor, the first French-speaking actor to win the Academy Award for Best Actor.

## Early life

Dujardin was born and raised in Rueil-Malmaison (Hauts-de-Seine, Île-de-France), a commune in the western suburbs of Paris, France. After attending high school, he went to work for his parents' construction company. Dujardin began contemplating a career in acting while serving his mandatory military service a few years later.

## Career

<table>
<thead>
<tr>
<th>Year</th>
<th>Film</th>
<th>Role</th>
<th>Award</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>Crâne</td>
<td>Valentin</td>
<td>Academy Award for Best Actor</td>
</tr>
<tr>
<td>2005</td>
<td>La Vie en Rose</td>
<td>Georges Valentin</td>
<td>Academy Award for Best Actor</td>
</tr>
<tr>
<td>2009</td>
<td>The Artist</td>
<td>George Valentin</td>
<td>Academy Award for Best Actor</td>
</tr>
</tbody>
</table>

Dujardin's success continued with roles in a series of critically acclaimed films, earning him accolades and recognition from audiences and critics alike. His portrayal of George Valentin in The Artist cemented his status as a leading man in Hollywood, and his success continued with roles in several more films, earning him additional awards and nominations.

## Past projects

Dujardin has also been involved in several past projects, including roles in films such as "OSS 117: Paris, Nos Amours," and "The Artist," as well as in television roles and image appearances.

## Filmography

- "The Artist" (2011)
- "The Intouchables" (2011)
- "The Disappearance of Eleanor Rigby: Him & Her" (2013)

## Awards

Dujardin has been nominated for several awards, including the Academy Award for Best Actor for his role in "The Artist," and has won the Cannes Film Festival Best Actor Award, the Screen Actors Guild Award for Outstanding Performance by a Male Actor in a Leading Role, and the Academy Award for Best Actor.

## Personal life

Dujardin and his wife, Nathalie Pechalat, have a son named Louis (born 2007). Dujardin is known for his role as the French ventriloquist on "C'est bien, Monsieur," a children's TV show. He and his wife have been married since 2009, and they reside in Paris, France.

## Further reading

Sentiment Analysis

- Identifying positive and negative opinions stated in a text
"A Complete Letdown" | Reviewed February 22, 2012

I was totally disappointed by this hotel. I made a reservation in my name and was charged over twice the amount I was quoted for. The city appeared to be deserted when I was there. The room was not cleaned and the service was terrible. I would not stay here again.

Sentiment Analysis

The Ritz-Carlton, Berlin

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
Outline

1. Administrative Information
2. Introduction
3. NLP Applications
   - Text Technologies
   - Speech Technologies
4. NLP Techniques
5. Linguistic Knowledge
6. Challenges
7. Course Materials
Speech Recognition

- Recognizing a spoken language and transforming it into a text
Welcome to the Microsoft Speech Recognition Training Wizard!

The Microsoft Speech Recognition Engine needs to collect some samples of your voice so that it can adjust to your speaking style.

This wizard will lead you through the training process. This will take a few minutes.
Speech Synthesis

- Producing a spoken language from a text
Speech Synthesis

1. Enter Text: (200 character limit, 50 character limit for Asian language)
   I'm your father, Luke

2. Translate From:
   To:
   English
   Spanish
   Voice:
   Alberto (Lat. Am.)
- Running a dialog between the user and the system
Spoken Dialog Systems

Computer, show me the shortest route to TSMC, HsinChu City
Applications’ Levels

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

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

- Difficult (still hard)
  - Question answering
  - Summarization
  - Dialog system
Outline

1. Administrative Information
2. Introduction
3. NLP Applications
4. NLP Techniques
5. Linguistic Knowledge
6. Challenges
7. Course Materials
Part Of Speech Tagging

“I saw the man on the roof.”


[PRON] Pronoun
[PREP] Preposition
[DET] Determiner
[V] Verb
[N] Noun
"I saw the man on the roof."

Parsing

S
  └── PRON
        ├── VP
        │    └── V
        │          └── DET
        │                └── saw
        │                        └── the
        └── NP
            └── N
                └── man
PP
  └── PP
      └── NP
          └── N
              └── on
                  └── the
                      └── roof
“Steven Paul Jobs, co-founder of Apple Inc, was born in California.”
“Jim flew his *plane* to Texas.”

“Alice destroys the item with a *plane*.”
“John grills a fish on an open fire.”

Cook
Food
Heating−Instrument
Outline

1. Administrative Information
2. Introduction
3. NLP Applications
4. NLP Techniques
5. Linguistic Knowledge
6. Challenges
7. Course Materials
Linguistic Knowledge

- Phonetics and Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse
Phonetics and Phonology

- The study of linguistic sounds and their relations to words
Morphology

- The study of internal structures of words and how they can be modified
- Parsing complex words into their components

**unbelievable**

**un+believ+able**

prefix: **un**  
root: **believe**  
suffix: **able**
Syntax

- The study of the structural relationships between words in a sentence

*I saw the man on the roof.*

```
  S
 /   \\
|     |
PRON VP

  VP
 /   \\
|     |
  V NP PP

  NP
 /   \\
|     |
 DET N DET N

  N
 /   \\
|     |
 saw the man on the roof
```

Semantics

- The study of the meaning of words, and how these combine to form the meanings of sentences
- Realizing lexical relations among words

<table>
<thead>
<tr>
<th>Hyponymy (is a)</th>
<th>dog &amp; animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meronymy (part of)</td>
<td>arm &amp; body</td>
</tr>
<tr>
<td>Synonymy</td>
<td>fall &amp; autumn</td>
</tr>
<tr>
<td>Antonymy</td>
<td>tall &amp; short</td>
</tr>
</tbody>
</table>
Pragmatics

- The study of how language is used to accomplish goals, and the influence of context on meaning
- Understanding the aspects of a language which depends on situation and world knowledge

*Do you know what time it is?*
Discourse

- The study of linguistic units larger than a single utterance

John reads a book. He borrowed it from his friend.
Challenges

- Different words/sentences express the same meaning
  - The third season of the year
    - *Fall*
    - *Autumn*
  - Book delivery time
    - *When will my book arrive?*
    - *When will I receive my book?*

- One word/sentence can have different meanings
  - *Fall*
    - The third season of the year
    - Moving down towards the ground or towards a lower position
  - *The door is open.*
    - Expressing a fact
    - A request to close the door
Challenges

- Moving down?
- The third season of the year?
  - What about *autumn*?
Computers can recognize speech

Computers can wreck a nice peach
Morphology

Unionize

Union + ize

Un + ion + ize
I saw the man with a telescope.
I saw the man with a telescope.
The astronomer loves the star.

Movie Star (Celebrity)

Sky Star
Every man loves a woman.

\[ \forall x : (\text{man}(x) \rightarrow \exists y : (\text{woman}(y) \& \text{love}(x, y))) \]

\[ \exists y : (\text{woman}(y) \& \forall x : (\text{man}(x) \rightarrow \text{love}(x, y))) \]
Can you give me the salt?

Would you please give me the salt?

Do you have the ability to give me the salt?
Alice understands that you like your mother, but she ...
Ambiguity

I made her duck.
Outline

1. Administrative Information
2. Introduction
3. NLP Applications
4. NLP Techniques
5. Linguistic Knowledge
6. Challenges
7. Course Materials
Topics

■ Introduction
  □ Introduction to Language Technology
  □ Language Modeling

■ Machine Learning for NLP

■ NLP Techniques

■ NLP Applications
Topics

- Introduction

- Machine Learning for NLP
  - Learning Techniques
  - Classification Algorithms
  - Clustering Algorithms

- NLP Techniques

- NLP Applications
Topics

- Introduction
- Machine Learning for NLP
- NLP Techniques
  - Part Of Speech Tagging
  - Parsing
  - Named Entity Recognition
  - Word Relations
  - Word Sense Disambiguation
  - Semantic Role Labeling
- NLP Applications
Topics

- Introduction
- Machine Learning for NLP
- NLP Techniques
- NLP Applications
  - Text Categorization
  - Information Retrieval
  - Information Extraction
  - Question Answering
  - Sentiment Analysis
  - Summarization
  - Machine Translation
SPEECH AND LANGUAGE PROCESSING

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

by Daniel Jurafsky and James H. Martin

Second Edition
Relevance Journal & Conference

- Journal
  - 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