Lecture 9: Natural Language Processing (NLP) (Introduction)

AI/ML Foundation Course with Python

Natural Language

(http://expertenough.com/2392/german-language-hacks)

日本語で

冬は世界各地でさまざまなお祝いが行われる時期です。ほんのいくつか例を挙げるだけでも、ハナカ、クリスマス、クワンザ、新年などさまざまなお祝いがあります。各文化によってその祝い方はさまざまですが、ほとんどのお祝いにはごちそうが欠かせません。

(http://www.transparent.com/learn-japanese/articles/dec_99.html)

Artificial Language

```
try {
    cMessage = messageQueue.take();
    for (AsyncContext ac : queue) {
        try {
            PrintWriter acWriter = ac.g
            acWriter.println(cMessage);
            acWriter.flush();
        } catch(IO
            System
            queue.
        }
    }
} catch(Interrupte
```

```
append(CharSequenc
append(char c)
append(CharSequenc
format(String form
format(Locale l, S
printf(String form
```

(https://netbeans.org/features/java/)

```
def add5(x):
    return x+5

def dotwrite(ast):
    nodename = getNodename()
    label=symbol.sym_name.get(int(ast[0]),ast[0])
    print '    %s [label="%s' % (nodename,label),
    if isinstance(ast[1], str):
        if ast[1].strip():
            print '= %s"];' % ast[1]
        else:
            print '"]'
    else:
        print '"];'
        children = []
        for n, child in enumerate(ast[1:]):
            children.append(dotwrite(child))
        print '    %s -> {' % nodename,
        for name in children:
            print '%s' % name,
```

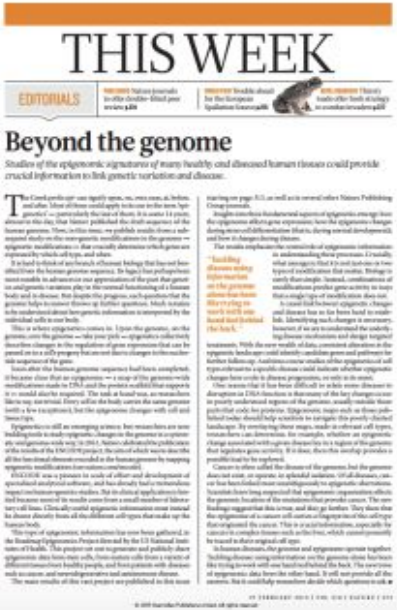
(http://noobite.com/learn-programming-start-with-python/)

Language

A vocabulary consists of a set of words (w_i)

http://learnenglish.britishcouncil.org/en/vocabulary-games)

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



THIS WEEK

EDITORIALS

Beyond the genome

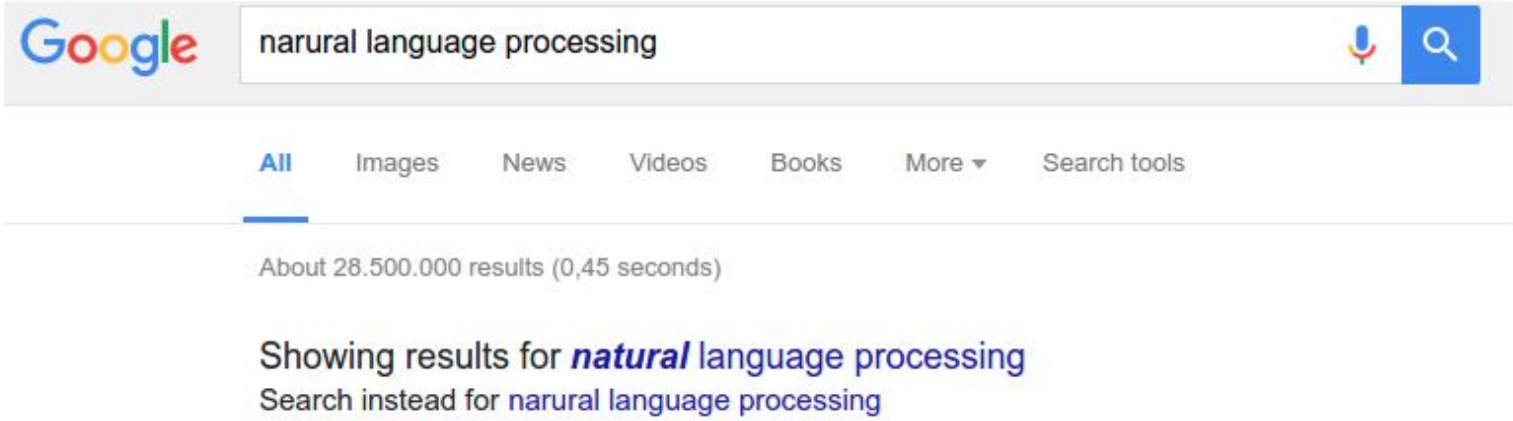
Studies of the epigenomic signatures of many healthy and diseased human tissues could provide crucial information to link genetic variation and disease.

A language is constructed of a set of all possible texts

(http://www.old-engli.sh/language.php)

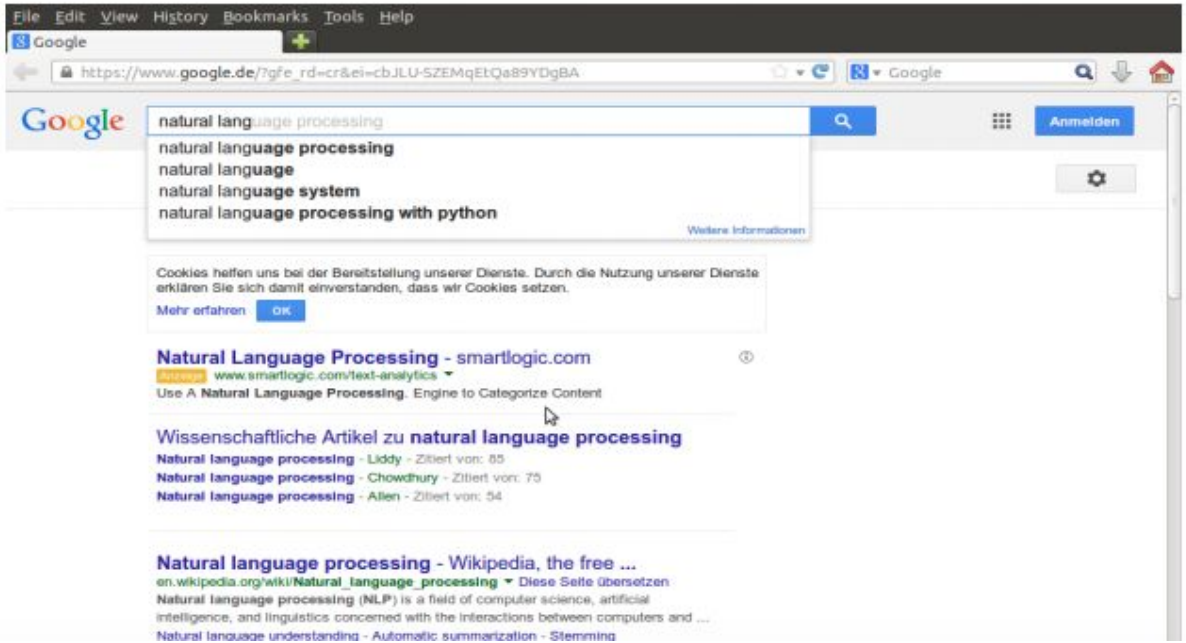
Spell and Grammar Checking

- Checking spelling and grammar
- Suggesting alternatives for the errors



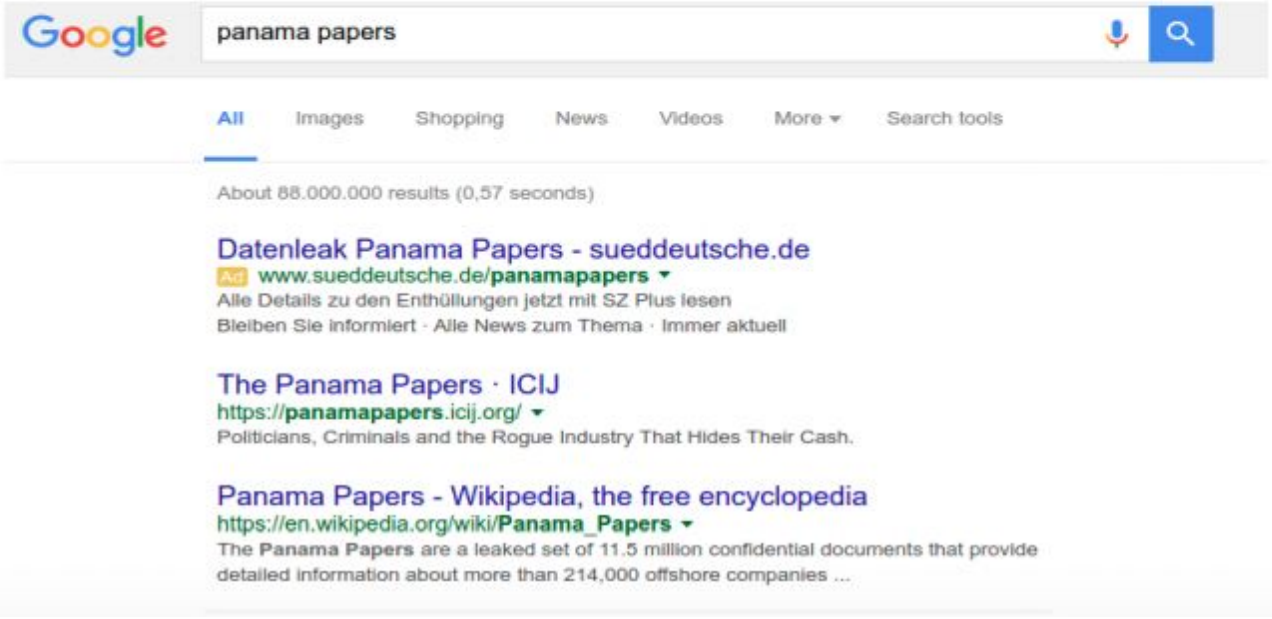
Word Prediction

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



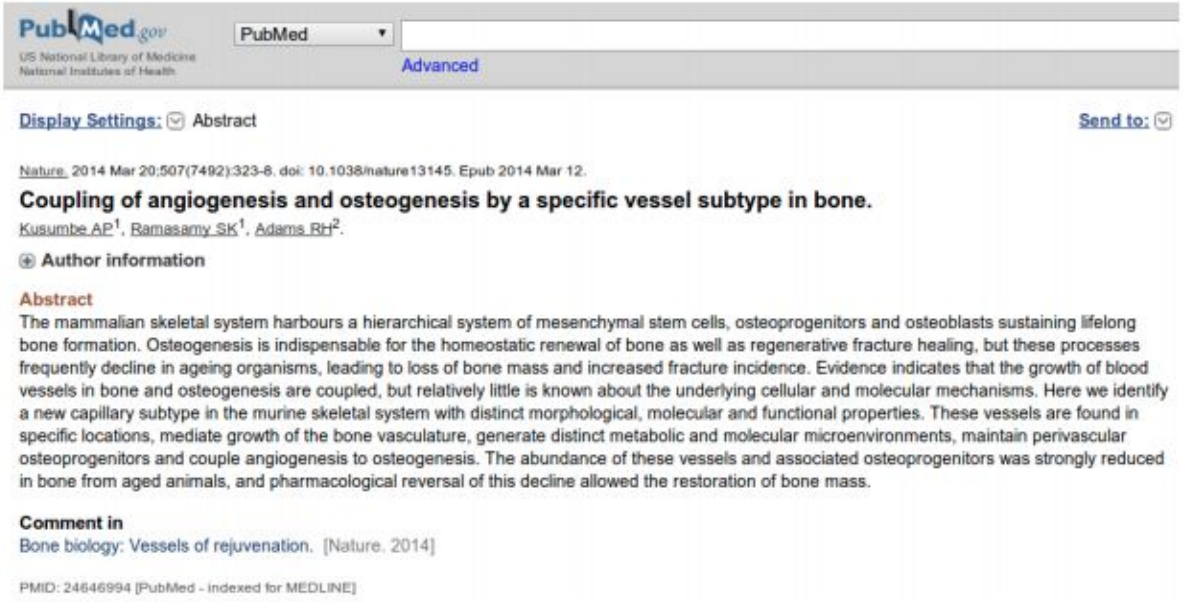
Information Retrieval

- Finding relevant information to the user's query



Text Categorization

- Assigning one (or more) pre-defined category to a text



PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed ▾ Advanced

Display Settings: Abstract Send to:

Nature. 2014 Mar 20;507(7492):323-8. doi: 10.1038/nature13145. Epub 2014 Mar 12.

Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone.

Kusumbe AP[1], Ramasamy SK[1], Adams RH[2].

Author information

Abstract

The mammalian skeletal system harbours a hierarchical system of mesenchymal stem cells, osteoprogenitors and osteoblasts sustaining lifelong bone formation. Osteogenesis is indispensable for the homeostatic renewal of bone as well as regenerative fracture healing, but these processes frequently decline in ageing organisms, leading to loss of bone mass and increased fracture incidence. Evidence indicates that the growth of blood vessels in bone and osteogenesis are coupled, but relatively little is known about the underlying cellular and molecular mechanisms. Here we identify a new capillary subtype in the murine skeletal system with distinct morphological, molecular and functional properties. These vessels are found in specific locations, mediate growth of the bone vasculature, generate distinct metabolic and molecular microenvironments, maintain perivascular osteoprogenitors and couple angiogenesis to osteogenesis. The abundance of these vessels and associated osteoprogenitors was strongly reduced in bone from aged animals, and pharmacological reversal of this decline allowed the restoration of bone mass.

Comment in

Bone biology: Vessels of rejuvenation. [Nature. 2014]

PMID: 24646994 [PubMed - indexed for MEDLINE]

MeSH Terms

Aging/metabolism
Aging/pathology
Animals
Blood Vessels/anatomy & histology
Blood Vessels/cytology
Blood Vessels/growth & development
Blood Vessels/physiology*
Bone and Bones/blood supply*
Bone and Bones/cytology
Endothelial Cells/metabolism
Hypoxia-Inducible Factor 1, alpha Subunit/metabolism
Male
Mice
Mice, Inbred C57BL
Neovascularization, Physiologic/physiology*
Osteoblasts/cytology
Osteoblasts/metabolism
Osteogenesis/physiology*
Oxygen/metabolism
Stem Cells/cytology
Stem Cells/metabolism

Text Categorization

Summarization

- Generating a short summary from one or more documents, sometimes based on a given query

This is a 7 sentence summary of http://hpi.de/en/news/jahrgaenge/2015/des...

Summary processing at low priority, upgrade to BOOST

Design Thinking Week: Students Improve the Daily Life Experience for People with Illiteracies

On the occasion of the World Literacy Day on September 8 more than 40 young innovators applied their Design Thinking skills in order to make life easier for these people.

Here, the focus was especially on the possibilities of using digital technologies and computers to better the daily obstacles in life of the people concerned.

Under the guidance of the D-School's coaches the teams researched, developed and prototyped - and could present many versatile solutions in the end: e.g. one of the groups came up with an idea for a software program that lets internet browsers read texts, functions and links out loud so that people with reading problems can still use news sites or social networks like Facebook.

Question answering

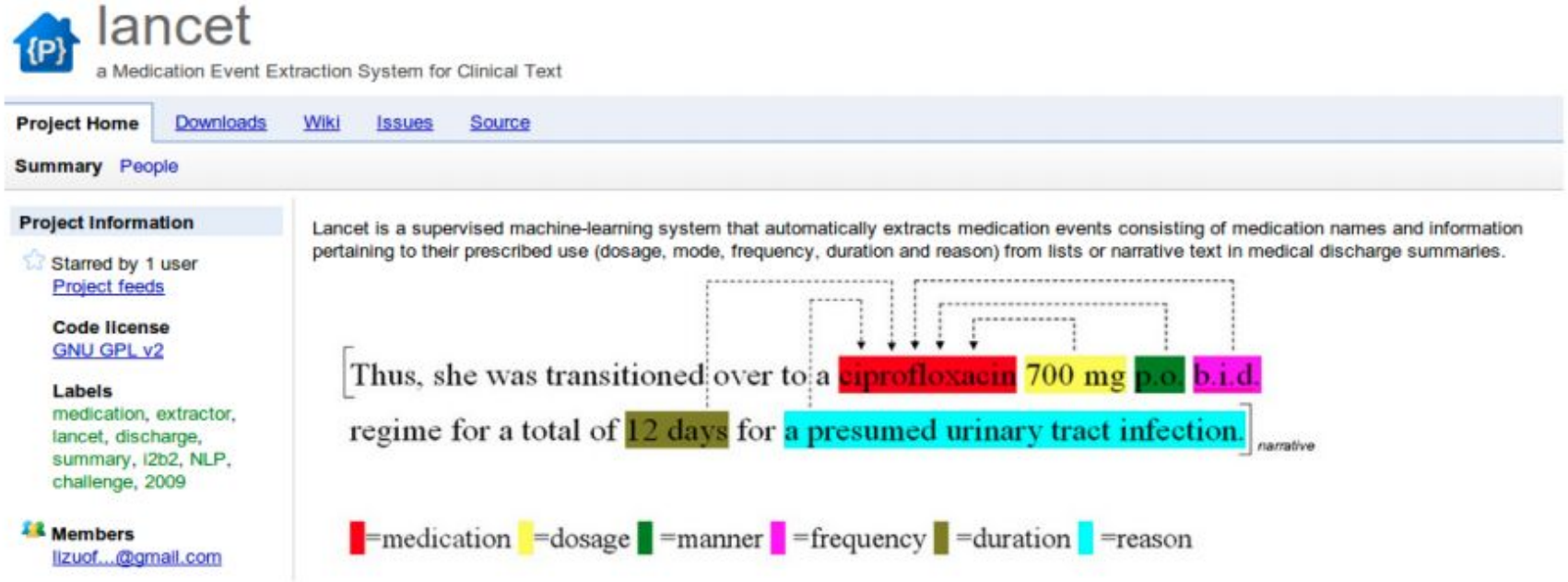
- Answering questions with a short answer

===> what countries speak Spanish

The language Spanish is spoken in Argentina, Aruba, Belize, Bolivia, Brazil, Canada, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Curacao, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Falkland Islands (Islas Malvinas), Gibraltar, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Martin, Sint Maarten, Spain, Switzerland, Trinidad and Tobago, United States, Uruguay, Venezuela, and Virgin Islands.

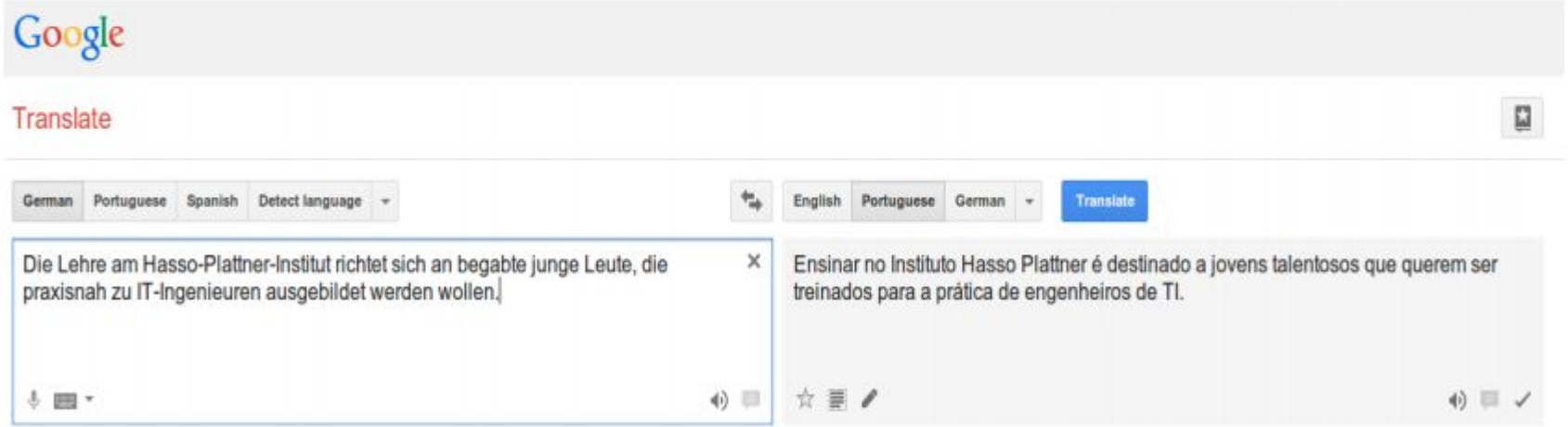
The language Castilian Spanish is spoken in Spain.

Information Extraction



Machine Translation

- Translating a text from one language to another



Sentiment Analysis

- Identifying sentiments and opinions stated in a text

Customer Reviews

Speech and Language Processing, 2nd Edition

15 Reviews

5 star: (8)
4 star: (3)
3 star: (3)
2 star: (0)
1 star: (1)

Average Customer Review

(15 customer reviews)

Share your thoughts with other customers

Create your own review

The most helpful favorable review

4 of 4 people found the following review helpful

Great introductions and reference book

I read the first edition of that book and it is terrific. The second edition is much more adapted to current research. Statistical methods in NLP are more detailed and some syntax-based approaches are presented. My specific interest is in machine translation and dialogue systems. Both chapters are extensively rewritten and much more elaborated. I believe this book is...

Read the full review ›

Published on August 9, 2008 by carheg

› See more **5 star**, **4 star** reviews

Vs.

The most helpful critical review

37 of 37 people found the following review helpful

Good description of the problems in the field, but look elsewhere for practical solutions

The authors have the challenge of covering a vast area, and they do a good job of highlighting the hard problems within individual sub-fields, such as machine translation. The availability of an accompanying Web site is a strong plus, as is the extensive bibliography, which also includes links to freely available software and resources.

Now for the...

Read the full review ›

Published on April 2, 2009 by P. Nadkarni

› See more **3 star**, 2 star, **1 star** reviews

Optical Character Recognition

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

Speech recognition

- Recognizing a spoken language and transforming it into a text

Siri.
Your wish is
its command.

Siri lets you use your voice to send messages, schedule meetings, place phone calls, and more. Ask Siri to do things just by talking the way you talk. Siri understands what you say, knows what you mean, and even talks back. Siri is so easy to use and does so much, you'll keep finding more and more ways to use it.

Speech synthesis

- Producing a spoken language from a text

Spoken dialog systems

- Running a dialog between the user and the system

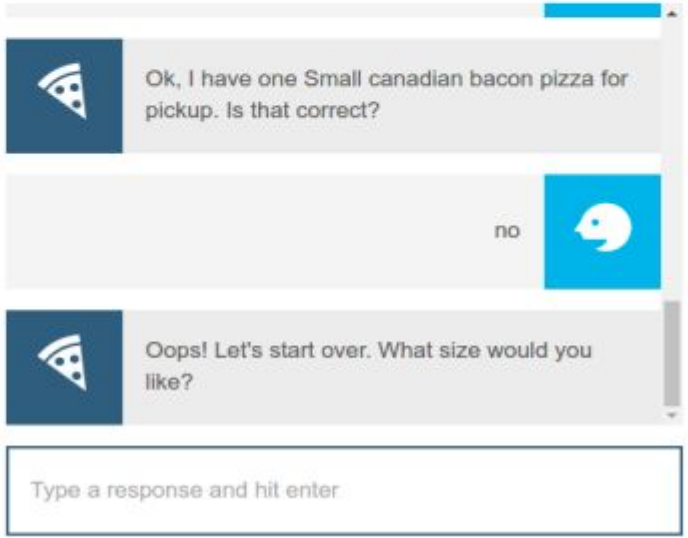
Siri.
Your wish is
its command.

Siri lets you use your voice to send messages, schedule meetings, place phone calls, and more. Ask Siri to do things just by talking the way you talk. Siri understands what you say, knows what you mean, and even talks back. Siri is so easy to use and does so much, you'll keep finding more and more ways to use it.

IBM Watson Developer Cloud

Spoken dialog systems

Try the service



Level of difficulties

- Easy (mostly solved)
  - Spell and grammar checking
  - Some text categorization tasks
  - Some named-entity recognition tasks

Level of difficulties

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

Level of difficulties

- Difficult (still hard)
  - Question answering
  - Summarization
  - Dialog systems

Part-of-speech tagging

- Assigning a syntatic tag to each word in a sentence

Stanford Parser

Please enter a sentence to be parsed:

Surgical resection specimens of 85 invasive ductal carcinomas of 85 women who had undergone 3D ultrasound were included.

Language: English ▾ Sample Sentence Parse

Your query

Surgical resection specimens of 85 invasive ductal carcinomas of 85 women who had undergone 3D ultrasound were included.

Tagging

```
Surgical/NNP resection/NN specimens/NNS of/IN 85/CD invasive/JJ
ductal/JJ carcinomas/NNS of/IN 85/CD women/NNS who/WP had/VBD
undergone/VBN 3D/CD ultrasound/NN were/VBD included/VBN ./.
```

Word sense disambiguation

Analysis with definitions(s)

*Bill Gates has developed an interest/***[readiness to give attention]** *in language technology and yesterday aquired a 10 % interest/***[a share (in a company, business, etc.)]** *in Torbjörn Lager 's sense disambiguation technology . Lager will retain a 90 % interest/***[a share (in a company, business, etc.)]** *in the new company which will be based in Göteborg , Sweden . Last year 's drop in interest/***[money paid for the use of money]** *rates will probably be good for the company . Finally , although all this may sound like an arcane maneuve of little interest/***[quality of causing attention to be given]** *outside Wall Street , it would set off an economical earthquake .*

These are the six senses of the noun *interest* according to the LDOCE:

Sense	Definition
1	readiness to give attention
2	quality of causing attention to be given
3	activity, subject, etc., which one gives time and attention to
4	advantage, advancement, or favour
5	a share (in a company, business, etc.)
6	money paid for the use of money

Paraphrasing

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

Ambiguity

- 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

Semantics

- The astronomer loves the star.
  - Star in the sky
  - Celebrity

Natural Language Understanding

- Input speech get transformed into useful representation in order to analyze the various aspects of language.
- Interpreting a natural language can be very ambiguous.
- Example
  - Lexical Ambiguity
  - Syntactic Ambiguity
  - Referential Ambiguity

Natural Language Generation

- It is the process of converting back a computer generated representation into natural language.
- Example
  - Text Planning
  - Sentence Planning
  - Text Realization

Natural Language Understanding

Lexical Ambiguity

It means Word Level Ambiguity.

{noun} or {verb}

Orange - color/fruit?

Orange (fruit) - noun

Mean - meaning/selfish?

Syntactic Ambiguity

It is about parsing the sentence.

You have orange t-shirt.

You have an orange.

Ambiguity While parsing a sentence

Referential Ambiguity

When meaning is not well referenced from the sentence.

Meera went to Geeta and said, “I am Hungry”

>> Here, who is hungry is not clear from the sentence.

Natural Language Generation

Text Planning

It includes extracting knowledge from knowledge base.

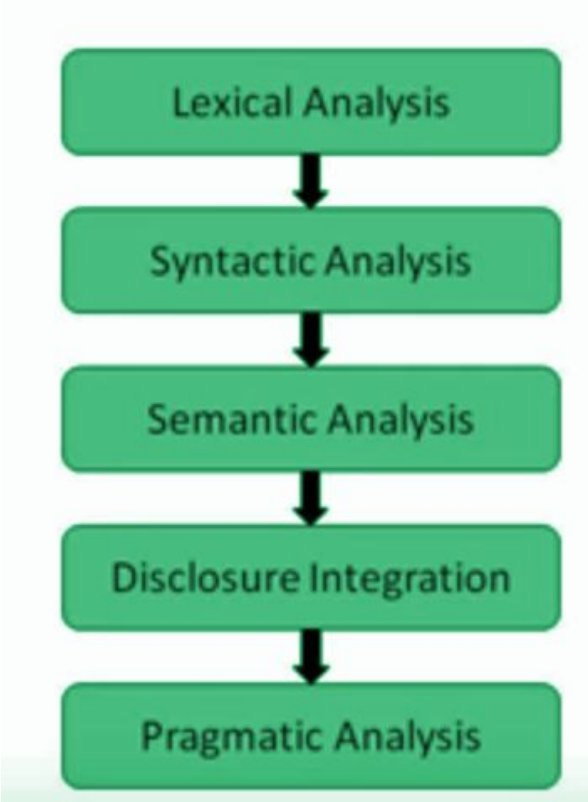
Sentence Planning

It includes selection of correct words and forming sentences which follow the grammar.

Text Realization

Mapping the planned sentence into reality

Steps in NLP


Lexical Analysis
Syntactic Analysis
Semantic Analysis
Disclosure Integration
Pragmatic Analysis

Lexical Analysis

- It deals with the recognition and identification of structure of the sentence
- It divides the paragraph onto sentences,phrases,words.

Syntactic Analysis

- Here the sentences are parsed as noun , verbs , adjectives and other part of sentences.
- Here the grammar of the sentence is analyzed in order to get the relationship among different words in a sentence

Semantic Analysis

- Here the actual meaning of the sentence is extracted from the words used.
- It checks whether the sentence makes any meaning.
- Eg. ‘Bitter sugar’ This is rejected in semantic phase because it doesn’t makes any meaning.

For example, Alex says to Bob, “Call me Cab”

Meaning of the sentence could be -

(1) Call Alex as cab

(2) Call a Cab for Alex (semantically correct)

Disclosure Integration

- Here the meaning of the sentence is verified with sentence before it.

Pragmatic Analysis

- Here the sentences are re-interpreted to verify the correctness of meaning in the given context.
- Here the real world knowledge of language is required.

For example, removing foul language etc.

Pragmatic Analysis

- Natural language generation
- Question answering
- Speech recognisation
- Sentiment analysis

Summary

The statistical approach to natural language is not limited to statistics per-se, but also to advanced inference methods like those used in applied machine learning.

> *… understanding natural language require large amounts of knowledge about morphology, syntax, semantics and pragmatics as well as general knowledge about the world. Acquiring and encoding all of this knowledge is one of the fundamental impediments to developing effective and robust language systems. Like the statistical methods … machine learning methods off the promise of automatic the acquisition of this knowledge from annotated or unannotated language corpora.*

— Page 377, The Oxford Handbook of Computational Linguistics, 2005.


THANKS