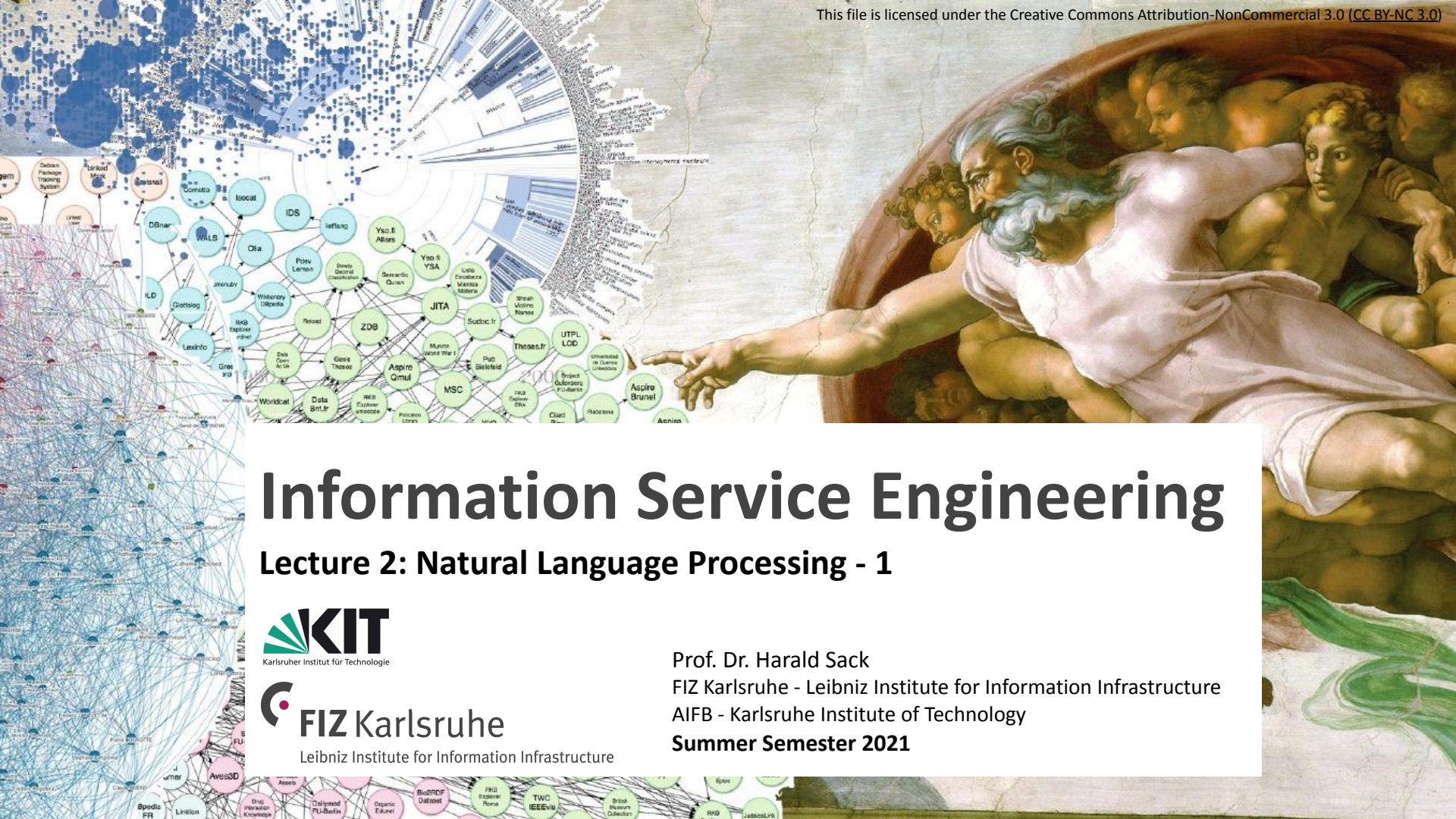


Information Service Engineering

Lecture 2: Natural Language Processing - 1

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

Prof. Dr. Harald Sack
FIZ Karlsruhe - Leibniz Institute for Information Infrastructure
AIFB - Karlsruhe Institute of Technology

Summer Semester 2021

Information Service Engineering

Last Lecture: Information, Natural Language, and the Web

1.1 How to get Information (from the Web)?

1.2 Communication, Language, and Understanding

1.3 How to measure Information?

1.4 The ever-growing Web of Information

1.5 Search Engines on the Web

1.6 The Meaning of Information

Information Service Engineering

Lecture 2: Natural Language Processing (1)

Karlsruher Institut für Technologie

FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

2.0 What is Natural Language Processing?

2.1 NLP and Basic Linguistic Knowledge

2.2 Morphology

2.3 NLP Applications

2.4 NLP Techniques

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

2.10 Language Model and N-Grams

2.11 Part-of-Speech Tagging

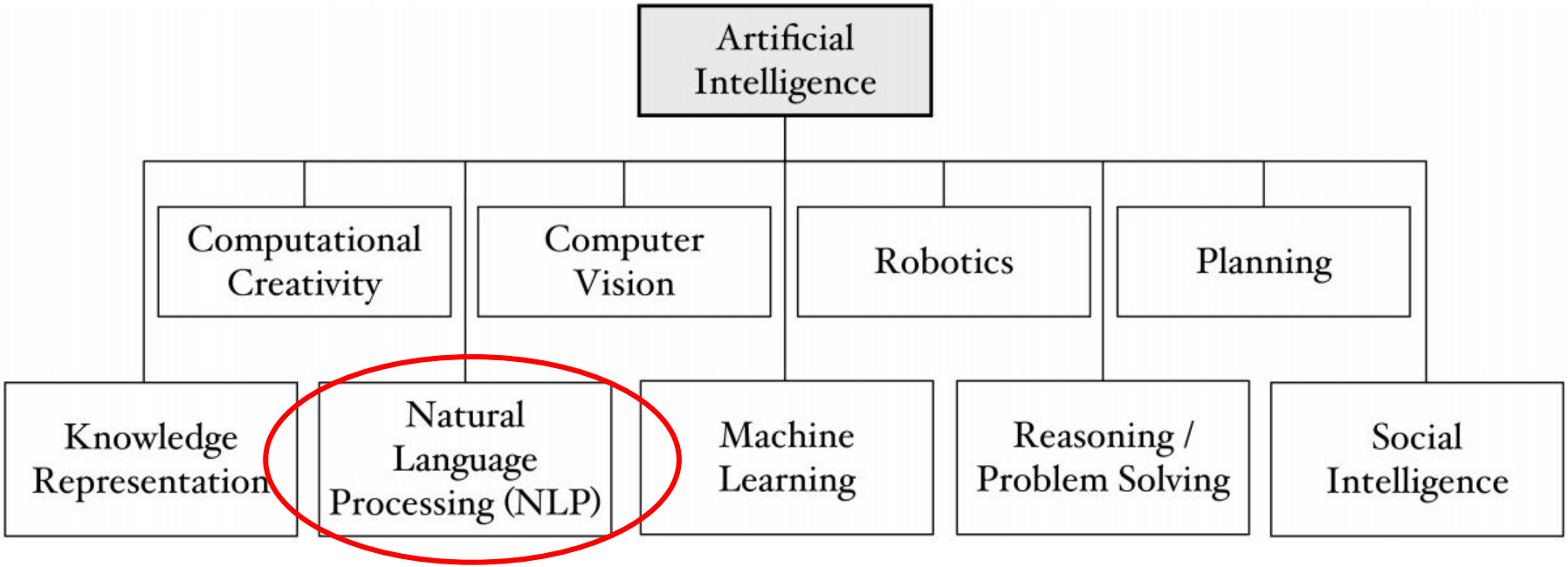
2.12 Word Embeddings

Karlsruher Institut für Technologie
FIZ Karlsruhe
Leibniz Institute for Information Infrastructure

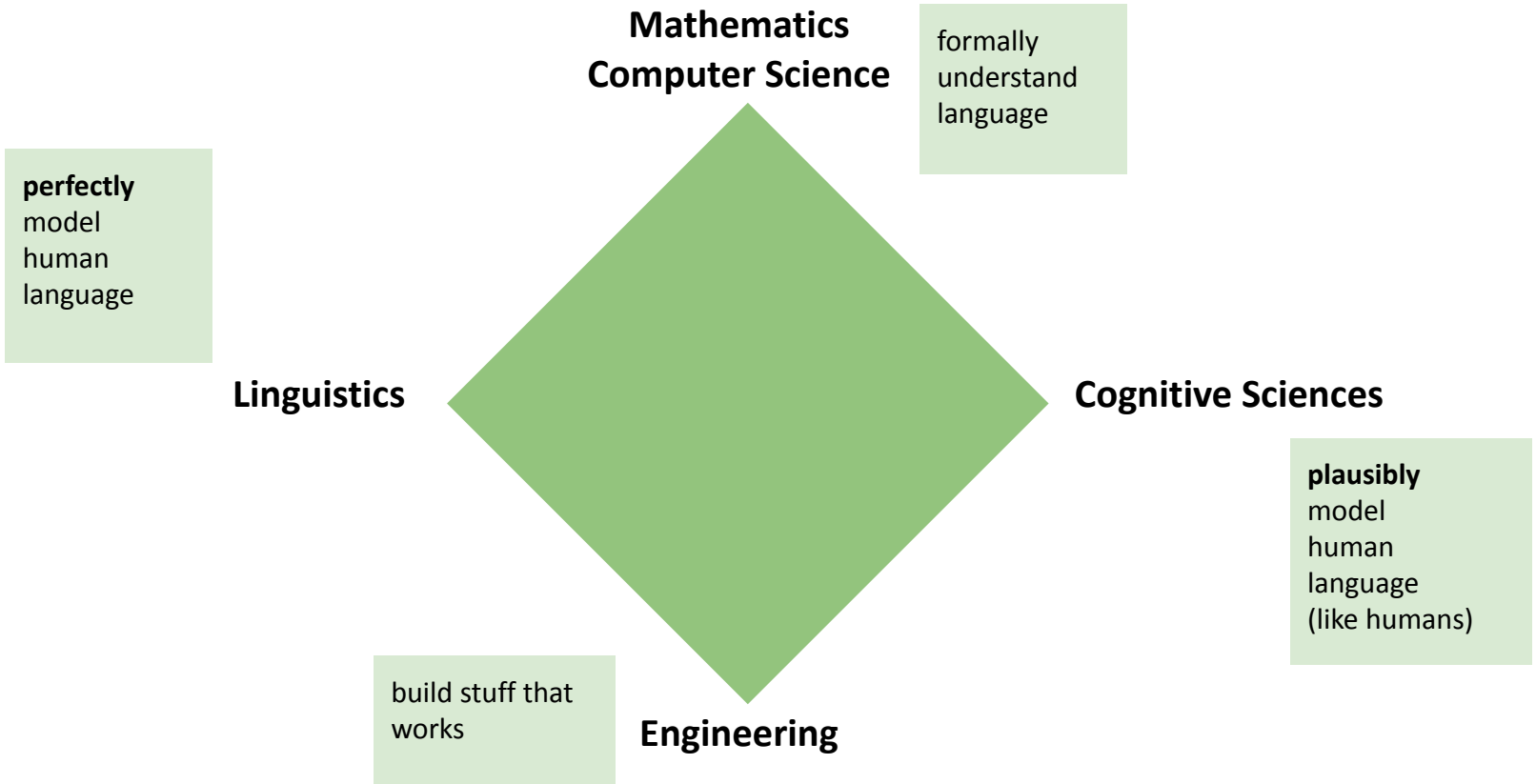
Natural Language Processing (NLP)

- is a field of computer science, artificial intelligence, and computational linguistics and
- is concerned with the **interactions between computers and human (natural) languages** and, in particular,
- is concerned with **programming computers to fruitfully process large natural language corpora**.
- Specifically the task **to extract meaningful information from natural language input** or **to produce natural language output.**

Positioning of NLP in Artificial Intelligence



NLP is an Interdisciplinary Science

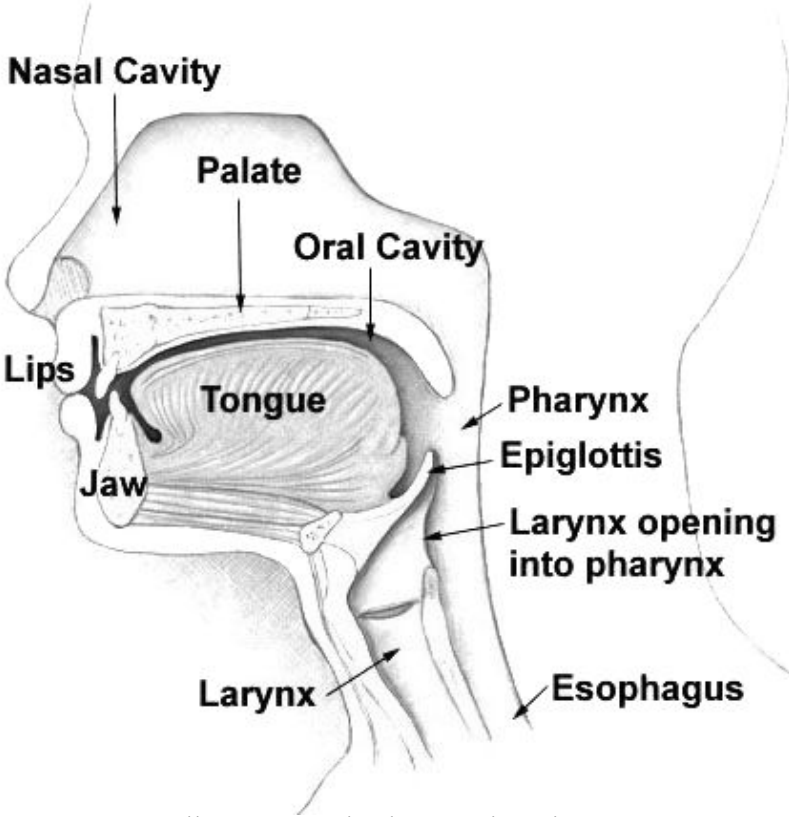


Natural Languages

- are mainly built on three different knowledge components:
  - **Phonology**: the **sound** of words
  - **Semantics**: the **meaning** of words
  - **Syntax**: the **grammatical rules** according to which words are put together

Phonetics vs Phonology

- **Phonetics** is a branch of linguistics that comprises
  - the **study of the sounds of (all) human speech**
  - and is concerned with the **physical properties of speech sounds**, i.e.
    - their physiological production,
    - acoustic properties,
    - auditory perception, and
    - neurophysiological status.

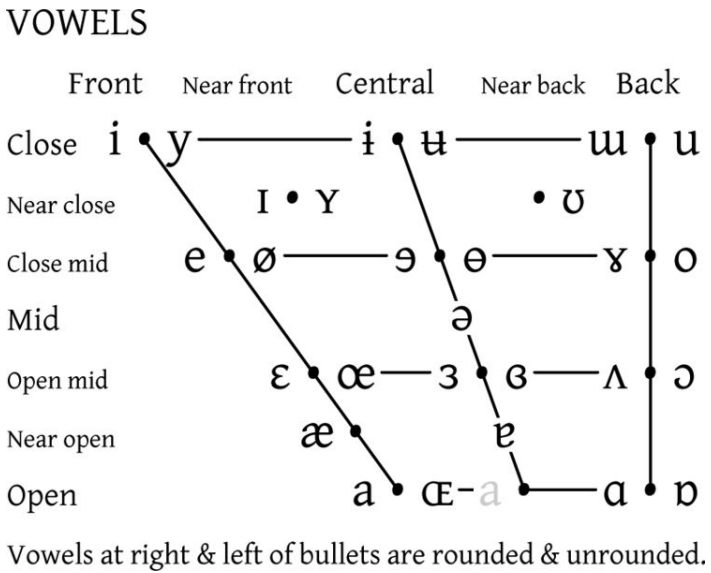


https://en.wikipedia.org/wiki/Language#/media/File:Illu01_head_neck.jpg

William J. Idsardi, Philip J. Monahan, Neurobiology of Language, 2016

Phonetics vs Phonology

- **Phonology** is defined as
  - **the study of speech sounds of a language or languages,**
  - **and the laws governing them**,
  - particularly the laws governing the **composition and combination of speech sounds in language.**



William J. Idsardi, Philip J. Monahan, Neurobiology of Language, 2016

Information Service Engineering

Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?

2.1 NLP and Basic Linguistic Knowledge

2.2 Morphology

2.3 NLP Applications

2.4 NLP Techniques

2.5 NLP Challenges

2.6 Evaluation, Precision and Recall

2.7 Regular Expressions

2.8 Finite State Automata

2.9 Tokenization

2.10 Language Model and N-Grams

2.11 Part-of-Speech Tagging

2.12 Word Embeddings

Phonology Basics

Phone

- **Any distinct speech sound**, regardless of whether the exact sound is critical to the meanings of words

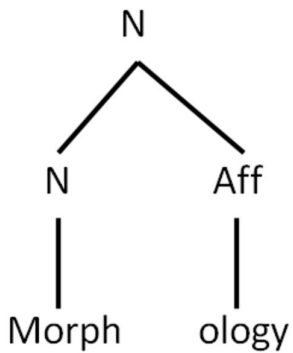
Phoneme

- **Smallest (abstract cognitive) sound unit** in a language that is able of **conveying a distinct meaning**
- **Example:**
  - “*s*” and “*r*” in “*sing*” and “*ring*”
  - “*ss*” and “*ll*” in “*kiss*” and “*kill*”

Morphology

- Morphology is the study of **internal structures (formation) of words** and **how they can be modified**.
- Morphology determines how to **parse complex words into their components**.

⇨ **What is a word?**



Linguistic Basics

A **word** (w_i) is the smallest **independent** unit of language.

“Independent”?

- do not depend on other words
- can be separated from other units
- can change position

https://pixabay.com/p-2041816/

Linguistic Basics

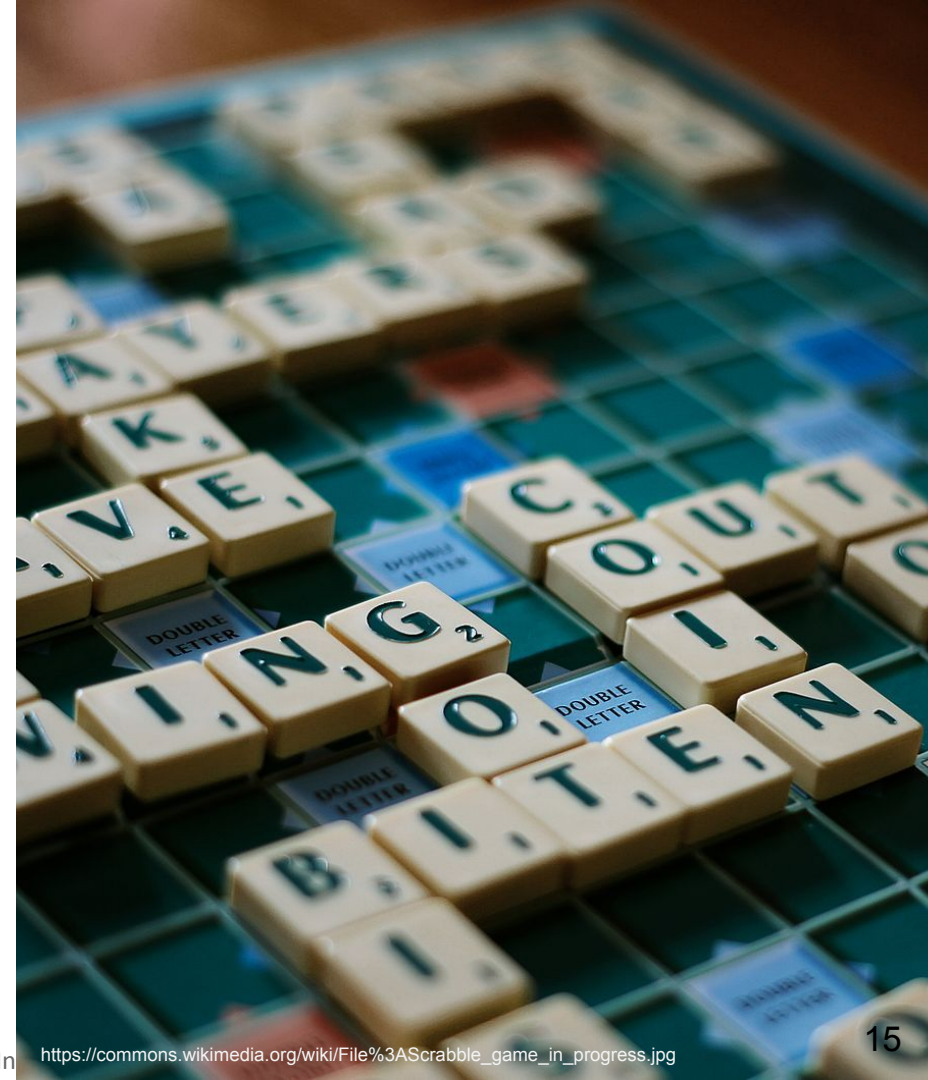
- **Example:**
  The man looked at the horses.

- ***s*** is the plural marker, dependent on the noun ***horse*** to receive meaning.
- ***Horses*** is a word: it can occur in other positions or stand on its own.

https://pixabay.com/p-2041816/

Linguistic Basics

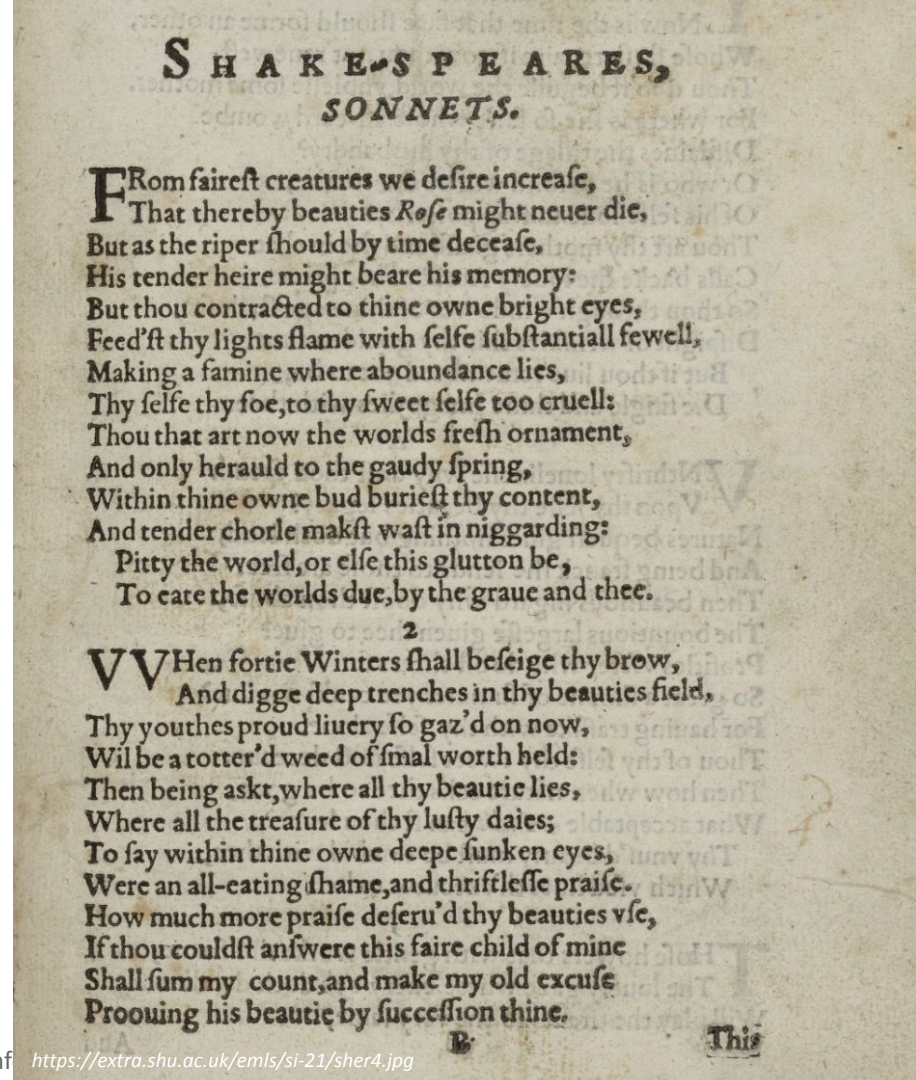
A **vocabulary** consists of a set of **words** (w_i).



https://commons.wikimedia.org/wiki/File%3AScrabble_game_in_progress.jpg

Linguistic Basics

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



SHAKE-SPEARES,
SONNETS.

FRom faireſt creatures we deſire increaſe,
That thereby beauties *Roſe* might neuer die,
But as the riper ſhould by time deceaſe,
His tender heire might beare his memory:
But thou contracted to thine owne bright eyes,
Feed'ſt thy lights flame with ſelfe ſubſtantiall fewell,
Making a famine where aboundance lies,
Thy ſelfe thy foe, to thy ſweet ſelfe too cruell:
Thou that art now the worlds freſh ornament,
And only herauld to the gaudy ſpring,
Within thine owne bud burieſt thy content,
And tender chorle makſt waſt in niggarding:
Pitty the world, or elſe this glutton be,
To eate the worlds due, by the graue and thee.

2

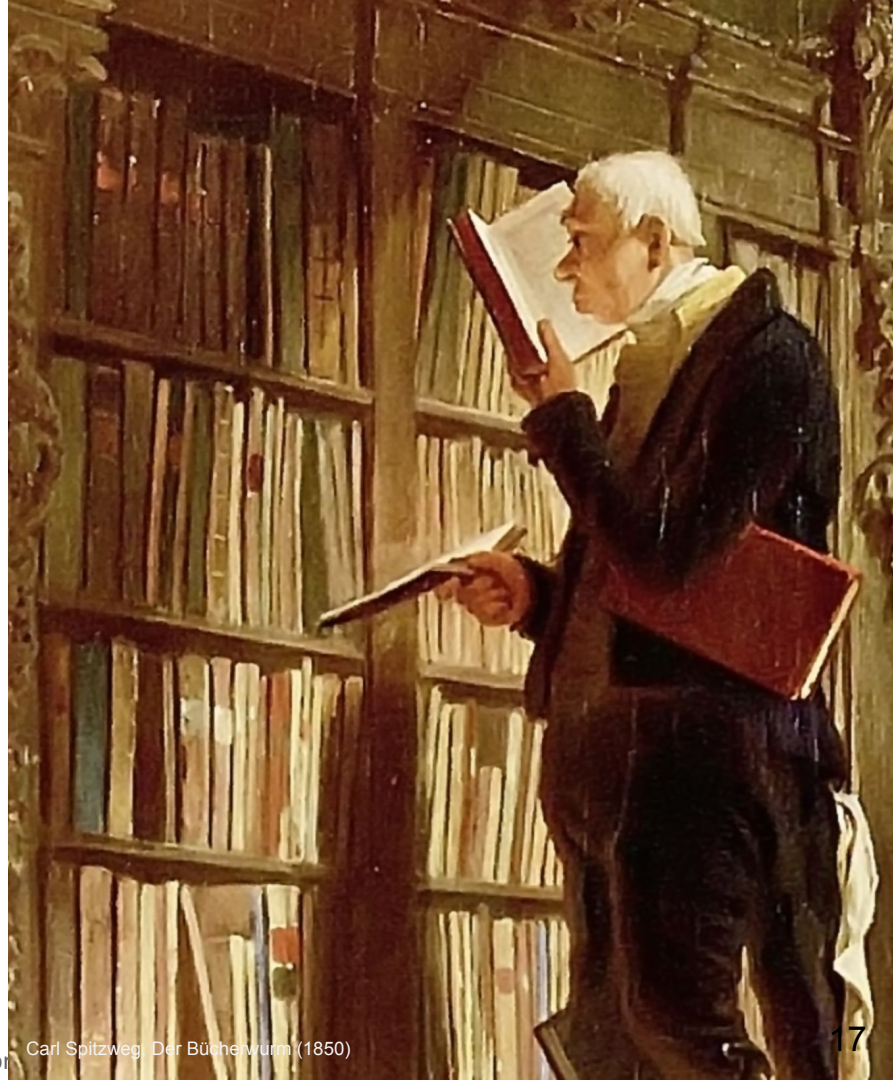
VVHen fortie Winters ſhall beſeige thy brow,
And digge deep trenches in thy beauties field,
Thy youthes proud liuery ſo gaz'd on now,
Wil be a totter'd weed of ſmal worth held:
Then being askt, where all thy beautie lies,
Where all the treaſure of thy luſty daies;
To ſay within thine owne deepe ſunken eyes,
Were an all-eating ſhame, and thriftleſſe praiſe.
How much more praiſe deſeru'd thy beauties vſe,
If thou couldſt anſwere this faire child of mine
Shall ſum my count, and make my old excuſe
Proouing his beautie by ſucceſſion thine.

B This

https://extra.shu.ac.uk/emls/si-21/sher4.jpg

Linguistic Basics

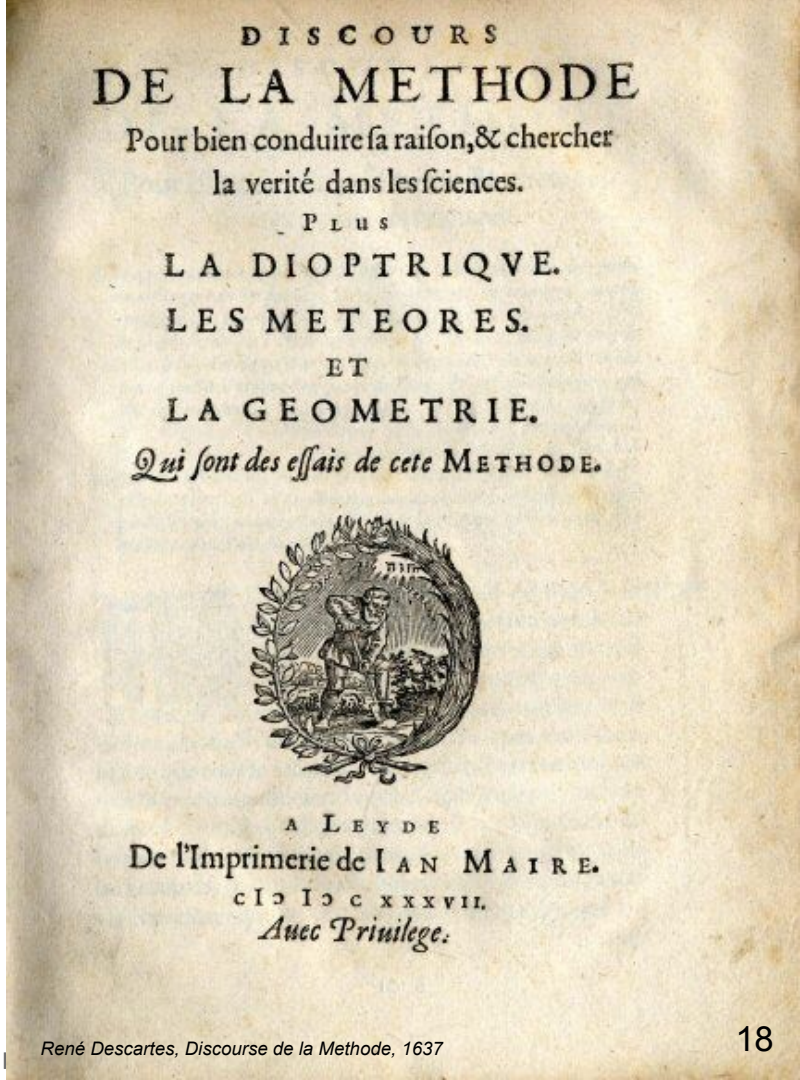
A **language** is
constructed of a
set of all possible texts.



Carl Spitzweg, Der Bücherwurm (1850)

Linguistic Basics

A **discourse** is
the study of linguistic units
larger than a single statement
(i.e. a coherent sequences of sentences).


DISCOURS
DE LA METHODE
Pour bien conduire ſa raiſon,& chercher
la verité dans les ſciences.
PLUS
LA DIOPTRIQVE.
LES METEORES.
ET
LA GEOMETRIE.
Qui ſont des eſſais de cete METHODE.
A LEYDE
De l'Imprimerie de IAN MAIRE.
cIↃ IↃ c xxxvii.
Auec Priuilege.

René Descartes, Discourse de la Methode, 1637

Information Service Engineering

Lecture 2: Natural Language Processing (1)

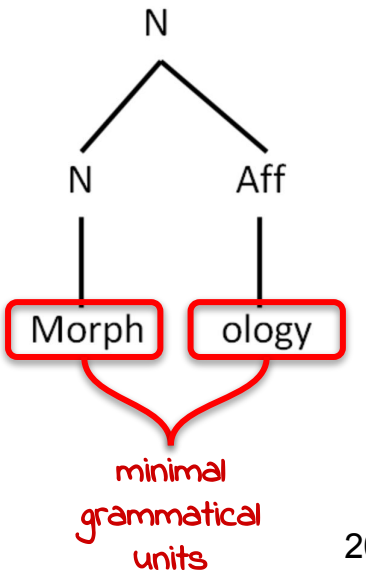
2.0 What is Natural Language Processing?
2.1 NLP and Basic Linguistic Knowledge
2.2 Morphology
2.3 NLP Applications
2.4 NLP Techniques
2.5 NLP Challenges
2.6 Evaluation, Precision and Recall
2.7 Regular Expressions
2.8 Finite State Automata
2.9 Tokenization
2.10 Language Model and N-Grams
2.11 Part-of-Speech Tagging
2.12 Word Embeddings

Morphology

Morphemes

- The **smallest grammatical unit** in a language,
  i.e. the smallest meaningful unit of a language

- We distinguish:
  - **Simple words**: consist of a single morpheme
    E.g. *work, build, run*, etc.
  - **Complex words**: have internal structure,
    i.e. consist of 2 or more morphemes
    E.g. *morphology*, **affix** *-ology* added to **root** *morph*



Morphology

Morphemes

- The **sm** .e. the smallest meanir

A bound morpheme that is part of a complex word but does not belong to any lexical category (i.e., is not a verb, a noun, an adjective)

- We distinguish:
  - **Simple words**: consist of a single morpheme
    E.g. *work, build, run*, etc.
  - **Complex words**: have internal structure,
    i.e. consist of 2 or more morphemes
    E.g. *morphology*, **affix** *-ology* added to **root** *morph*

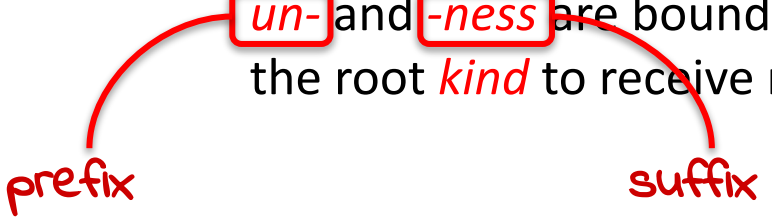
Core part of a complex word, the part that carries the major component of its meaning

Morphology

Free vs. Bound Morphemes

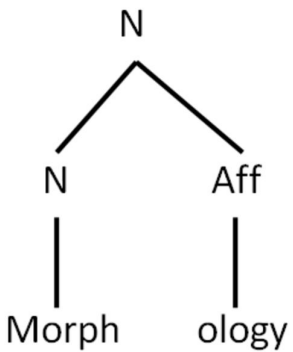
- **Free morpheme**: a simple word, consisting of one morpheme,
  - e.g. *house, work, high, chair, wrap*
- **Bound morpheme**: morphemes that must be attached to another morpheme to receive meaning,
  - e.g. *unkindness*

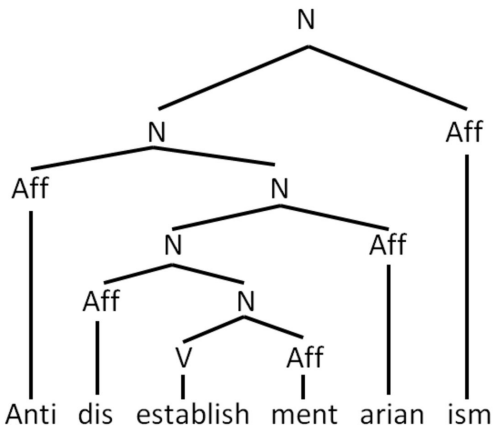
un- and *-ness* are bound morphemes that require the root *kind* to receive meaning



Morphological Parsing

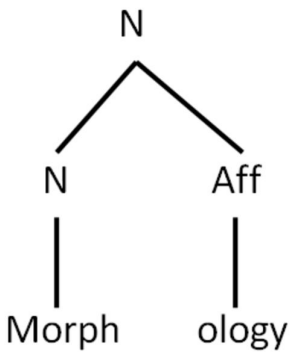
- = the process of determining the morphemes (and their purpose) from which a given word is constructed.
- Can be visualized in a tree diagram (**morphology tree**).



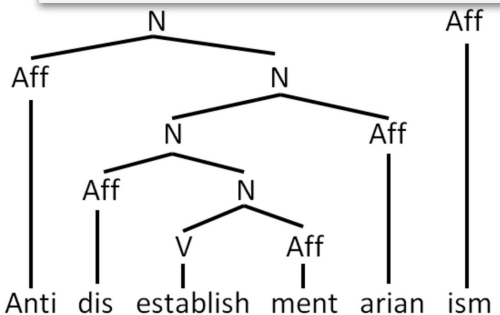


Morphological Parsing

- In many languages, words can be made up of a main **stem** (carrying the basic dictionary meaning) plus one or more **affixes** carrying grammatical information.
- **Surface form:** morphology
- **Lexical form:** morph+N+Aff

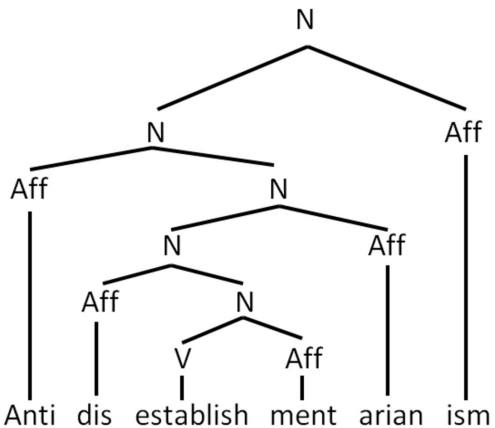


Morphological Parsing is the problem of **extracting the lexical form from the surface form.**



Morphological Parsing - Applications

- Any NLP tasks involving **grammatical parsing** will typically involve morphology parsing as a prerequisite.
- **Search engines**: e.g. a search for *'fox'* should return documents containing *'foxes'*, and vice versa.
- Even a humble task like **spell checking** can benefit: e.g. is "*morphology*" a possible word form?



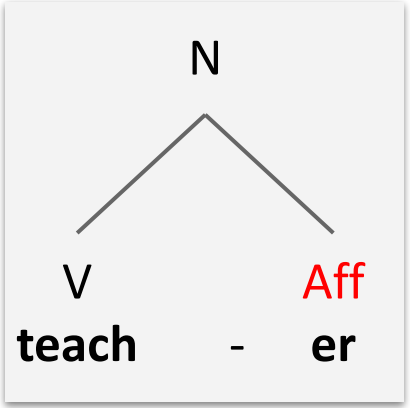
Morphological Rules

- Many languages build more complex words out of morphemes via
  - **Derivation** (with derivational morphemes)
  - **Compound**
  - **Inflection** (with inflectional morphemes)

Derivation

- The process of forming a new word from an existing word by **adding derivational morphemes as affixes**.
- The meaning of the resulting word is different from that of its root.
- Very often there is a change in **word category** involved.

- **Example**
  - *teach* *-er* → *teacher*
    stem **affix** **resulting word**
    verb derivational morpheme noun



Compounding

- **Combination of already existing words** into a new one
- There is no affixation but each of the parts can be assigned to a specific **word category**

- Examples:
  - N + N → N : *lawn mower*
  - P + N → N : *up shot*
  - N + V → V : *blow dry*
  - P + Adj → Adj : *over grown*

Compounding

- **Combination of already existing words** into a new one
- There is no affixation but each of the parts can be assigned to a specific **word category**

- Examples:
  - N + **N** → **N** : *lawn mower*
  - P + **N** → **N** : *up shot*
  - N + **V** → **V** : *blow dry*
  - P + **Adj** → **Adj** : *over grown*

Head

Inflection

- Modification of a word to **indicate aspects of the grammatical function of a word** such as
  tense, **case**, **voice**, **aspect**, **person**, **number**, **gender**, and **mood**.

- In **English, inflection** is predominantly expressed via affixation with **inflectional morphemes**.

Inflection

English has only eight inflectional morphemes:

● noun plural {-s}	*He has three dessert**s**.*
● noun possessive {-s}	*This is Betty’**s** dessert.*
● verb present tense {-s}	*Bill usually eat**s** dessert.*
● verb past tense {-ed}	*He bak**ed** the dessert yesterday.*
● verb past participle {-en}	*He has always eat**en** dessert.*
● verb present participle {-ing}	*He is eat**ing** the dessert now.*
● adjective comparative {-er}	*His dessert is larg**er** than mine.*
● adjective superlative {-est}	*Her dessert is the larg**est**.*

Morphology

Inflection vs. Derivation

- **Derivation** often changes the category of the root, **inflection** never does that.
- **Derivation** changes the meaning of the root, **inflection** does not.
- **Derivation** applies before **inflection**.

Morphology

Inflection vs. Derivation

1. The farmer’s cows escaped.
2. It was raining.
3. Those socks are inexpensive.
4. Jim needs the newer copy.
5. The strongest rower continued.
6. The pitbull has bitten the cyclist.
7. She quickly closed the book.
8. The alphabet-iz-ation went well.

Stemming vs. Lemmatization

- **Stemming**
  The process of reducing inflected or sometimes derived words to their **word stem**
  - *Example:* *cats* → *cat*

- **Lemmatization**
  The process of grouping together the inflected forms of a word so they can be analyzed as a single item, identified by the word's **lemma**, or **dictionary form**
  - *Example:* *better* → *good*

morphological parse of cats: **cat + N + PL**

morphological parse of better: **good + Adj + Com**

surface form

lexical form

Basic Morphology - Summary

- Morphemes - free vs. bound (affixes, suffixes, prefixes)
- Morphological parsing
  - Surface form and lexical form
- Morphological rules
  - Inflection,
  - Derivation, and
  - Compounding
- Stemming vs. Lemmatization

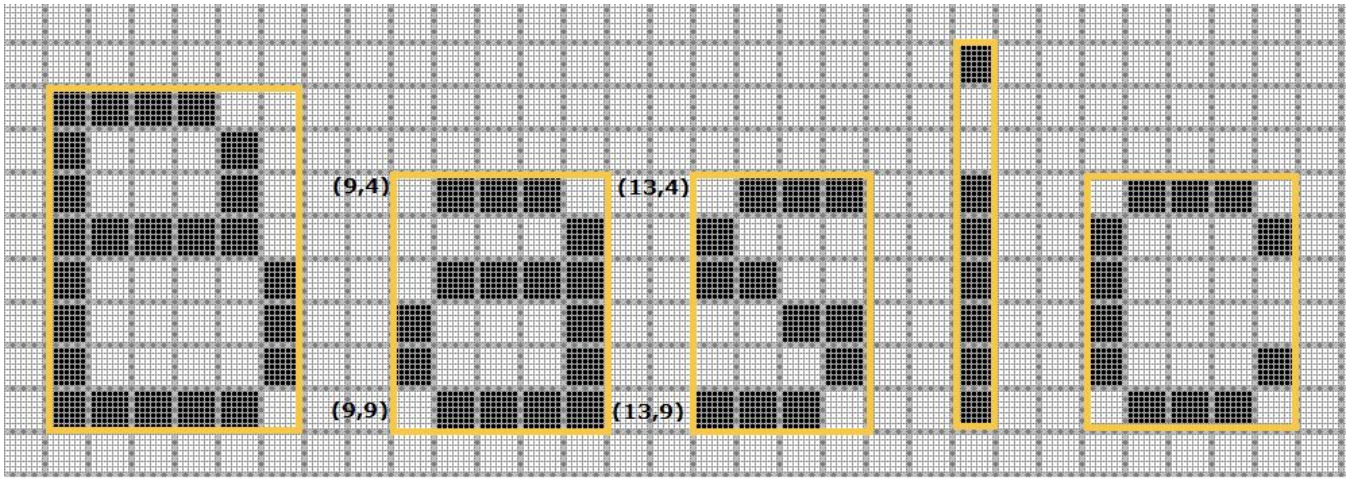
Information Service Engineering

Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?
2.1 NLP and Basic Linguistic Knowledge
2.2 Morphology
2.3 NLP Applications
2.4 NLP Techniques
2.5 NLP Challenges
2.6 Evaluation, Precision and Recall
2.7 Regular Expressions
2.8 Finite State Automata
2.9 Tokenization
2.10 Language Model and N-Grams
2.11 Part-of-Speech Tagging
2.12 Word Embeddings

Optical Character Recognition

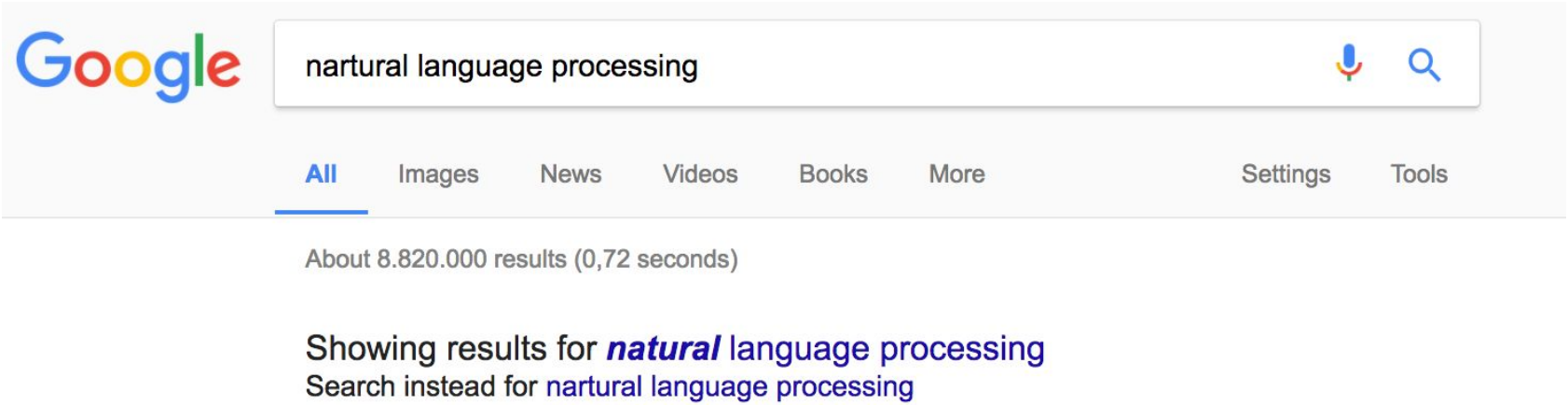
- Conversion of **images** of typed, handwritten or printed text into machine-encoded **text**
- In combination with computer vision, pattern recognition, and artificial intelligence



https://commons.wikimedia.org/wiki/File:BasicBoundry.png

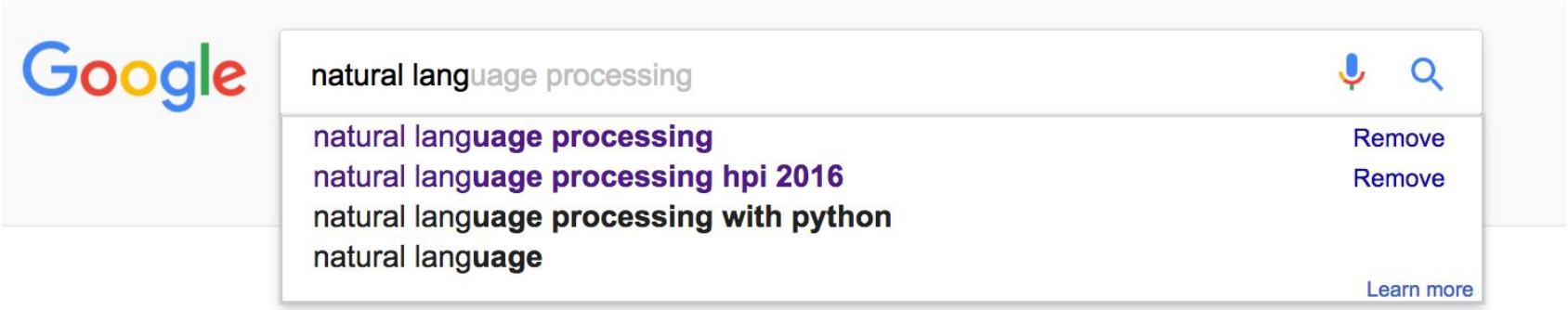
Spelling and Grammar Checking

- Checking spelling and grammar
- Suggesting alternatives for the errors



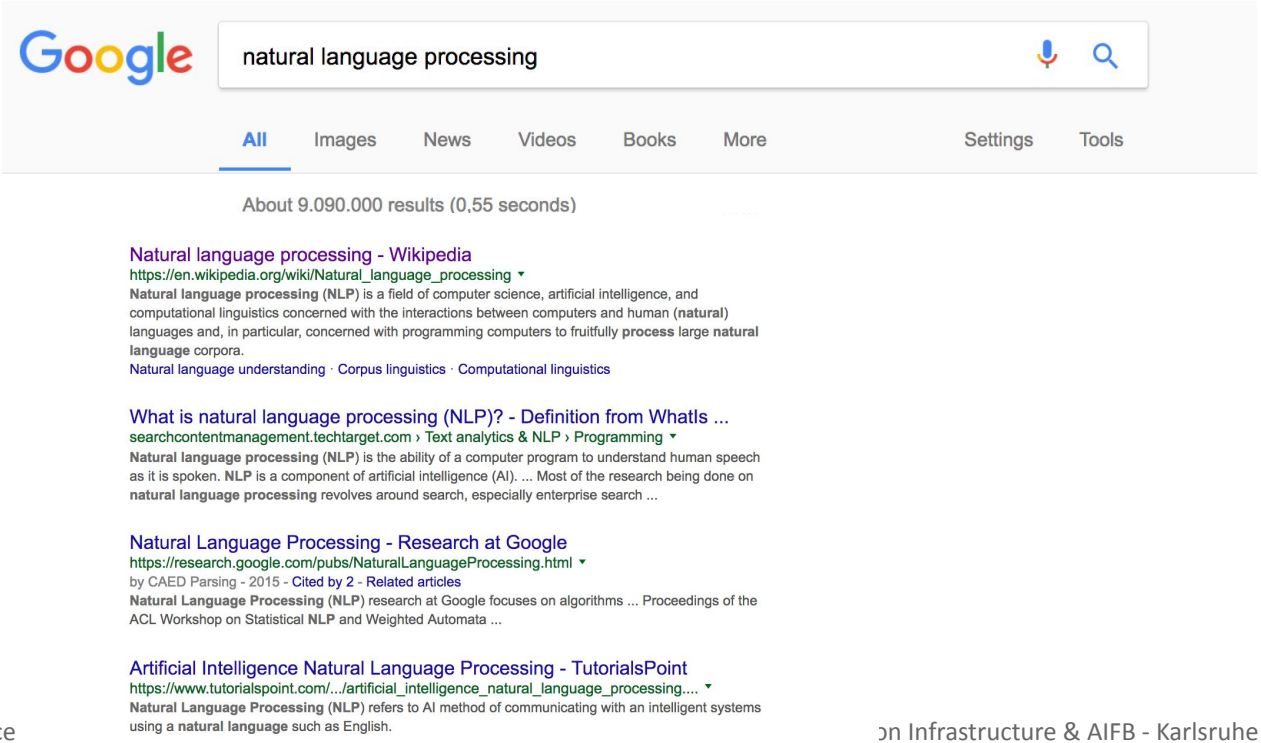
Word Prediction

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



Information Retrieval

- Finding relevant information according to the user's information need expressed in a query



Text Categorization

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

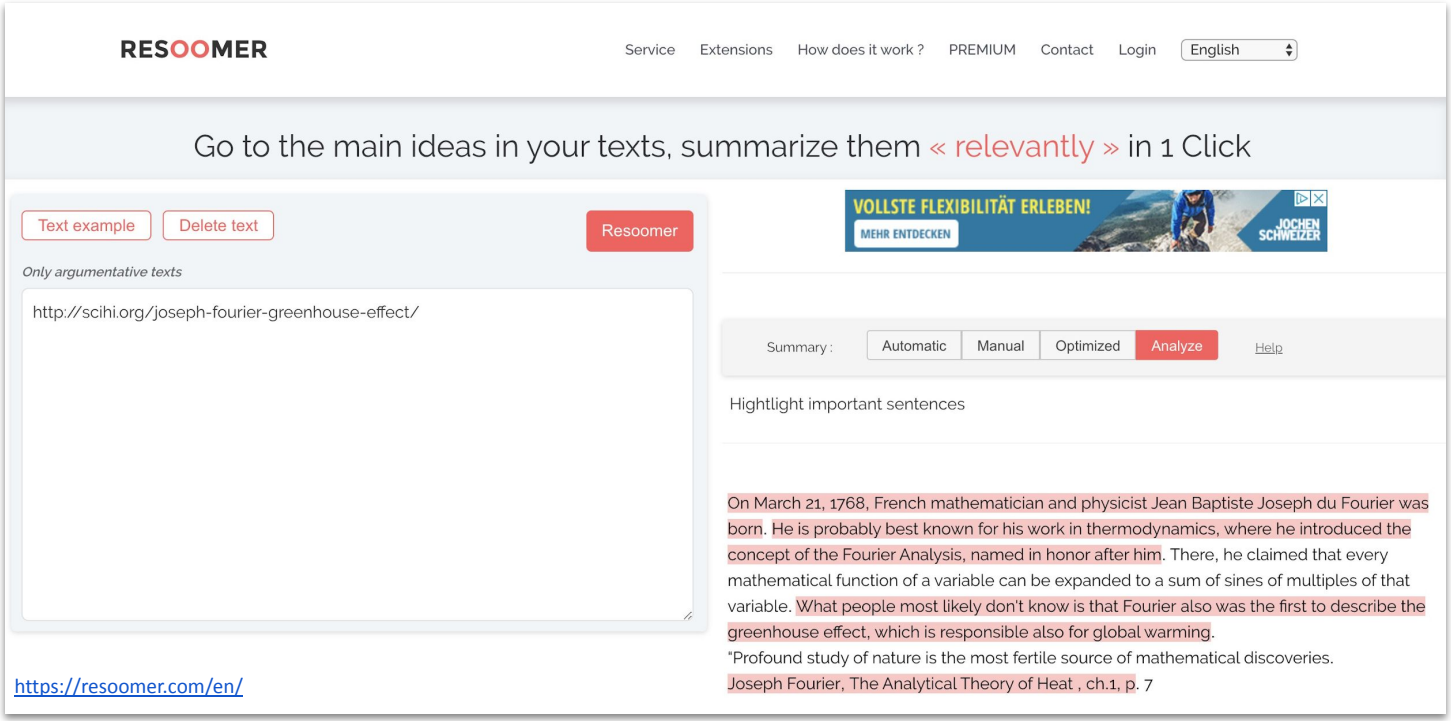
https://upload.wikimedia.org/wikipedia/commons/0/09/Spam_can.png

Text Categorization

https://www.flickr.com/photos/ragesoss/2401468931

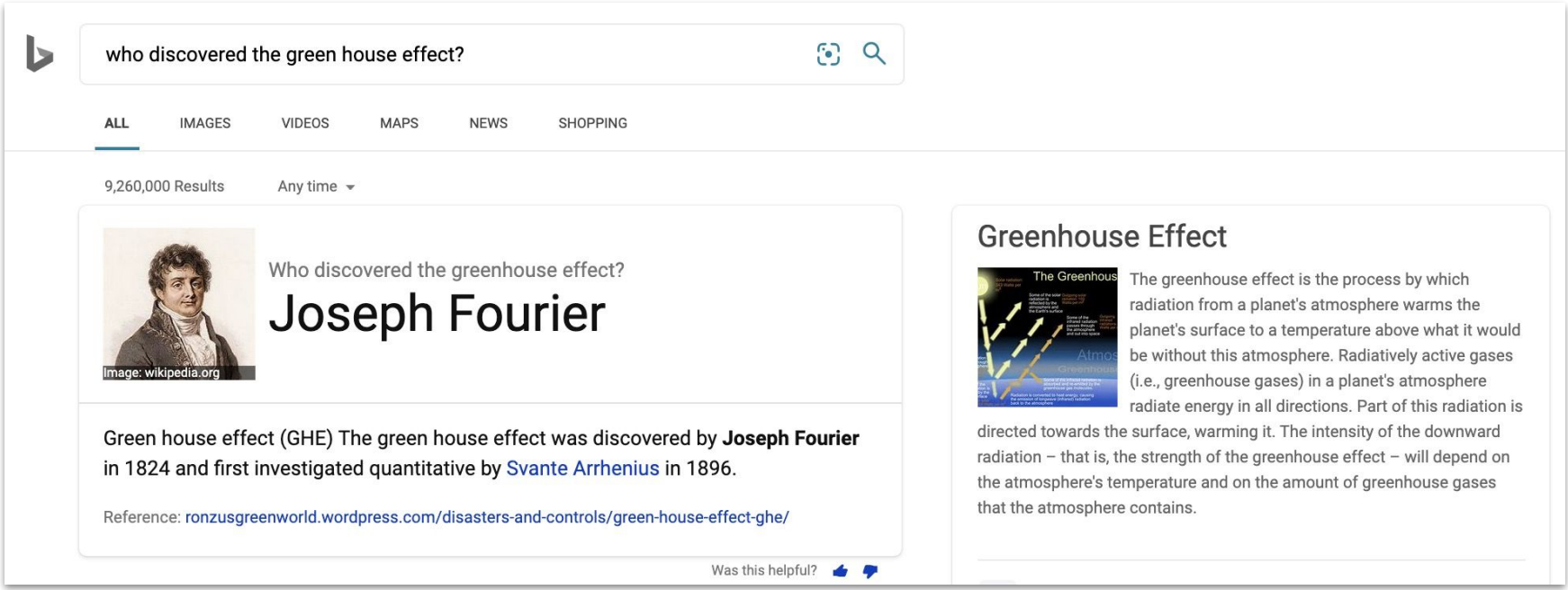
Text Summarization

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



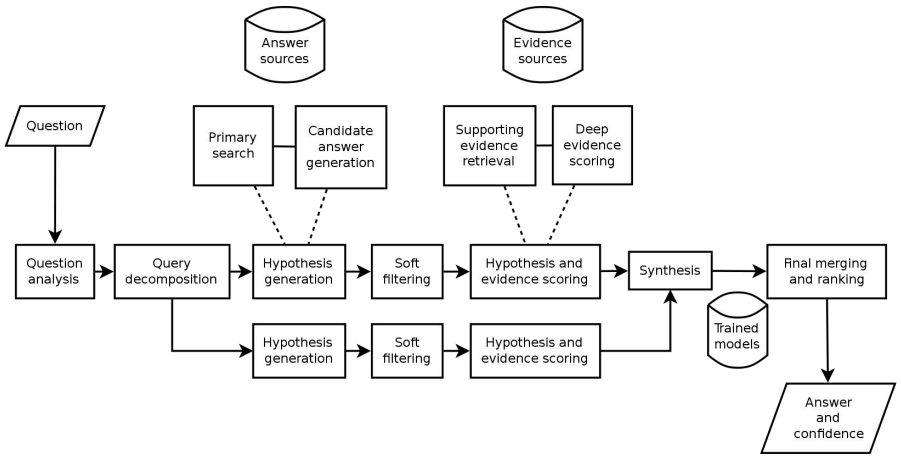
Question Answering

- Automatically answer questions posed by humans in a natural language

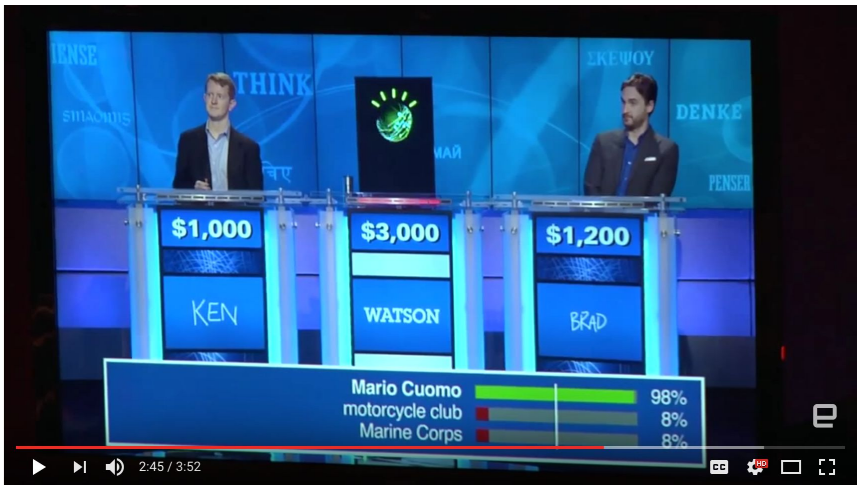


Question Answering

- IBM Watson in Jeopardy!



High-level architecture of IBM's DeepQA used in Watson



https://www.youtube.com/watch?v=WFR3lOm_xhE

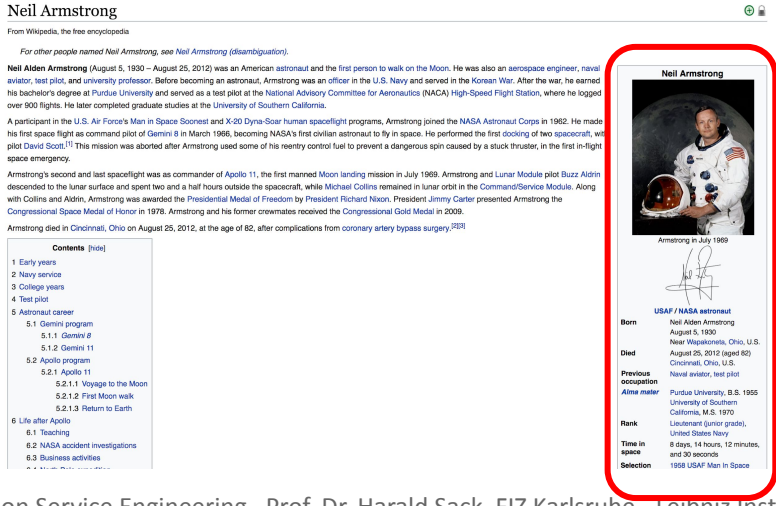
*Ferrucci, D.; et al. (2010). "Building Watson: An Overview of the DeepQA Project". AI Magazine. **31** (3). Retrieved February 19, 2011.*

Information Extraction

- Automatically extracting **structured information** from unstructured and/or semi-structured machine-readable documents

WIKIPEDIA
The Free Encyclopedia

https://en.wikipedia.org/wiki/Neil_Armstrong



Neil Armstrong

From Wikipedia, the free encyclopedia

For other people named Neil Armstrong, see Neil Armstrong (disambiguation).

Neil Alden Armstrong (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California.

A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight as command pilot of Gemini 8 in March 1966, becoming NASA's first civilian astronaut to fly in space. He performed the first docking of two spacecraft, with pilot David Scott.[1] This mission was aborted after Armstrong used some of his reentry control fuel to prevent a dangerous spin caused by a stuck thruster, in the first in-flight space emergency.

Armstrong's second and last spaceflight was as commander of Apollo 11, the first manned Moon landing mission in July 1969. Armstrong and Lunar Module pilot Buzz Aldrin descended to the lunar surface and spent two and a half hours outside the spacecraft, while Michael Collins remained in lunar orbit in the Command/Service Module. Along with Collins and Aldrin, Armstrong was awarded the Presidential Medal of Freedom by President Richard Nixon. President Jimmy Carter presented Armstrong the Congressional Space Medal of Honor in 1978. Armstrong and his former crewmates received the Congressional Gold Medal in 2009.

Armstrong died in Cincinnati, Ohio on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery.[2][3]

Contents [hide]

1 Early years
2 Navy service
3 College years
4 Test pilot
5 Astronaut career
5.1 Gemini program
5.1.1 *Gemini 8*
5.1.2 Gemini 11
5.2 Apollo program
5.2.1 Apollo 11
5.2.1.1 Voyage to the Moon
5.2.1.2 First Moon walk
5.2.1.3 Return to Earth
6 Life after Apollo
6.1 Teaching
6.2 NASA accident investigations
6.3 Business activities

Neil Armstrong

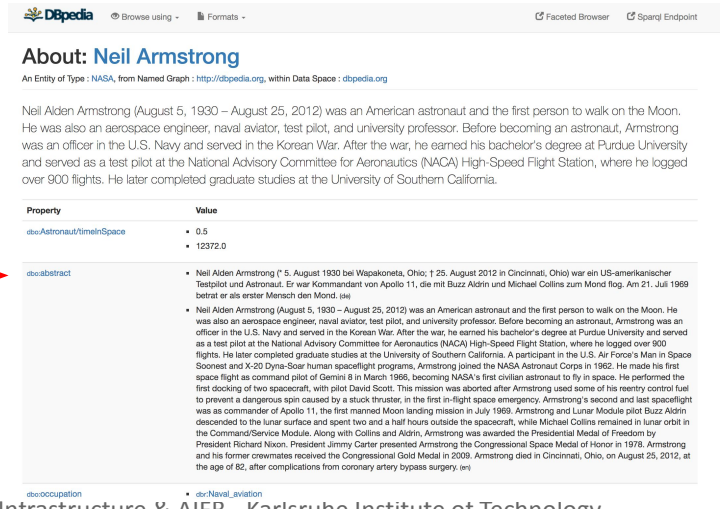
Armstrong in July 1969

USAF / NASA astronaut

Born	Neil Alden Armstrong August 5, 1930 Near Wapakoneta, Ohio, U.S.
Died	August 25, 2012 (aged 82) Cincinnati, Ohio, U.S.
Previous occupation	Naval aviator, test pilot
Alma mater	Purdue University, B.S. 1955 University of Southern California, M.S. 1970
Rank	Lieutenant (junior grade), United States Navy
Time in space	8 days, 14 hours, 12 minutes, and 30 seconds
Selection	1958 USAF Man In Space

DBpedia

http://dbpedia.org/page/Neil_Armstrong



DBpedia | Browse using | Formats | Faceted Browser | Sparql Endpoint

About: Neil Armstrong

An Entity of Type : NASA, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

Neil Alden Armstrong (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California.

Property	Value
dbo:Astronaut/timeInSpace	• 0.5 • 12372.0
dbo:abstract	• Neil Alden Armstrong (* 5. August 1930 bei Wapakoneta, Ohio; † 25. August 2012 in Cincinnati, Ohio) war ein US-amerikanischer Testpilot und Astronaut. Er war Kommandant von Apollo 11, die mit Buzz Aldrin und Michael Collins zum Mond flog. Am 21. Juli 1969 betrat er als erster Mensch den Mond. (de) • Neil Alden Armstrong (August 5, 1930 – August 25, 2012) was an American astronaut and the first person to walk on the Moon. He was also an aerospace engineer, naval aviator, test pilot, and university professor. Before becoming an astronaut, Armstrong was an officer in the U.S. Navy and served in the Korean War. After the war, he earned his bachelor's degree at Purdue University and served as a test pilot at the National Advisory Committee for Aeronautics (NACA) High-Speed Flight Station, where he logged over 900 flights. He later completed graduate studies at the University of Southern California. A participant in the U.S. Air Force's Man in Space Soonest and X-20 Dyna-Soar human spaceflight programs, Armstrong joined the NASA Astronaut Corps in 1962. He made his first space flight as command pilot of Gemini 8 in March 1966, becoming NASA's first civilian astronaut to fly in space. He performed the first docking of two spacecraft, with pilot David Scott. This mission was aborted after Armstrong used some of his reentry control fuel to prevent a dangerous spin caused by a stuck thruster, in the first in-flight space emergency. Armstrong's second and last spaceflight was as commander of Apollo 11, the first manned Moon landing mission in July 1969. Armstrong and Lunar Module pilot Buzz Aldrin descended to the lunar surface and spent two and a half hours outside the spacecraft, while Michael Collins remained in lunar orbit in the Command/Service Module. Along with Collins and Aldrin, Armstrong was awarded the Presidential Medal of Freedom by President Richard Nixon. President Jimmy Carter presented Armstrong the Congressional Space Medal of Honor in 1978. Armstrong and his former crewmates received the Congressional Gold Medal in 2009. Armstrong died in Cincinnati, Ohio, on August 25, 2012, at the age of 82, after complications from coronary artery bypass surgery. (en)
dbo:occupation	• dbr:Naval_aviation

Entity Suggestion / Autosuggestion

- Predicting a potential knowledge base entity from (ambiguous) text input according to the intention of the user

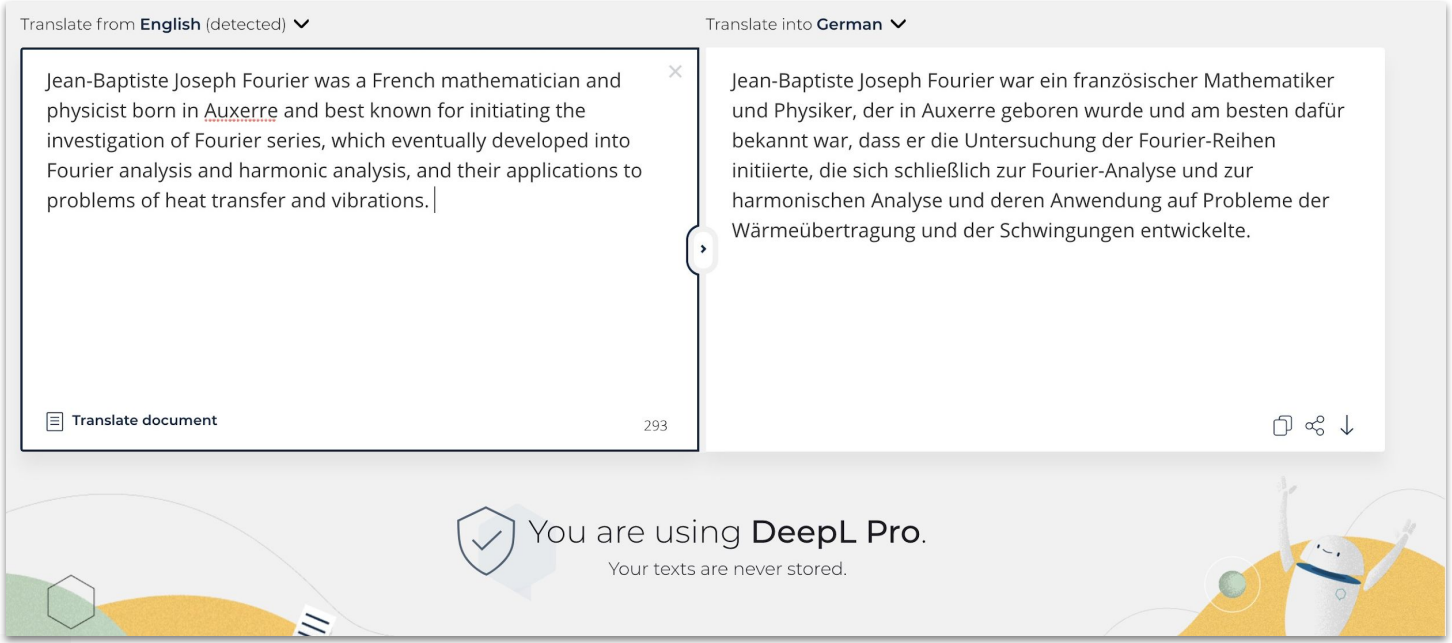
refer

http://refer.cx

http://scihi.org/around-the-world-in-80-days/

Machine Translation

- Translating a text from one language to another



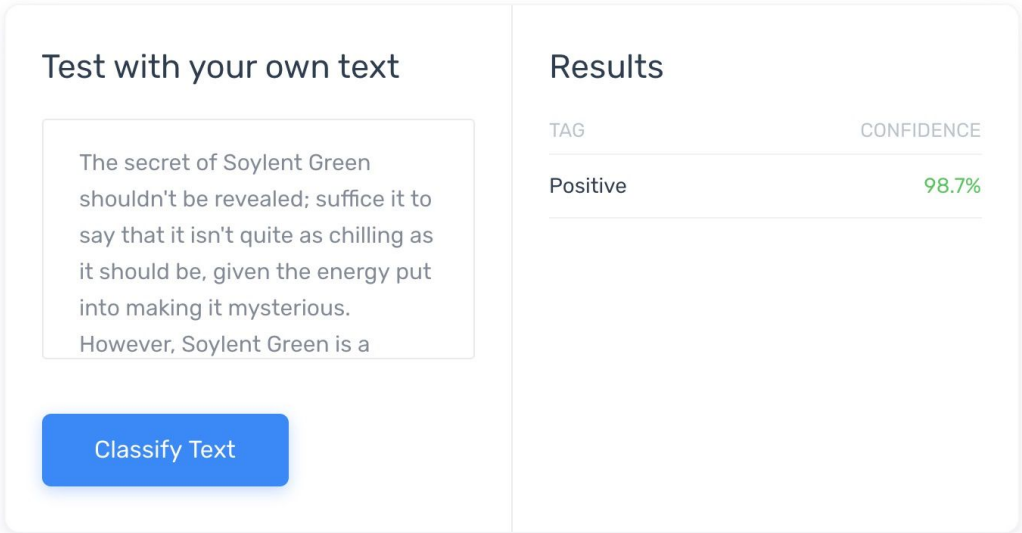
DeepL

https://www.deepl.com/

Sentiment Analysis

- Identifying sentiments and opinions stated in a text

http://rottentomatoes.com/m/soylent_green



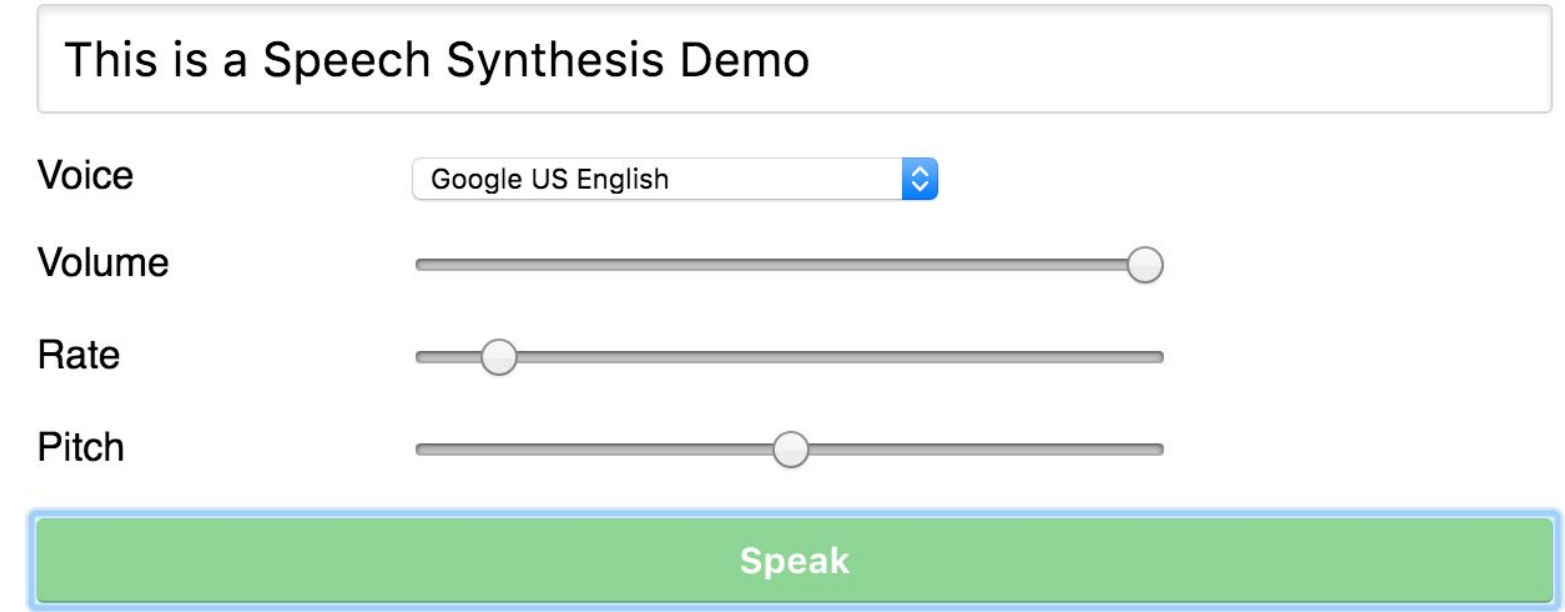
https://monkeylearn.com/sentiment-analysis-online/

Speech Recognition (Speech to Text, STT)

- Recognizing a spoken language and transforming it into a text

Speech Synthesis (Text to Speech, TTS)

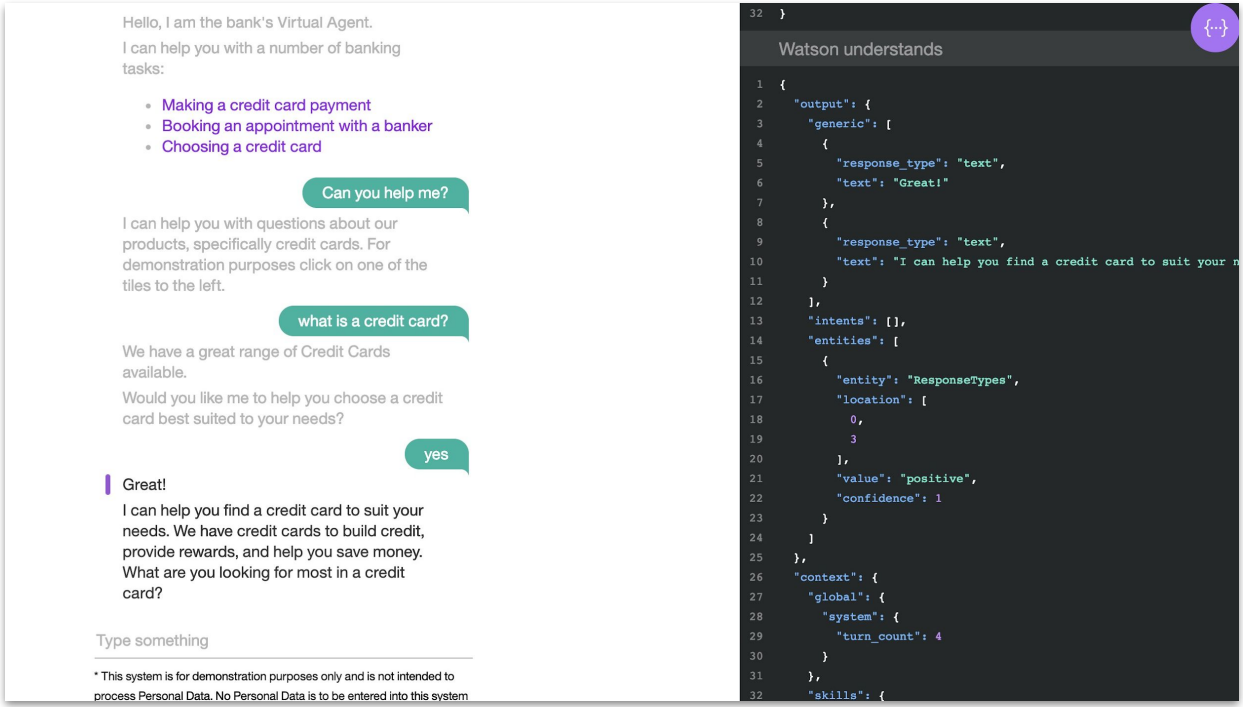
- Producing a spoken language from a text



https://codepen.io/matt-west/pen/wGzuJ

Dialog Systems

- Running a dialog between the user and the system



https://assistant-simple.ng.bluemix.net/

Information Service Engineering

Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?
2.1 NLP and Basic Linguistic Knowledge
2.2 Morphology
2.3 NLP Applications
2.4 NLP Techniques
2.5 NLP Challenges
2.6 Evaluation, Precision and Recall
2.7 Regular Expressions
2.8 Finite State Automata
2.9 Tokenization
2.10 Language Model and N-Grams
2.11 Part-of-Speech Tagging
2.12 Word Embeddings

Sentence Splitting

- Splitting a **text** into **sentences**

Original Text

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born. He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him. There, he claimed that every mathematical function of a variable can be expanded to a sum of sines of multiples of that variable. What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global warming.

Analysis Result

On March 21, 1768, French mathematician and physicist Jean Baptiste Joseph du Fourier was born.

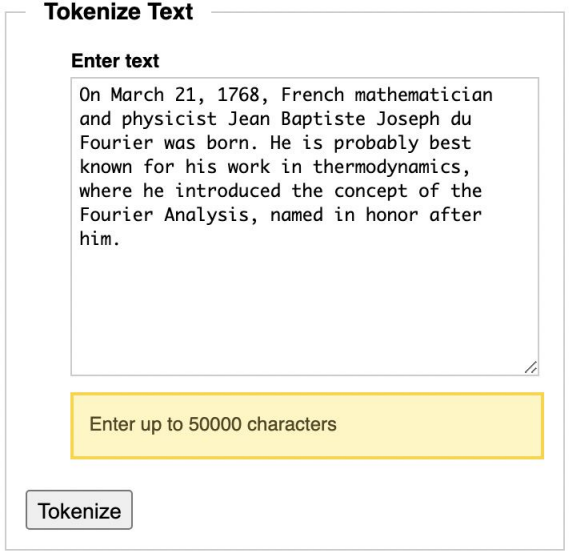
He is probably best known for his work in thermodynamics, where he introduced the concept of the Fourier Analysis, named in honor after him.

There, he claimed that every mathematical function of a variable can be expanded to a sum of sines of multiples of that variable.

What people most likely don't know is that Fourier also was the first to describe the greenhouse effect, which is responsible also for global warming.

http://textanalysisonline.com/nltk-sentence-segmentation

Tokenization



- Splitting a **stream of text** up into words, phrases, symbols, or other meaningful elements called **tokens**

TreebankWordTokenizer

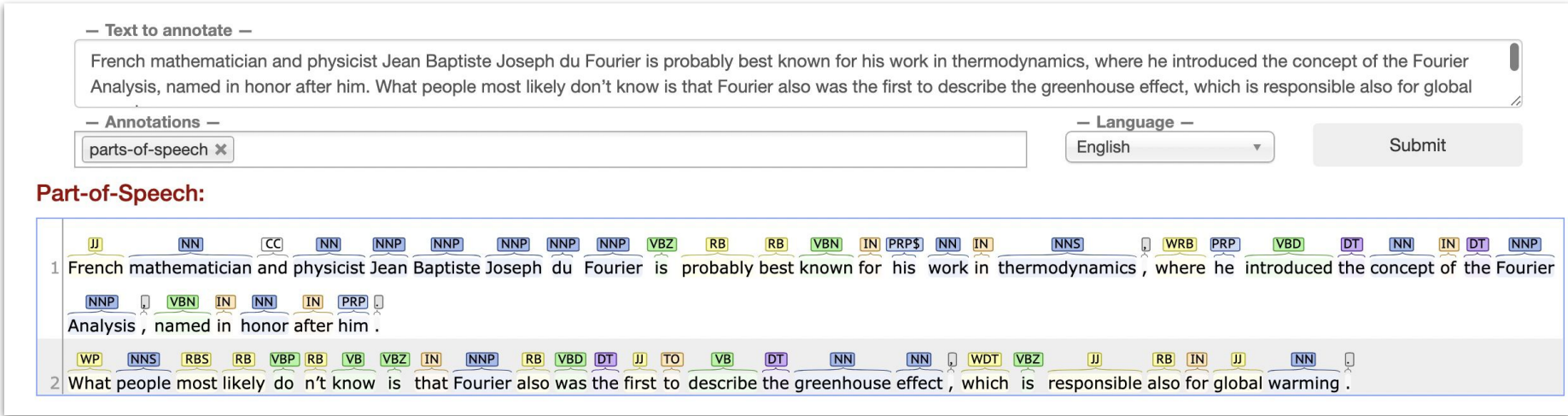
1.

On	March	21	,	1768	,	French	mathematician	and	physicist
Jean	Baptiste	Joseph	du	Fourier	was	born	.		

http://text-processing.com/demo/tokenize/

Part-of-Speech Tagging (POS Tagging)

- Marking up a word in a text as corresponding to a particular **part of speech**, based on both its definition and its context



http://corenlp.run/

Constituency Parsing

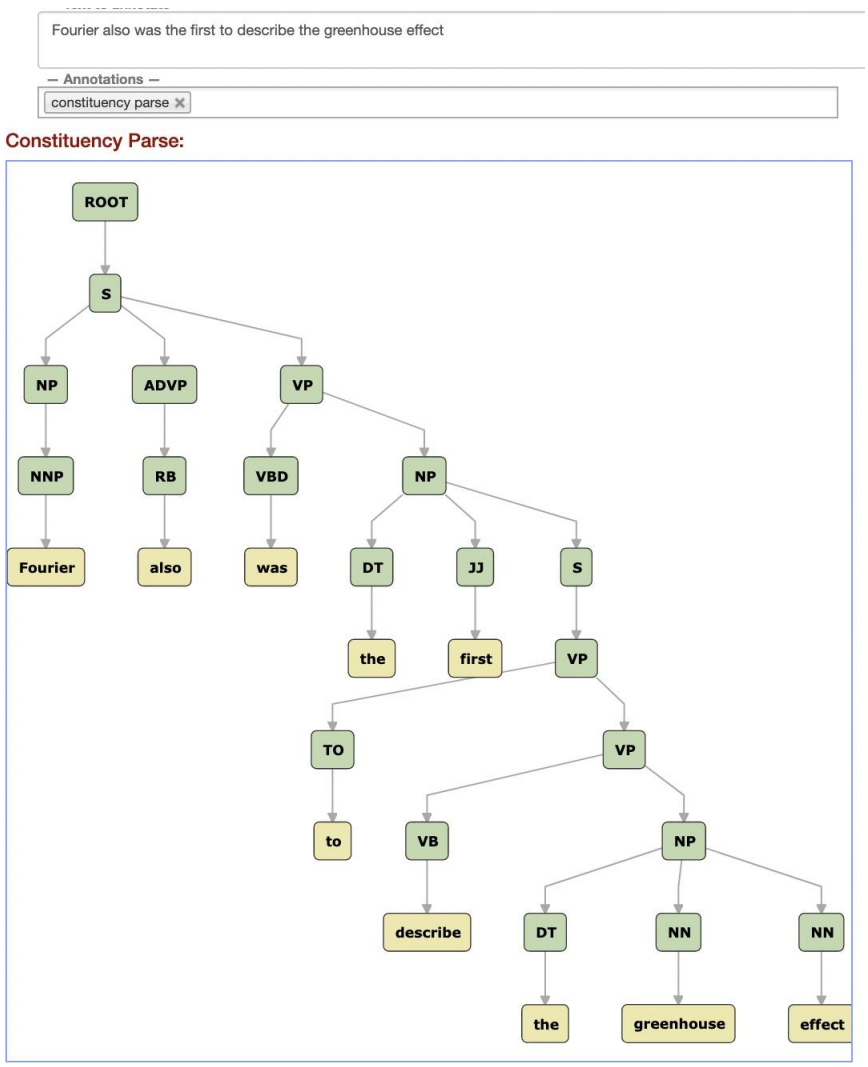
- Analyzing a string of symbols **conforming to the rules of a formal grammar** and
- building a **syntactic tree of a sentence**

Selected Syntactic Categories:

- Noun Phrase (NP)
- Verb Phrase (VP)
- Prepositional Phrase (PP)
- Adverbial Phrases (AdvP)
- Determiner (DT)
- Auxiliary (MD): *have, may*
- Conjunction (CC): *and*
- Noun (NN), Proper Noun (NNP), Verb (VB), Preposition (IN)
- Adjective (JJ), Adverb (RB)
- ...

[Jurafsky, chap. 13]

http://corenlp.run/



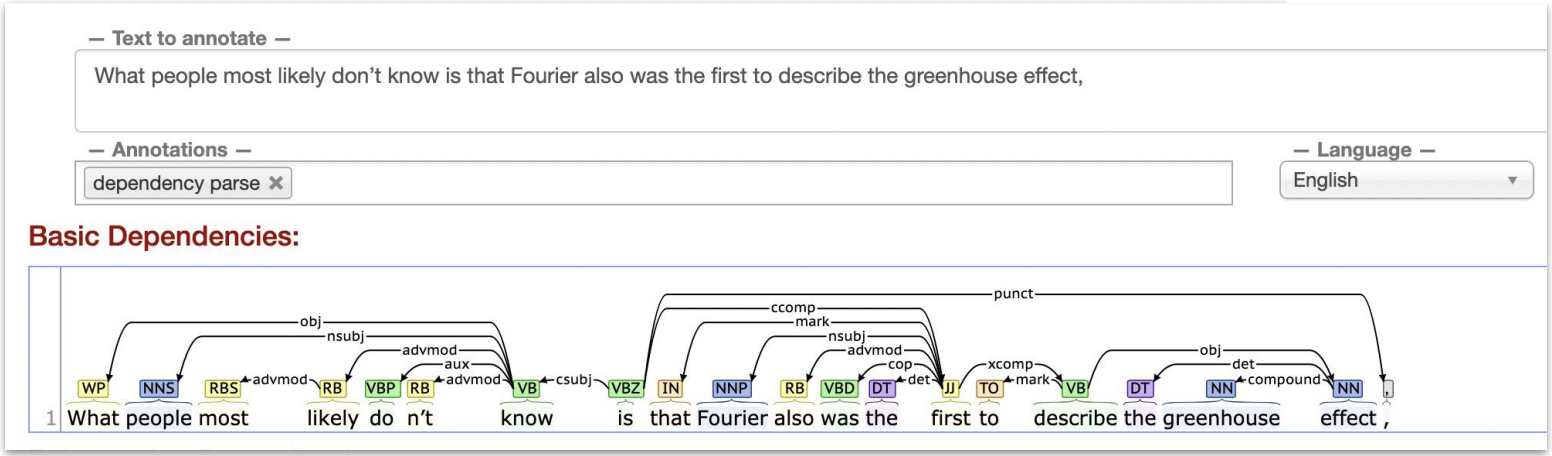
Dependency Parsing

- **Approximation of semantic relations** between arguments
- Relies on **direct binary grammatical relations among words**
- Draws relations from **fixed inventory of grammatical relations**
- Can handle morphologically rich languages

[Jurafsky, chap. 15]

Selected Dependency Relations:

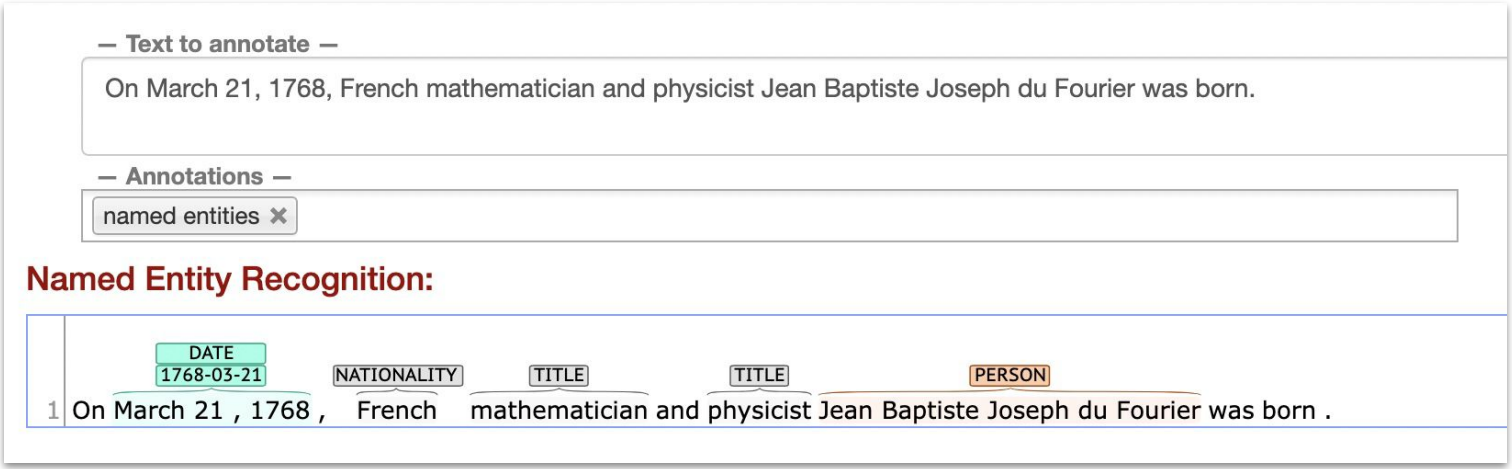
- Clausal Argument Relations
  - **NSUBJ** Nominal subject
  - **DOBJ** Direct object
  - **IOBJ** Indirect object
- Nominal Modifier Relations
  - **NMOD** Nominal modifier
  - **AMOD** Adjectival modifier
  - **DET** Determiner
  - **CASE** Preposition, postpositions, other markers
- Other Relations
  - **CONJ** Conjunct
  - **CC** Coordinating conjunction



http://corenlp.run/

Named Entity Recognition

- Locating and classifying atomic elements into predefined categories such as **names**, **persons**, **organizations**, **locations**, expressions of **time**, **quantities**, **monetary values**, etc. C.J.Rijsbergen, Information Retrieval (1979)



http://corenlp.run/

Named Entity Resolution / Word Sense Disambiguation

- Finding out the exact meaning of an entity, i.e. **linking a text segment to its corresponding entity** in a knowledge base

Jean Baptiste Joseph Baron du Fourier (1768-1830)

On March 21 , 1768 , French mathematician and physicist **Jean Baptiste Joseph du Fourier** was born. He is probably best known for his work in therm ncept of the Fourier Analysis , named in honor after him. There unction of a variable can be expanded to a sum of sines of mu st likely don't know is that Fourier also was the first to descri ponsible also for global warming .

> "Profound study of nature is the most fertile so
> Joseph Fourier, The Analytical Theory of Heat

Joseph Fourier
Jean Baptiste Joseph Fourier (21 March 1768 – 16 May 1830) was a French mathematician and physicist born in Auxerre and best

birth year	1768
death year	1830
is influenced by of	Adolphe Quetelet
birth place	Auxerre
death place	Bourbon Restoration
birth place	Burgundy

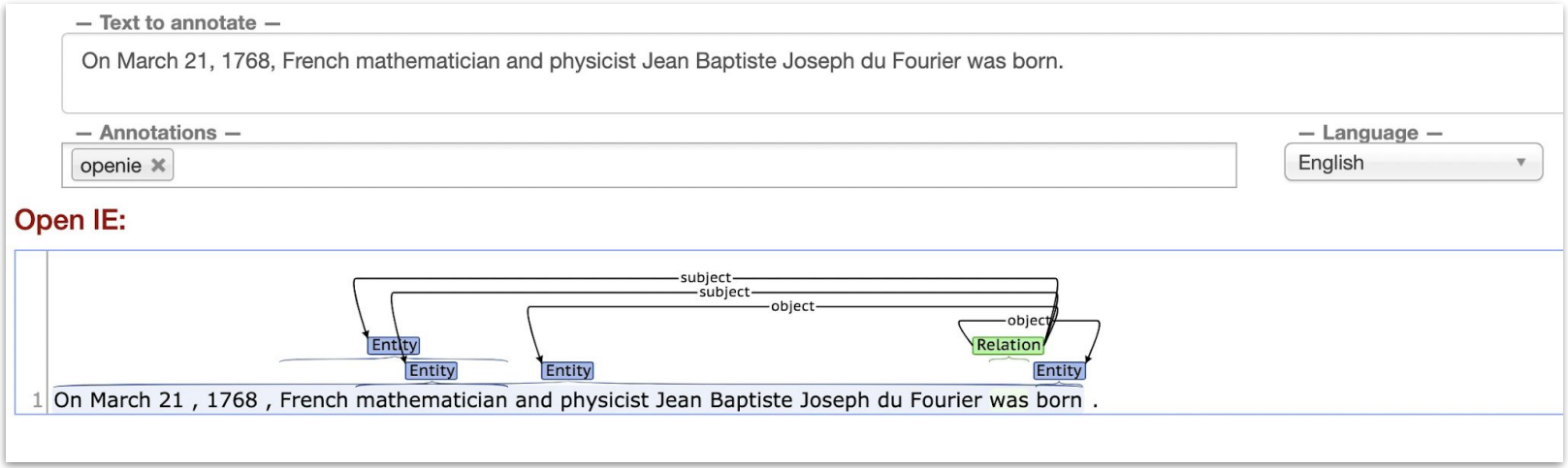
Joseph Fourier – Early Years

Jean Baptiste Joseph Fourier was born on March 21, 176 e, as the son of a tailor. Orphaned already at age nine, Fourier w rre, and through this introduction, he was educated by the Benvenistes of the Convent of St. Mark. While he

http://scihi.org/joseph-fourier-greenhouse-effect/

Semantic Role Labeling

- Extracting **subject-predicate-object triples** from a sentence

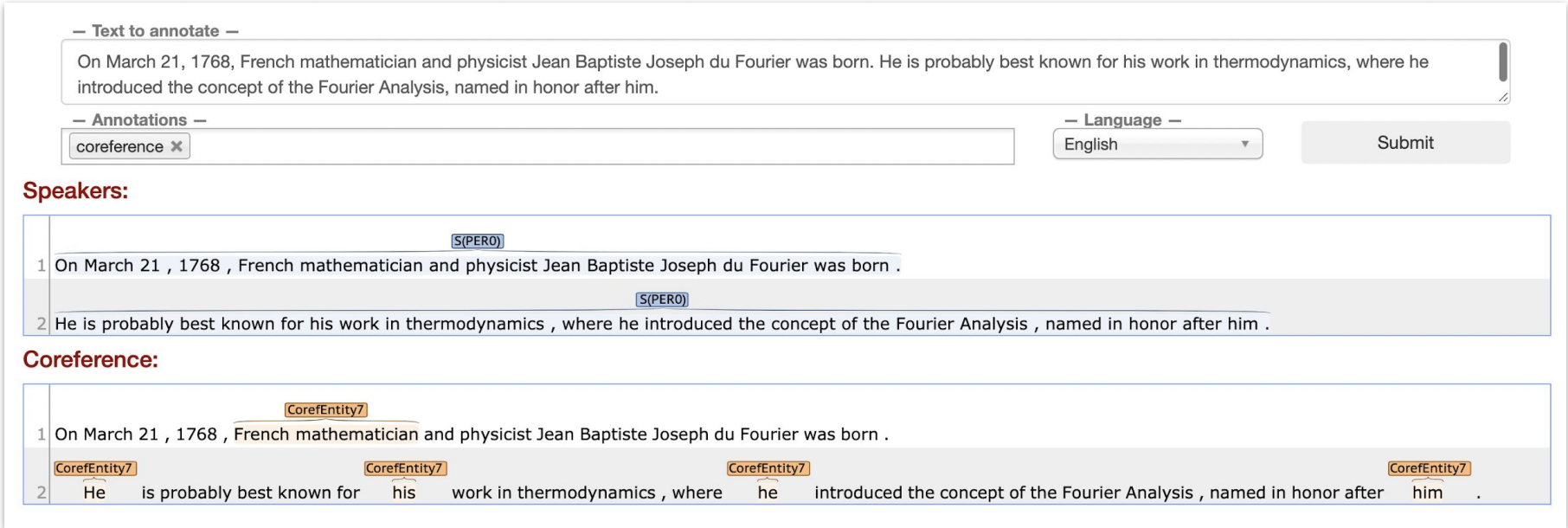


http://corenlp.run/

Co-Reference Resolution

- Also known as **anaphora resolution**
- Determine which phrases in a document or discourse **refer to the same underlying entity**

http://corenlp.run/



Information Service Engineering

Lecture 2: Natural Language Processing (1)

2.0 What is Natural Language Processing?
2.1 NLP and Basic Linguistic Knowledge
2.2 Morphology
2.3 NLP Applications
2.4 NLP Techniques
2.5 NLP Challenges
2.6 Evaluation, Precision and Recall
2.7 Regular Expressions
2.8 Finite State Automata
2.9 Tokenization
2.10 Language Model and N-Grams
2.11 Part-of-Speech Tagging
2.12 Word Embeddings

2. Natural Language Processing - 1

Bibliography

- K.-U. Carstensen et al. (eds.), *Computerlinguistik und Sprachtechnologie, eine Einführung*, 3. Aufl., Spektrum Akademischer Verlag, 2010.
  (Teile von Kap.3, Morphologie)
- D. Jurafsky, J. H. Martin, *Speech and Language Processing*, 3nd ed. (draft), 2019,
  - *13 Constituency Parsing, p. 232ff*
  - *15 Dependency Parsing, p. 273ff*
- Ferrucci, D.; et al. (2010). "*Building Watson: An Overview of the DeepQA Project*". AI Magazine. 31 (3). Retrieved February 19, 2011.

2. Natural Language Processing - 1

Syllabus Questions

- What is natural language processing?
- Why is language a central concept of intelligence?
- What are the main components natural languages are built on?
- What are words, vocabulary, text, language, and discourse?
- What is the difference between phonetics and phonology?
- What is the difference between a phone and a phoneme in phonology?
- What is morphology?
- What are morphemes, free morphemes, and bound morphemes?
- How can new words in a language be created from existing words?
- What is the difference between compounding and derivation?
- What is the difference between derivation and inflection?
- Would you rather use lemmatization or stemming for frequency-based analysis of text? Why?
- What are the most important techniques in NLP?
- What is the difference between Named Entity Recognition and Named Entity Resolution?