

Introduction to Intelligent User Interfaces

Text and Natural Language Processing



Overview - NLP

Definition and Motivation

Applications

- Terms, basic concepts and algorithms
- Milestones in the history of NLP
- Text Analytics

Artificial languages and Natural Languages

- Syntax and semantic
- Vocabulary as set of words
- Text a sequence of words
- Language all "valid" texts





"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications."

"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications."

"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications."

"Natural Language Processing is a theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications."



https://www.youtube.com/watch?v=Sp4q60BsHoY

Text and Natural Language Processing

Challenges in NLP Examples

- Paraphrase an input text
- Translate text from one language into another
- Answer questions about the contents of the text (or corpus)
- Draw inferences and conclusions from the text (or corpus)

Typical algorithms for

- Responding to phrases (e.g. chat bot)
- Discover topics and concepts a text describes
- Summarizing texts (to different length)
- Extracting relevant keywords from texts
- Identify the sentiment of a text or phrase

Break up texts in tokens

. . .

Reduce words to the root / stem

https://algorithmia.com/blog/introduction-natural-language-processing-nlp

lmu r	nünche		Q			
Web	Bilder	Videos	Nachrichten	Karten	Einstellungen 🔻	

Deutschland
 Sichere Suche: Moderat
 Irgendwann

Ludwig Maximilian University - Hotels. Best price guarantee. (WERBUNG)

www.booking.com/Munich/Hotels 🛛 🗭 Werbung melden

Hotels near Ludwig Maximilian University. Save up to 50% on your reservation.

Luxury Hotels

Easy and Secure Online Booking. Read Real Reviews and Book Now!

Book for Tomorrow

Easy, fast and secure booking! New deals listed every day

Book for Tonight Your booking instantly confirmed! Around-the-clock customer service

New deals listed every day! Easy

Budget Hotels

and Secure Booking

LMU München

https://www.uni-muenchen.de

Musikalisches Engagement an der LMU. Freude, schöner Götterfunken Ob Chöre, Ensembles, Orchester - für LMU-Studierende gibt es unzählige Möglichkeiten, sich während der Studienzeit musikalisch zu engagieren. Neben dem Spaß am Musizieren ergeben sich neue Freundschaften oder Konzertreisen rund um den Globus.

Studien- und Lehrangebot - LMU München

tttps://www.uni-muenchen.de/studium/studienangebot/index.html Über die LMU Einrichtungen Studium Studien- und Lehrangebot Studienangebote ...

Biomedizinisches Centrum München - LMU München

https://www.bmc.med.uni-muenchen.de/index.html

Das neue Biomedizinische Centrum München (BMC) ist einer der deutschlandweit größten Forschungsbauten der letzten Jahre - mit Laboren für derzeit etwa 60 Forschergruppen und insgesamt ca. 450 Mitarbeiter. In der Strategie der LMU, Wissenschaft und Klinik eng zu verzahnen, nimmt das BMC einen zentralen Platz ein. Mit seinem Profil und ...

Zentrum Seniorenstudium - LMU München

https://www.seniorenstudium.uni-muenchen.de/index.html

Die Ludwig-Maximilians-Universität bietet akademisch Interessierten ein umfangreiches Studienangebot aus allen Fakultäten. Es kommt den Wünschen nach ...

Text and Natural Language Processing

Institut für Musikwissenschaft - LMU München

https://www.musikwissenschaft.uni-muenchen.de/index.html Von Montag, 28.10.2019, bis Donnerstag, 31.10.2019, können interessierte Schülerinnen und Schüler in das Fach Musikwissenschaft an der Uni hineinschnuppern.

Herzlich Willkommen am Brustzentrum der LMU München!

www.klinikum.uni-muenchen.de/Brustzentrum/de/index.html

Brustzentrum der LMU München Brustkrebs zählt zu den häufigsten bösartigen Erkrankungen der Frau. Nur durch eine sichere Diagnosestellung gefolgt von einer von Anfang an unter Berücksichtigung der Tumorbiologie interdisziplinär geplanten und qualitativ hochwertigen Therapie können heutzutage die besten Überlebenschancen erreicht werden.

Klassische Archäologie - LMU München

https://www.klass-archaeologie.uni-muenchen.de/index.html

11.11.2019 China und Rom. Archäologisches zu einer antiken Fernhandelsbeziehung Vortrag von Prof. Dr. Lorenz E. Baumer ...

Ethikkommission - Medizinische Fakultät - LMU München

https://www.med.uni-muenchen.de/ethik/index.html

Vor der Durchführung eines biomedizinischen Forschungsvorhabens oder einer klinischen Prüfung am Menschen hat sich jeder Forscher, der Mitglied der LMU ist, durch die Ethikkommission bei der Med. Fakultät der LMU beraten zu lassen und ein zustimmendes Votum einzuholen.

Alphabetische Liste aller Personen - Juristische Fakultät - LMU Mün...

https://www.jura.uni-muenchen.de/personen/index.html

Hinweise zur Datenübertragung bei der Google[™] Suche. Links und Funktionen. www.lmu.de; LMU-Portal; Personen ... Sprachumschaltung. English

11

Google

LMU münchen

R Maps 🖾 Bilder Einstellungen Q Alle News Videos : Mehr Tools

Ungefähr 5.880.000 Ergebnisse (0,47 Sekunden)

LMU München

https://www.uni-muenchen.de -

Die LMU ist eine der renommiertesten und traditionsreichsten Universitäten Europas. Sie verbindet hervorragende Forschung mit einem anspruchsvollen .

Ergebnisse von uni-muenchen.de

Studien- und Lehrangebot

Studienfächer und -Studienangebote - Sprachkurse

Studienfächer und

Studienangebote

Bitte beachten Sie: Die auf diesen Seiten eingestellten

Fakultäten

Medizinische Fakultät - Fakultät für Sprach - Juristische Fakultät - ...

Studienfächer und Studiengänge von A bis Z. Studiengänge .

Studium

Studieninteressierte - Studium AZ -Hochschulzugang - Studierende

Einrichtungen

Fakultäten - Medizinische Fakultät - Bibliotheken - ...

Ludwig-Maximilians-Universität München – Wikipedia

https://de.wikipedia.org > wiki > Ludwig-Maximilians-Universität_München -

Die Ludwig-Maximilians-Universität München (kurz Universität München oder LMU) ist eine Universität in der bayerischen Landeshauptstadt München.

Gründung: 1472 in Ingolstadt, seit 1826 in Mü... Präsident: Bernd Huber Land: Deutschland Davon Professoren: 762 (2017)

Schlagzeilen

2	FETE
1	

Oliver Welke an der LMU - "Du darfst nichts Halbgares servieren, auch in der Satire nicht"

Süddeutsche · vor 1 Tag



SZ-Veranstaltung - Was kann Satire? Ein Gespräch mit Oliver Welke

→ Mehr zu LMU münchen

Q

Q

J

Ludwig-Maximilians-Universität München - Home | Facebook https://www.facebook.com > Places > Munich, Germany

★★★★★ Bewertung: 4,4 - 589 Abstimmungsergebnisse Sie begleiten "Herr der Ringe" live im Gasteig und standen schon unter der Leitung von Ennio Morricone in der Olympiahalle auf der Bühne - der Münchner

Department Psychologie - Fakultät für Psychologie und ... https://www.fak11.lmu.de> dep psychologie

... der Psychologie, der Pädagogik/Bildungswissenschaft und des Lehramts bei und ist an übergreifenden Programmen wie der "LMU excellent Graduate School .

Zentrale Lernplattform • LMU München

https://moodle.lmu.de Zentrale Lernplattform der Ludwig-Maximilians-Universität München.

Ludwig-Maximilians-Universität München - Das offizielle ...

https://www.muenchen.de > sehenswuerdigkeiten > orte ★★★★★ Bewertung: 5 - 6 Abstimmungsergebnisse München hat nicht nur die zweitgrößte Universität Deutschlands, sondern auch eine der schönsten: Das beeindruckende Hauptgebäude der ...

Ähnliche Suchanfragen zu LMU münchen

Imu münchen klinikum

Imu münchen adresse

ludwig maximilians universität münchen namhafte absolventen

Imu portal

Isf Imu

Imu münchen online portal

Imu münchen jura

Imu münchen stellenangebote



\sim	A bing com	/soorsh2a-loout	münchen Plac - n Pife		1 Pup q - Insu	Impii Picc-9 6 Pick-Pi
\sim	 bing.com 	/search:q=imu+	munchenœqs=nœic	onn=QbLndsp=-	- recpq=imu	+muccsc=o-occsk=co

Alle	Bilder	Videos	Karten	News	Shopping	I	Meine gespeicherten I
1 370 0	00 Fraebnisse	Datum	*	Sprache -	Region -		

LMU München

https://www.uni-muenchen.de -

Musikalisches Engagement an der LMU. Freude, schöner Götterfunken Ob Chöre, Ensembles, Orchester für LMU-Studierende gibt es unzählige Möglichkeiten, sich während der Studienzeit musikalisch zu engagieren. Neben dem Spaß am Musizieren ergeben sich ...

Studien- und Lehrangebot

Ringvorlesung LMU Studium Generale Seniorenstudium Frauenstudien und Gender ...

Studium

Von Ägyptologie bis Zahnmedizin bietet die LMU München fast 200 Studiengänge mit ...

LSF

Hier sollte eine Beschreibung angezeigt werden, diese Seite lässt dies jedoch nicht zu.

Einrichtungen

Organisation der LMU Hochschulleitung Organigramm der LMU Gremien Beauftragte ...

Stellenangebote

Über die LMU Einrichtungen Studium Forschung Kooperationen Weiterbildung ...

LMU Intern

Die internen Webangebote der LMU sind vielfältig - Portale, Lern-Management ...

Online-Katalog

Funktionen | Katalog der UB | Aufsätze & mehr | Web Opac App und mobile Oberfläche. ...

Immatrikulation

Sobald die Zahlung der Semesterbeiträge auf dem Konto der LMU eingegangen ist (im ...

Preis für LMU-Forscher Mehr Nachhaltigkeit im Unterricht! 30.10.2019. Weitere News ...

werden, diese Seite lässt dies jedoch nicht zu.

Ergebnisse von uni-muenchen.de suchen

Lokale Ergebnisse für Imu münchen Bing Lokale Suche

Geschwister-Scholl-Platz 1 · 80539 München · 089 21800

Routenplaner · Details · O O O O TripAdvisor Bewertungen

In Partnerschaft mit Das Örtliche

Suche

en.uni-muenchen.de - LMU Munich Diese Seite übersetzen https://www.en.uni-muenchen.de/index.html -

LMU Research Fellowship. Getting to Grips with Clouds Postdoc Linda Forster is an Outgoing LMU Research Fellow and is now engaged on a project on cloud formation at the California Institute of Technology. The results should help to improve forecasts of future climate change.

Aktuelles

OPACplus

Hier sollte eine Beschreibung angezeigt

München: Oliver Welke spricht an der LMU über Satire



Abendzeitung · 4 T.



aus Schließfach

Biomedizinisches Centrum München - LMU München

Das neue Biomedizinische Centrum München (BMC) ist einer der deutschlandweit größten Forschungsbauten der letzten Jahre - mit Laboren für derzeit etwa 60 Forschergruppen und insgesamt ca. 450 Mitarbeiter. In der Strategie der LMU, Wissenschaft und Klinik eng zu verzahnen, nimmt das BMC

https://www.qualitative-sozialforschung.soziologie.uni-muenchen.de/index.html -

Lehr- und Forschungsbereich für Qualitative Methoden der empirischen Sozialforschung, Prof. Dr. Hella

Die Ludwig-Maximilians-Universität bietet akademisch Interessierten ein umfangreiches Studienangebot

Orientierung und Zusammenführung der Generationen entgegen und möchte einen Beitrag zur sinnvollen

aus allen Fakultäten. Es kommt den Wünschen nach wissenschaftlicher Information, geistiger

https://www.bmc.med.uni-muenchen.de/index.html -

Qualitative Sozialforschung - LMU München

Zentrum Seniorenstudium - LMU München

https://www.seniorenstudium.uni-muenchen.de/index.html -

einen zentralen Platz ein. Mit seinem Profil und ...

von Unger - Institut für Soziologie - LMU München

Gestaltung des Lebens nach der Zeit aktiver ...

News über Lmu München

bing.com/news



Weitere Nachrichten zu Imu münchen anzeigen

Herzlich Willkommen am Brustzentrum der LMU München! www.klinikum.uni-muenchen.de/Brustzentrum -

Als zertifiziertes universitäres Brustzentrum in einem Tumorzentrum der Spitzenklasse (Comprehensive Cancer Center CCC München) bieten wir unter Leitung von Frau Prof. Harbeck an zwei Standorten in München (Frauenkliniken Maistrasse-Innenstadt und Großhadern) alle Bestandteile einer modernen Brustkrebstherapie unter einem Dach



meiner Frau!"

Text and Natural Language Processing

Albrecht Schmidt & Sven Mayer

Mittelbayerische · 21 Std.

Discussion: Naive implementation of search

Given

- 10000 text documents with an average length of 1000 words
- A search term of up to 4 words
- Aim: the 10 text documents that best match the query
- Approach?
- Where can Machine Learning and AI help?

Scenario:

You have stored all articles of the New York Times from 2010 to 2020

Someone asks for the article about "Merkel visiting the US"

Which articles do you give back?

Query = "media informatics"

- Transfering the query into a coordinate
 - One cell / dimension for each word
 - Count occurances



Example based on: https://livebook.manning.com/book/essential-natural-language-processing/chapter-1/55

Text and Natural Language Processing

Query = "media informatics"

- Transfering documents into coordinates
 - D1: media informatics media media media
 - D2: media informatics media informatics media informatics
 - Count occurances



Example based on: https://livebook.manning.com/book/essential-natural-language-processing/chapter-1/55

Text and Natural Language Processing

Query = "media informatics"

- Transfering documents into coordinates
 - D1: media informatics media media media
 - D2: media informatics media informatics media informatics
 - Count occurances



Example based on: https://livebook.manning.com/book/essential-natural-language-processing/chapter-1/55

Text and Natural Language Processing

17

```
d1 = "media ... informatics media ... media ... media"
d2 = "media informatics ... media ... informatics ... media informatics"
. . .
vector = [0, 0]
for word in d1.split(" "):
       if word=="media":
            vector[0] = vector[0] + 1
       if word=="informatics":
            vector[1] = vector[1] + 1
print (vector)
```

Example based on: https://livebook.manning.com/book/essential-natural-language-processing/chapter-1/55

Overview - NLP

Definition and Motivation

Applications

- Terms, basic concepts and algorithms
- Milestones in the history of NLP
- Text Analytics

- Where do you use NLP on a daily basis?
- What are typical tasks?

- Where do you use NLP on a daily basis?
- What are typical tasks?



Nutzer fragen auch	
Who is the 52 president?	~
Who is the richest president?	~
Is the President of the United States a federal employee?	~
Is a former president still called President?	~
	Eoodback gabon

List of presidents of the United States - Wikipedia States - Wikipedia States - Wikipedia States - Wikipedia.org > wiki > List_of_presidents_o... Diese Seite übersetzen

The **president of** the **United States** is the head **of** state and head **of** government **of** the **United** ... **of** the **United States**; the 45th and current **president** is **Donald Trump**. ... **Of** those **who** have served as the nation's **president**, four died in office **of** ...

List of presidents of the United ... · List of vice presidents · Disambiguation

President of the United States - Wikipedia

https://en.wikipedia.org > wiki > President_of_the_Un... Diese Seite übersetzen Donald Trump is the 45th and current president of the United States. Member of: Cabinet; Domestic Policy Council; ... Term length; Four years, renewable once

Salary: \$400,000 annually Appointer: Electoral College

Text and Natural Language Processing

21

- Where do you use NLP on a daily basis?
- What are typical tasks?



https://www.tomsguide.com/how-to/how-to-transcribe-and-export-written-notes-on-the-galaxy-note-10

Text and Natural Language Processing

- Where do you use NLP on a daily basis?
- What are typical tasks?



Text and Natural Language Processing

Overview - NLP

- Definition and Motivation
- Applications
- Terms, basic concepts and algorithms
- Milestones in the history of NLP
- Text Analytics

Tokenization

- Separating a text into individual words
- Words are called tokens
- Removing punctuation, (multiple) spaces, separators
- Approach:
 - Search along the text and extract tokens separated by space and punctuation
 - Store all tokens in a list
- Any difficulties?

Tokenization

- Separating a text into individual words
- Words are called tokens
- Removing punctuation, (multiple) spaces, separators
- Approach:
 - Search along the text and extract tokens separated by space and punctuation
 - Store all tokens in a list
- Any difficulties?
 Dr. Max von Mayer-Hauser is today in New York and we will meet him-hopefully.

Stop Words Removal

- Remove "small" words in the text, such as articles, pronouns, and prepositions
- Examples
 - English: the, and, a, to, …
 - German: der, die, das, und, es, ...
- Approaches include stop word lists or stop word learning (based on frequency)
- Discussion: Approach based on ML/AI?

Text Normalization

- Aim: match the "same" words
- Syntactic matching
- What is the problem? How to do this?

Text Normalization

- Aim: match the "same" words
- Upper case / lower case letters (e.g. all lower case)
 - Tricky if upper case is required to detect names, grammar
- Normalizing word forms (stemming, lemmatization)
- Acronyms (U.K. \rightarrow UK)
- Umlauts (für \rightarrow fuer or fuer \rightarrow für)
- Dealing with numbers and symbols in text
- Correcting misspelling

Stemming

- Reducing the word to its stem
- Extract the morphological root
- removing affixes and suffixes
- Heuristic process (works most of the time)
- Approaches and algorithms
 - Set of rules (language dependent)
 - E.g. as automaton
- Typical algorithms: Porter Stemmer, Snowball Stemmer
- Example:
 - player, playing, playful, plays, played \rightarrow play
 - newer, newest \rightarrow new ... what happens to news

30

Example Porter Stemmer

- Set of rules for removing/changing suffixes
- Rules are grouped
- Rules for their application of the rules

In a set of rules written beneath each other, only one is obeyed, and this will be the one with the longest matching S1 for the given word. For example, with:

 $SSES \rightarrow SS$

- IES $\rightarrow I$
- $SS \rightarrow SS$
- S →

(here the conditions are all null) CARESSES maps to CARESS since SSES is the longest match for S1. Equally CARESS maps to CARESS (S1 = 'SS') and CARES to CARE (S1 = 'S').

In the rules below, examples of their application, successful or otherwise, are given on the right in lower case. The algorithm now follows:

Step 1a		
$SSES \rightarrow SS$	caresses	\rightarrow caress
IES $\rightarrow I$	ponies	\rightarrow poni
	ties	→ ti
$SS \rightarrow SS$	caress	→ caress
S →	cats	\rightarrow cat
Step 1b		
$(m > 0) EED \rightarrow EE$	feed	\rightarrow feed
	agreed	→ agree
$(^*v^*)$ ED \rightarrow	plastered	→ plaster
	bled	\rightarrow bled
$("v")$ ING \rightarrow	motoring	\rightarrow motor
	sing	\rightarrow sing

If the second or third of the rules in Step 1b is successful, the following is done:

$AT \rightarrow ATE$ $BL \rightarrow BLE$ $IZ \rightarrow IZE$ (*d and not (*L or *S or *Z))	conflat(ed) troubl(ing) siz(ed)	$\begin{array}{l} \rightarrow \text{ conflate} \\ \rightarrow \text{ trouble} \\ \rightarrow \text{ size} \end{array}$
→ single letter	hopp(ing) tann(ed) fall(ing) hiss(ing)	\rightarrow hop \rightarrow tan \rightarrow fall \rightarrow hiss
$(m \ = \ 1 \ and \ ^*o) \ \rightarrow E$	fail(ing) fil(ing)	\rightarrow fizz \rightarrow fail \rightarrow file

Porter, Martin F (1980) "An algorithm for suffix stripping." *Program,* 14(3) Reprinted 2006: <u>https://cl.lingfil.uu.se/~marie/undervisning/textanalys16/porter.pdf</u>

Text and Natural Language Processing

31

Lemmatization

- Return the base (dictionary) form of a word
- Uses linguistic knowledge (vocabulary, grammar, morphological analysis)
- "Lemmatization, unlike Stemming, reduces the inflected words properly ensuring that the root word belongs to the language. In Lemmatization root word is called Lemma. A lemma (plural lemmas or lemmata) is the canonical form, dictionary form, or citation form of a set of words." <u>https://www.datacamp.com/community/tutorials/stemminglemmatization-python</u>
- Examples: was→be, running→run, has→have, swims→swim, caring→care, …

Stemming/Lemmatization based on ML? How would you do this?

Part-Of-Speech Tagging

- Determining the type of a word in the context of a sentence
- Identifying words as nouns, verbs, adjectives, adverbs, …

Named-Entity Disambiguation and Entity linking

- Determining the meaning of a word, if word have more meanings
- Using context and knowledge
- Example

I did not use an apple to make these slides.

VS.

I really like to bake apple crumple.

Named-Entity Recognition

- Gets entities from unstructured texts
- Assigning entities/words to categories
- Examples of named entities: people, places, companies, organizations, industries, products, product categories, time, location, brands, etc.
- Application and domain specific, e.g. abbreviations for trading stocks, medical conditions, addresses
- Library: <u>https://spacy.io/api/annotation#section-named-entities</u>
- Example in Python: <u>https://nlpforhackers.io/named-entity-extraction/</u>
- Software: Stanford Named Entity Recognizer (NER) <u>https://nlp.stanford.edu/software/CRF-NER.shtml</u>

Bag of Words

- Texts are considered a set of words
- Simplified representation
- Ignores grammar (and generally word order)
- Typical calculation:
 - Frequency of occurrence of words
 - Frequency of n-grams (preserves some word order)
- "My cat likes to sleep. I sleep a lot. Do you have a cat?" cat 3x, sleep 2x, a 2x, ...

Corpus, Corpora

- Monolingual corpora
 - data from one single language
- Parallel corpora
 - original texts in one language
 - translations in other languages
- Examples:
 - Gutenberg, archive of free electronic books, <u>https://www.gutenberg.org/</u> or <u>https://www.projekt-gutenberg.org/</u>
 - <u>https://www.collinsdictionary.com/api/collins-english-dictionary,61,HCA.html</u>

Corpus

- "A corpus is a large body of natural language text used for accumulating statistics on natural language text. The plural is corpora. Corpora often include extra information such as a tag for each word indicating its part-of-speech, and perhaps the parse tree for each sentence." NLP Dictionary <u>http://www.cse.unsw.edu.au/~billw/nlpdict.html</u>
- Further Examples:
 - Swiss SMS Corpus, <u>https://sms.linguistik.uzh.ch/</u>
 - The National University of Singapore SMS Corpus <u>https://www.kaggle.com/rtatman/the-national-university-of-singapore-sms-corpus</u>

Live-Coding

Text and Natural Language Processing

Overview - NLP

- Definition and Motivation
- Applications
- Terms, basic concepts and algorithms
- Milestones in the history of NLP
- Text Analytics