

Natural Language Processing

Prof. Dr. Jannik Strötgen
jannik.stroetgen@h-ka.de
Summer 2024

Hochschule Karlsruhe
University of
Applied Sciences



Fakultät für
Informatik und
Wirtschaftsinformatik

Preliminary Agenda



Preliminary Agenda

Date	Topic
19.03.2024	Organisation & motivation
26.03.2024	Introduction to NLP and
02.04.2024	<i>no lecture (Easter)</i>
09.04.2024	Pre-Processing and Part-of-Speech Tagging
16.04.2024	Parsing
23.04.2024	Named Entity Recognition and Disambiguation
30.04.2024	Similarity and Search
07.05.2024	Language Models: Static Word Embeddings

Date	Topic
14.05.2024	Contextual Embeddings
21.05.2024	<i>no lecture (Whitsun break)</i>
28.05.2024	Text Mining and Sentiment Analysis
04.06.2024	Information Extraction & QA
11.06.2024	Applications exploiting NLP
18.06.2024	NLP with LLMs
25.06.2024	My Research Topics
02.07.2024	Recap, exam preparation

Recap

Parsing

The Chomsky Hierarchy

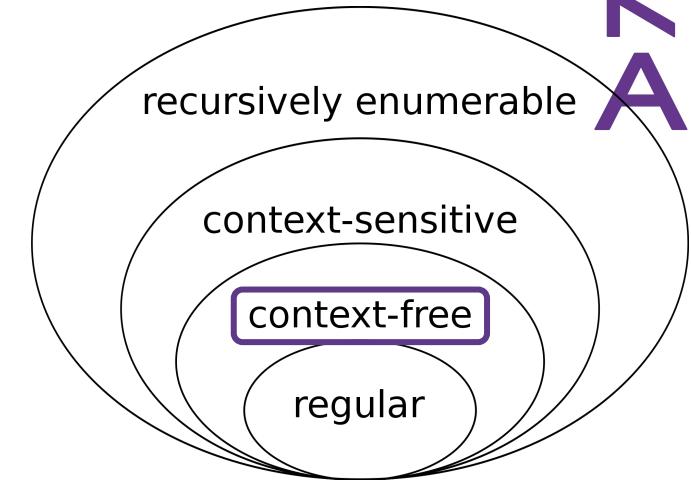
H
K
A

Four types of languages:

- Type-0: Recursively enumerable languages
Rules have the form: $\gamma \rightarrow \alpha$ (no constraints)
- Type-1: Context-sensitive languages
Rules have the form: $\alpha A \beta \rightarrow \alpha \gamma \beta$
- Type-2: Context-free languages
Rules have the form: $A \rightarrow \alpha$
- Type-3: Regular languages
Rules have the form: $A \rightarrow a$ or $A \rightarrow aB$

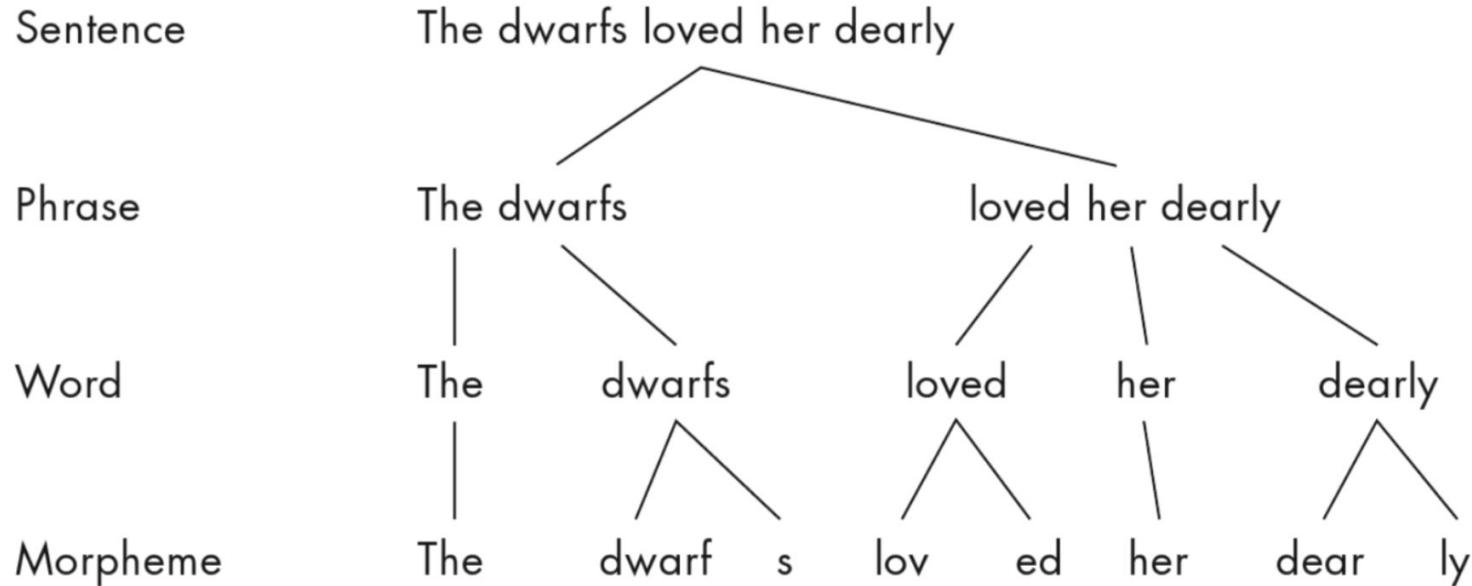
Where:

a is a terminal symbol. A, B are non-terminal symbols. α, β, γ are arbitrary sequences of terminal / non-terminal symbols. α and β maybe empty. γ cannot be empty.



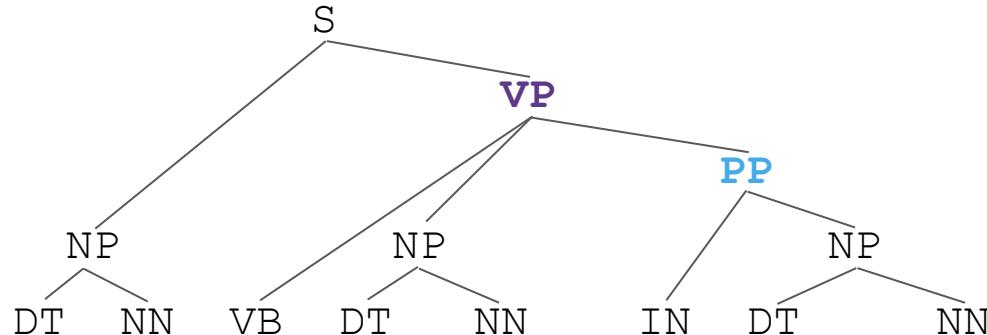
Natural languages can be derived from (mostly) context free grammars (CFGs).

What is Parsing?

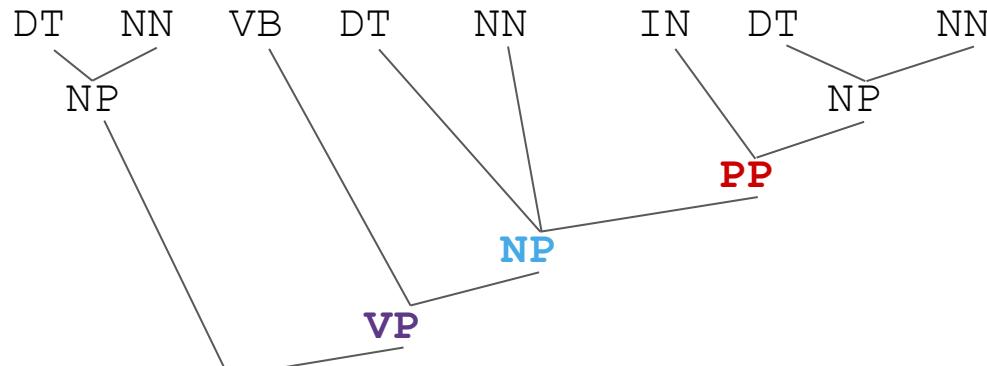


Parsing is the process of determining the **syntactical structure** of a sentence with respect to a given **grammar** (from which the sentence was generated).

Syntactic Ambiguity Parsing Example



The boy saw the woman with the telescope.



Derivation:

- $S \rightarrow NP \ VP$
- $VP \rightarrow VB \ NP \ PP$

Derivation:

- $S \rightarrow NP \ VP$
- $VP \rightarrow VB \ NP$
- $NP \rightarrow DT \ NN \ PP$

Outline

Outline of the Course



1. Foundations and Pre-processing
2. Part-of-speech Tagging
3. Parsing
4. **Named Entity Recognition and Linking**
5. Similarity and Search
6. Language Models: Static Word Embeddings
7. Contextual Language Models
8. Text Mining (Classification, Clustering, and Topic Models)
9. Opinion Mining and Sentiment Analysis
10. Relation Extraction and Question Answering
11. Applications in Document Analysis

Named Entity Recognition and Linking

1. Named Entities and Information Extraction
2. Named Entity Recognition (NER)
 - Sequence Labeling for NER
 - Features for NER
3. Named Entity Linking
 - Named Entity Disambiguation
 - Named Entity Linking
 - Normalizing Temporal Expressions
4. NE-based Applications

Named Entities & Information Extraction

4. Named Entity Extraction & Disambiguation

Some words are more important than others. **Named entities** are **uniquely identified** by their name, for example specific persons, locations, or times that are important in a story.

Ah, distinctly I remember it was in the bleak December;
And each separate dying ember wrought its ghost upon the floor.
Eagerly I wished the morrow;—vainly I had sought to borrow
From my books surcease of sorrow—sorrow for the lost Lenore—
For the rare and radiant maiden whom the angels name Lenore—
Nameless here for evermore.



Elon Musk

Elon Musk offers to buy Twitter for more than \$40bn

Tech entrepreneur makes offer of \$54.20 a share in cash to ‘unlock potential’ of social media site

Elon Musk has launched an audacious bid to buy Twitter for \$43.4bn (£33bn), saying he wants to release its “extraordinary potential” to boost free speech and democracy across the world.

The Tesla chief executive and world’s richest person revealed in a regulatory filing on Thursday that he had launched a hostile takeover of [Twitter](#). He further confirmed the move in a public appearance at the TED conference in Vancouver later that day.

“Having a public platform that is massively trusted and broadly inclusive is extremely important to the future of civilization,” Musk said during an interview with Chris Anderson, Ted conferences curator.

Event Reporting: The 5 Ws of Journalism

H
K
A

In journalism, the report of an event is typically considered complete if (and only if) it answers five questions starting with:

- Who was involved? → Elon Musk
- What happened? → Twitter was bought for \$43 billion
- When did it happen? → Thursday (whatever that means...)
- Where? → Vancouver (maybe?)
- Why?



Events: Formal Perspectives



Linguistic perspective:

- A thing that someone (or something) does
- An action that happens
- ⇒ An (action) **verb**

Information Retrieval perspective:

- Something that happens to an **actor** (or a **group** of actors) at a specific **place** at a specific **time**.
- ⇒ A relation of entities

In practice, **both** views are useful in extracting information from documents...

...but we already know how to identify verbs, so let's focus on entities!

News Annotations: Entities

Elon Musk

Elon Musk offers to buy Twitter for more than \$40bn

Tech entrepreneur makes offer of \$54.20 a share in cash to ‘unlock potential’ of social media site

Elon Musk has launched an audacious bid to buy Twitter for \$43.4bn (£33bn), saying he wants to release its “extraordinary potential” to boost free speech and democracy across the world.

The Tesla chief executive and world’s richest person revealed in a regulatory filing on Thursday that he had launched a hostile takeover of Twitter. He further confirmed the move in a public appearance at the TED conference in Vancouver later that day.

“Having a public platform that is massively trusted and broadly inclusive is extremely important to the future of civilization,” Musk said during an interview with Chris Anderson, Ted conferences curator.

Person

Location

Organization

Number

Time / Date

Definitions: Entities and Named Entities

Entity:

- Philosophically: A thing that exists as itself, either concretely or abstractly
- Pragmatically: **A physical object**

Named Entity:

- A unique entity that is identified by its name
- Examples:
 - Lake Constance (**but not:** water)
 - Heidelberg University (**but not:** university)
 - Elon Musk (**but not:** people)
 - Excalibur (**but not:** sword)

Common Types of Named Entities



The core set (by definition):

- Persons (e.g., Elon Musk)
- Locations (e.g., Karlsruhe)
- Organizations (e.g., EU)

Further useful named entity types:

- Dates (e.g., Friday, 13.05.)
- Times (e.g., 13:37)
- Numeric expressions (e.g., \$43bn)

Domain-specific entity types:

- Chemicals (e.g., $\text{C}_2\text{H}_5\text{OH}$)
- Genes (e.g., TP53)
- Stock symbols (e.g., AAPL)
- Laws (e.g., StGB)
- URLs (e.g., www.h-ka.de)
- etc...

Information Extraction (IE) Systems



Goals:

- Identify and understand **relevant parts** of texts (with limited scope)
- Gather, collate, and **link information** within and between documents in the corpus
- Produce a **structured representation** of relevant information
- **Organize information** so that it is useful to people
- Store information in a **semantically precise format** that is usable by algorithms

IE pipeline:

- Recognition of named entities (e.g., Elon Musk)
- Extraction of relations between entities (e.g., Elon Musk owns Twitter)
- Knowledge base population (e.g., owner(Twitter, Elon Musk))
- For more on relations: see later lecture(s)

Named Entity Recognition

The Named Entity Recognition (NER) Task

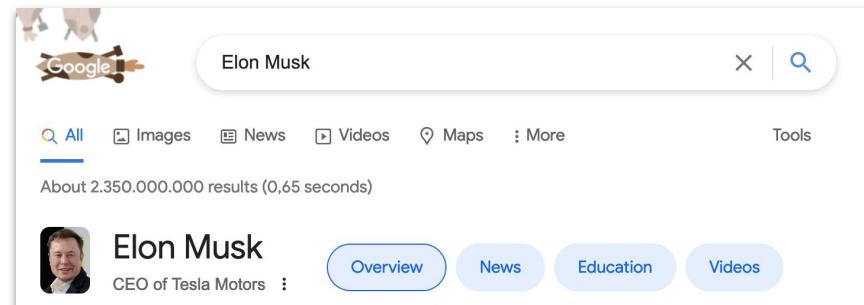


The goal:

- Identify named entities in a document and tag them with a type

Applications:

- Named entities in a text can be indexed, linked to external databases, etc.
- Sentiment can be attributed to companies or products
- Identify relations between named entities
- In question answering: answers are often named entities (Factoids)
- Google knowledge graphs answers
- etc...



Training:

1. Collect a set of representative training documents
2. Manually label each token for its entity class (PER, LOC, ORG, etc.) or other (O)
3. Design appropriate feature extractors for the text and classes
4. Train a sequence classifier to predict the labels from the data

Testing / Classification:

- For a set of new, unlabeled input documents:
 - Run the sequence model in inference mode to label each token
 - Output the recognized entities

A Sample Architecture

HKA

- We often used BiLSTMs with a CRF layer on top for sequence tagging tasks such as NER

Closing the Gap: Joint De-Identification and Concept Extraction in the Clinical Domain

Lukas Lange^{1,2,3} Heike Adel¹ Jannik Strötgen¹
¹ Bosch Center for Artificial Intelligence, Renningen, Germany
² Spoken Language Systems (LSV), Saarland University, Saarbrücken, Germany
³ Saarbrücken Graduate School of Computer Science, Saarbrücken, Germany
`{Lukas.Lange,Heike.Adel,Jannik.Ströetgen}@de.bosch.com`

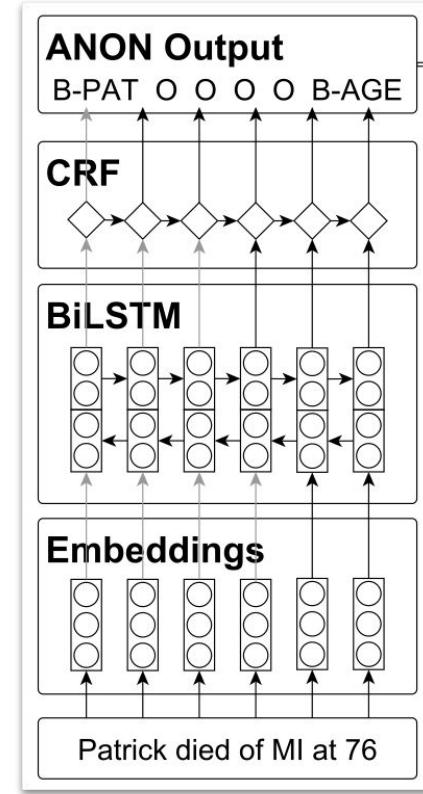
Abstract

Exploiting natural language processing in the clinical domain requires de-identification, i.e., anonymization of personal information in texts. However, current research considers de-identification and downstream tasks, such as concept extraction, only in isolation and does not study the effects of de-identification on other tasks. In this paper, we close this gap by reporting concept extraction performance on automatically anonymized data and investigating joint models for de-identification and concept extraction. In particular, we propose a stacked model with restricted access to privacy-sensitive information and a multi-task model. We set the new state of the art on benchmark datasets in English (96.1% F1 for de-identification and 89.9% F1 for concept extraction) and Spanish (91.4% F1 for concept extraction).

In this paper, we argue that to evaluate the ef-

PATIENT \leftarrow PHI terms \rightarrow **AGE**
Patrick died of myocardial infarction (MI) at 76.
concept \rightarrow PROBLEM

Figure 1: Sentence with annotations of the two tasks.



Encoding Classes for NER Sequence Labeling



Example:

Fred	showed	Mary	Mengqiu	Huang	's	new	painting	
PER	O	PER	PER	PER	O	O	O	(IO encoding)
B-PER	O	B-PER	B-PER	I-PER	O	O	O	(IOB encoding)

In practice:

- Inside-Outside (IO) encoding is less precise
- Inside-Outside-Beginning (IOB) encoding is more precise but requires a larger tagset (2c+1 classes vs. c+1 classes in IO encoding)
- But: the benefit is limited in practice so IO is often used

Token Features for NER Sequence Labeling



Token-level features

- Current token
 - Effectively, this is dictionary learning
 - Example: Barbara is typically part of a name
- Previous/next word (context)

Tag-level features

- Inferred linguistic classification (e.g., POS tags)

Label-level features:

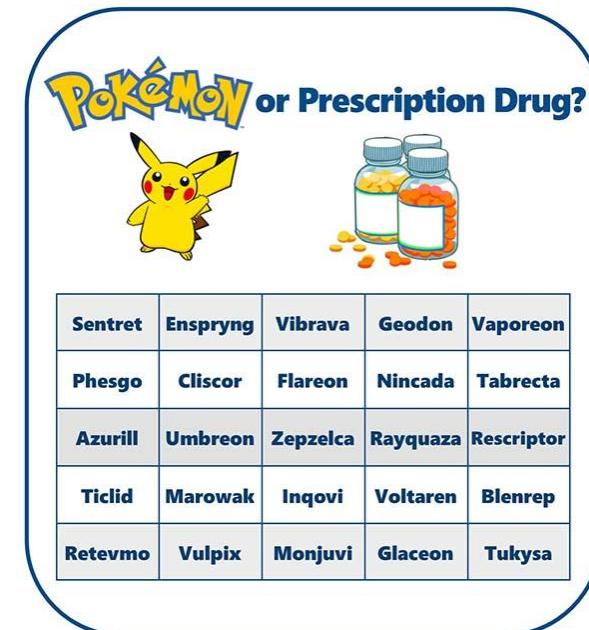
- Previous (and perhaps next) named entity label in the current sequence
- Example: Jon [PER] Snow [PER]

Token Substring Features

H
K
A

Many types of named entities have characteristic name structures in which substrings increase the likelihood for a certain type of entity:

- Substring **oxa** in drug names
 - Oflo**oxac**in, Cotrim**oxa**zole, **Oxa**zepam, etc.
- Substring **field** in location names
 - Clover**field**, Bloom**field**, Wake**field**, Spring**field** etc.
- Special characters like a colon (:) in movie titles
 - Terminator 2: Judgment Day



Token Shape Features



Some named entity names tend to follow patterns that can be mapped to a simplified representation based on attributes such as:

- Token length,
- Capitalization,
- Numerals,
- Greek letters,
- Internal punctuation,
- etc.

Example:

Token	Shape encoding
Varicella-zoster	Xx-xxx
mRNA	xxxx
CPA1	XXXd

Sequence Labeling Task

Once we have labeled training sequences and token features, NER is not very different from some tasks we already know. In many problems in NLP, data is a sequence of characters, words, phrases, lines, or sentences...

Example:

Fred	showed	Mary	Mengqiu	Huang	's	new	painting	
PER	O	PER	PER	PER	O	O	O	(NE tags)
NNP	VBD	NNP	NNP	NNP	POS	JJ	NN	(POS tags)

We can think of NER in similar terms.

Models for Named Entity Extraction



We can use many different kinds of sequence models for NER:

- Markov Models
 - Conditional Markov Model make a single decision at a time, conditioned on evidence from observations and previous decisions
- Conditional Random Fields
 - A whole-sequence conditional model, rather than a chain of local models
- Deep learning models
 - Bidirectional Long Short Term Memory models (LSTMs)
 - Transformers

➡ More on the Deep Learning models in later lectures

Named Entity Disambiguation and Linking

Example of an Entity Relation



Studied at (



,



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Example of Location Ambiguity



Sample locations named Heidelberg

Named Entity Ambiguities

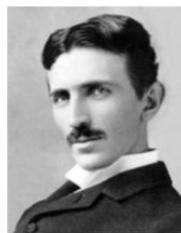
Language is ambiguous. Named entities are no exception...

Example:

The **Tesla** chief executive revealed on Thursday that he had launched a hostile takeover of Twitter.



Company?



Person?

$$T = \frac{Vs}{m^2}$$

Unit?

Handling Ambiguity: Normalization

Normalization: Reducing or rewriting something to a common (normal) form.

Examples:

- Mathematics: normalizing a fraction
 - $1/2 = 2/4 = 128/256 = \dots$
- Mathematics: normalizing a vector to unit length
 - $v' = v / |v|$
- Formal grammars:
 - Chomsky **Normal** Form for CFGs
- Databases:
 - 1NF, 2NF, 3NF, EKNF, etc.
- Person names:
 - Donald Trump, also known as:
Donald J. Trump, Trump, POTUS 45, President Trump...

Named Entity Normalization: Definitions



Donald John Trump Jr. is an American political activist, businessman, conspiracy theorist, author, and former television presenter. He is the eldest child of the 45th president of the United States, **Donald J. Trump**.

NE disambiguation

The task of deciding whether two entity mentions refer to the same entity.



\neq



(NED): NE

linking

The task of linking an entity mention to a unique identifier.



WIKIPEDIA
The Free Encyclopedia

[Donald Trump Jr.](#)

From Wikipedia, the free encyclopedia

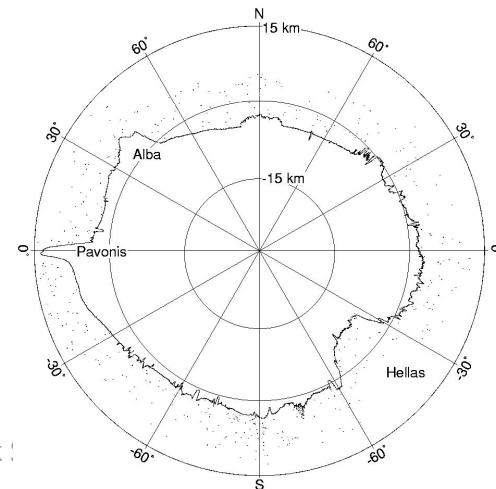
Selecting Normal Forms / Linking Targets

H
K
A

Selecting proper normal forms or linking targets is not trivial.

Some historic examples:

- Y2K Problem:
 - Encoding years with 2-digit numbers during the late 1900s
 - 1998 → '98 → 1998 is reasonably clear. 2001 → '01 → 1901 not so much.
- Normalizing locations via geocoordinates
 - Latitude / longitude works great and is unique...
 - ...until you start sending things to Mars.



How do we avoid these problems in NEL?

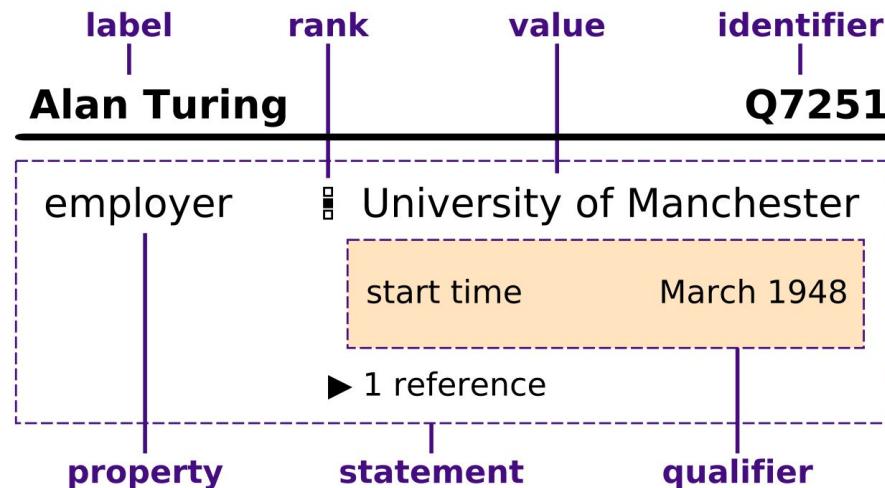
Linking Targets for Named Entities



For **persons**, **organizations**, and **locations**, we typically link to an entry in a **knowledge graph**. The most commonly used resource is Wikidata: an open community-curated knowledge graph and “the database behind Wikipedia”.

Entities have

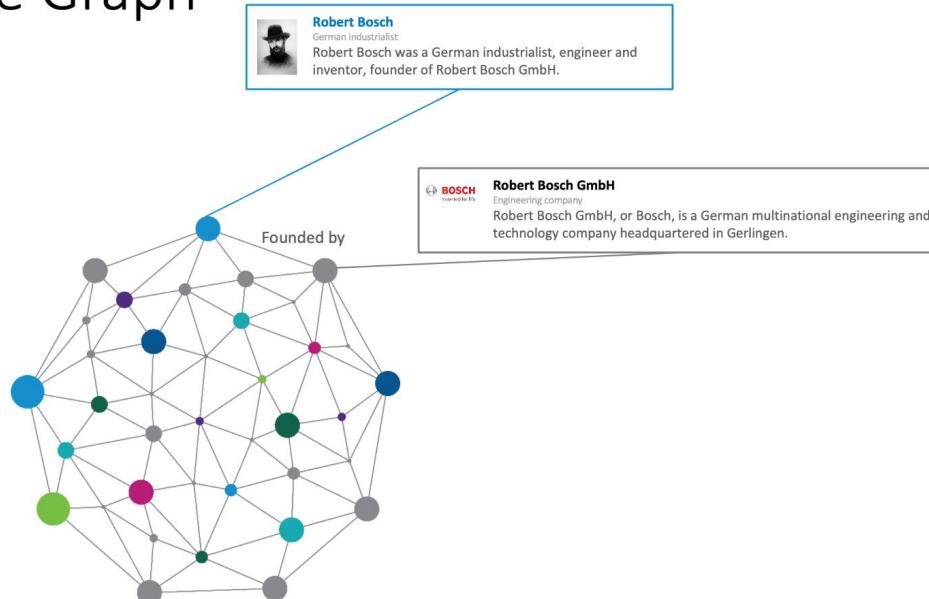
- A unique Q-identifier
- Properties that connect them to other entities
- Validity times for properties
- A ranking mechanism for statements



Reminder: Knowledge Graphs



Knowledge Graph



March 19, 2024

Natural Language Processing @ HKA - Prof. Dr. Jannik Strötgen

28

Named Entity Linking: Algorithmic Approach



How can we match entity mentions in a text to identifiers in a knowledge base?

This can be modeled as a ML prediction task:

Given an entity mention in a text, predict the most likely entity in a knowledge base.

Core steps:

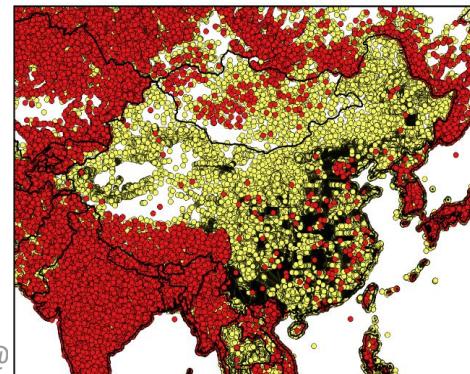
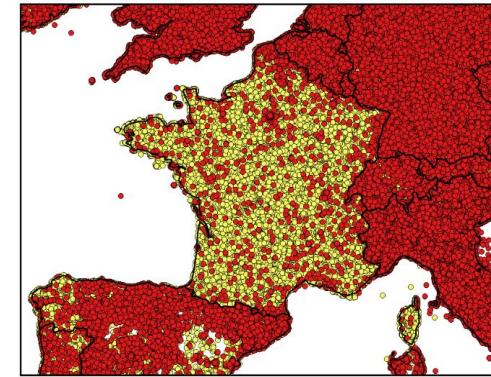
- Candidate **generation** (= named entity recognition)
- Candidate **ranking** using features:
 - Context features: tokens around the entity mentions
 - Knowledge graph context: text features from the knowledge graph
 - Knowledge graph structural features (e.g., graph centrality)
 - Unambiguous entity mentions in the context
 - Heuristic features (e.g., overall popularity of an entity)

Drawbacks of Knowledge Graphs

H
I
K
A

Engineering the structure of a knowledge graph (an **ontology**) is difficult and subjective, but directly affects how useful it is. Examples:

- **Place categories** in Wikidata:
Are they populated places or levels
in an organizational hierarchy?
- **Jesus** in Wikidata:
 - A person? God? A prophet?
 - Real or fictional?



Human settlements
in Wikidata (red) vs.
populated places in
GeoNames (yellow)

Normalization of Temporal Expressions



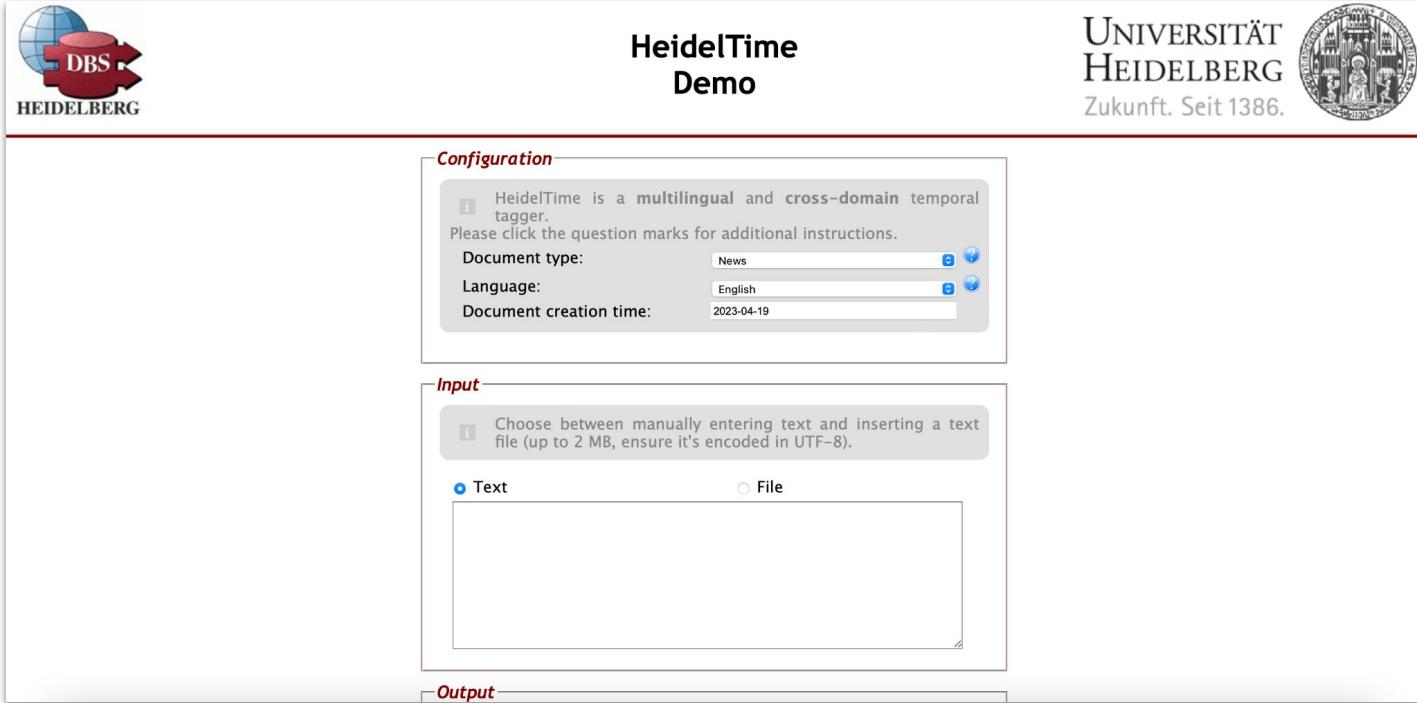
For temporal expressions, we can use a rule-based normalization rather than linking them to a knowledge graph. For example:

April 23, 2024 at 2pm → 2024-04-23 14:00 UTC

But different types of temporal expressions require different rules for normalization:

- **Absolute** temporal expressions
 - April 23, 2024; International Worker's Day 1997;
during the Council of Constance
- **Relative** temporal expressions
 - Yesterday; this afternoon; one year later

Task



The screenshot shows the HeidelTime Demo interface. At the top left is the DBS Heidelberg logo, which consists of a globe icon and the text "DBS HEIDELBERG". The top center features the title "HeidelTime Demo". To the right is the University of Heidelberg logo, which includes the text "UNIVERSITÄT HEIDELBERG Zukunft. Seit 1386." and a circular seal.

Configuration

HeidelTime is a multilingual and cross-domain temporal tagger.
Please click the question marks for additional instructions.

Document type:  

Language:  

Document creation time:

Input

Choose between manually entering text and inserting a text file (up to 2 MB, ensure it's encoded in UTF-8).

Text File

Output

Domain Dependence of Temporal Expressions

For relative temporal expressions, a reference time is necessary, which requires domain knowledge to retrieve.

News-style texts: Use publication metadata

- Elon Musk revealed on Thursday that ...
- News article published on Saturday, 16.04.2022
- Thursday → 2022-04-14

Narrative texts: (often) use preceding Information in paragraph

- Keep track of last mentioned date(s)
- Use as reference for normalization

Hungarian Astronaut ... **1998-04-18**
Tired of being sidelined, Hungarian astronaut Bertalan Farkas is leaving for the United States to start a new career, he **said**
Saturday
... On **May 22, 1995** Farkas was made a brigadier general, and **the following year** he was appointed military attaché at the Hungarian embassy in Washington. However, **cited by** District of Columbia traffic police in **December** for driving under the influence of alcohol, Farkas was ordered home and retired.

TimeBank corpus, APW19980418.0210.tml

Soviet-Afghan War **1979** ← Soviet deployment
The Afghan government, having secured a treaty in **the spring** and **summer of 1979**, They requested Soviet troops to provide security and to assist in the fight against the mujahideen rebels. On **April 14, 1979**, the Afghan government requested that the USSR send 15 to 20 helicopters with their crews to Afghanistan, and on **June 16** ... The operation was fully complete by **the morning of December 28, 1979**. ... According to the Soviet Politburo they were complying with the **1978** Treaty of Friendship, Cooperation and Good Neighborliness ... Soviet ground forces, under the command of Marshal Sergei Sokolov, entered Afghanistan from the north on **December 27**. In **the morning**, the 103rd Guards ...

WikiWars corpus, 18.SovietsInAfghanistan.key.xml

NE-based Application

Temporal Question Answering



TEQUILA: Temporal Question Answering over Knowledge Bases

Ask a question

Answer

Sample Question

TempQuestions

Advanced Options

Dataset

who was the president of us when ezra taft benson was born?

who was governor of minnesota when maathaad maathaadu mallige was released?

which album did neko case release in march 2006?

who is the first husband of julia roberts?

when did the ny knicks last win a championship?

Time-aware Search

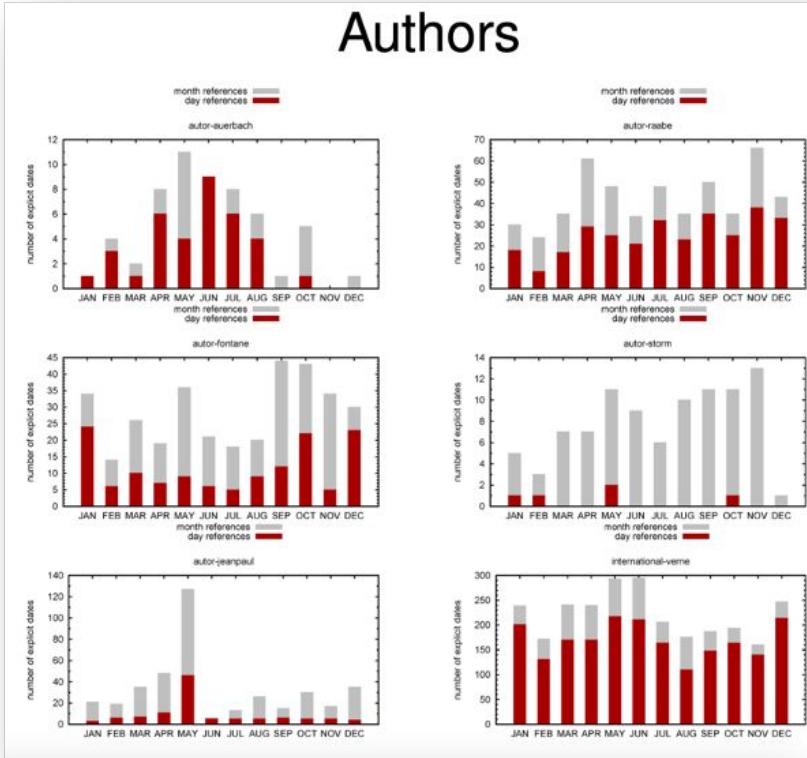
HKA



Today in World Literature

HKA

Authors



« Previous

Quote of the Day (23th of April)

Next »



Edgar Wallace: The Dark Eyes of London (Source)

"I'll put you into the cell after I've charged you," said Larry quietly, and the woman looked up sharply.

"You have charged me with breaking and entering," she said.

"That is not the crime which I shall bring against you," said Larry. "If I get no satisfaction from you now I shall take you back to the pen and charge you with being an accessory to the murder of Gordon Stuart on the night of **the twenty-third of April**."

[Continue reading »](#)

Edgar Wallace

Image Attribution: Unknown (CC BY-SA 3.0 de)

Quotes for date:

#1

Home About Tiwoli Random Quote

DE EN ES

<https://tiwoli.spinfo.uni-koeln.de/view?id=1069&lang=en>

All Dates Lead to Rome



Table 1: Example streets (S) with explanations (E).

S Straße des 13. Januar (January 13), Germany (Saarland)

E On January 13, 1935, the Saar status referendum took place; 90% voted for reunification with Germany.

S 23 Nisan Caddesi (April 23), Turkey (e.g., Ankara)

E National sovereignty and children's day; opening of the Grand National Assembly of Turkey at Ankara in 1920.

S Rue du 8 Mai 1945 (May 8), France (e.g., Paris)

E May 8 is a French holiday; this day in 1945, de Gaulle announced the end of WWII in France.

S Via XX Settembre (September 20), Italy (e.g., Rome)

E This day in 1870, the capture of Rome ended the reign of the Papal States (754–1870).

S Estádio 11 de Novembro (November 11), Angola (Luanda)

E Stadium and surrounding street are named after the date of Angola's independence in 1975.

Exercise Sheet and homework



- In the exercise sheet, you will use NER tools
- But is this really needed in times of ChatGPT?

Exercise Sheet and homework



Wird das funktionieren?

H
K
A

ChatGPT 3.5 ▾

JA

You

Was sind die Zeitausdrücke und ihre Bedeutung in dem folgenden Text?

From: <student>@h-ka.de

To: jannik.stroetgen@h-ka.de

Time: April 15, 2024 - 23:50

Subject: Meeting

Hi Prof,

tut mir leid, dass ich es am Montag nicht geschafft hab. Wie wärs am Freitag? Uhrzeit ist mir egal – vielleicht um 2?

Danke!

<Student>

Exercise Sheet and homework



Wird das funktionieren?

ChatGPT 3.5 ▾

You
Was sind die Zeitausdrücke und ihre Bedeutung in dem folgenden Text?
From: <student>@h-kade.de
To: jannik.strötgen@h-kade.de
Time: April 15, 2024 - 23:50
Subject: Meeting
Hi Prof,
tut mir leid, dass ich es am Montag nicht geschafft hab. Wie wärs am Freitag? Uhrzeit ist mir egal –
vielleicht um 2?
Danke!
<Student>

ChatGPT
Die Zeitausdrücke und ihre Bedeutungen im Text sind:

1. "am Montag" - Bezug auf einen vergangenen Montag, an dem etwas nicht geschafft wurde.
2. "am Freitag" - Vorschlag für einen zukünftigen Freitag als möglichen Termin für ein Treffen.
3. "Uhrzeit ist mir egal" - Bedeutet, dass die genaue Uhrzeit nicht wichtig ist.
4. "vielleicht um 2" - Vorschlag für eine mögliche Uhrzeit, etwa um 2 Uhr, für das Treffen.

Diese Ausdrücke beziehen sich jeweils auf I
über vergangene, gegenwärtige oder zukün
□ ☰ ☱ ☲ ☳ ☴

17. April 2024 Prof. Dr. Jannik Strötgen @ H-KA 66

ChatGPT
Die Zeitausdrücke und ihre Bedeutungen im Text sind:

1. "am Montag" - Bezug auf einen vergangenen Montag, an dem etwas nicht geschafft wurde.
2. "am Freitag" - Vorschlag für einen zukünftigen Freitag als möglichen Termin für ein Treffen.
3. "Uhrzeit ist mir egal" - Bedeutet, dass die genaue Uhrzeit nicht wichtig ist.
4. "vielleicht um 2" - Vorschlag für eine mögliche Uhrzeit, etwa um 2 Uhr, für das Treffen.

Diese Ausdrücke beziehen sich jeweils auf bestimmte Tage und Zeiten und geben Informationen
über vergangene, gegenwärtige oder zukünftige Ereignisse.
□ ☰ ☱ ☲ ☳ ☴

Exercise Sheet and homework



Was wir brauchen: Prompt Engineering

JA You

Du bist ein Temporal Tagger und extrahierst und normalisierst Zeitausdrücke im TimeML Format mit TIMEX3 Tags. Die Tags sollen inline sein. Mach das für den folgenden Text:

From: <student>@h-ka.de
To: jannik.stroetgen@h-ka.de
Time: April 15, 2024 - 23:50
Subject: Meeting

Hi Prof,
tut mir leid, dass ich es am Montag nicht geschafft hab. Wie wärs am Freitag? Uhrzeit ist mir egal – vielleicht um 2?
Danke!
<Student>

Was wir brauchen: Prompt Engineering

ChatGPT

From: <student>@h-ka.de
To: jannik.stroetgen@h-ka.de
Time: <TICKET3 tid="t1" type="DATE" value="2024-04-15">April 15, 2024</TICKET3> - <TICKET3 tid="t2" type="TIME" value="23:50">23:50</TICKET3>
Subject: Meeting

Hi Prof,

tut mir leid, dass ich es am <TICKET3 tid="t3" type="WEEKDAY" value="MONDAY">Montag</TICKET3> nicht geschafft hab. Wie wärs am <TICKET3 tid="t4" type="WEEKDAY" value="FRIDAY">Freitag</TICKET3>? Uhrzeit ist mir egal – vielleicht um <TICKET3 tid="t5" type="TIME" value="14:00">2</TICKET3>?

Danke!

<Student>

Exercise Sheet and homework

Was wir brauchen: Prompt Engineering

< 4 / 4 >

Klare Anweisungen

ChatGPT

- Was ist die E-Mail-Adresse des Professors?
- Was ist die Uhrzeit der nächsten Sitzung?
- Wie sieht der Titel der Sitzung aus?

From: <student>@h-ka.de
To: jannik.stroetgen@h-ka.de
Time: <TIMEX3 tid="t1" type="DATE" value="2024-04-15T23:50:00">April 15, 2024 - 23:50</TIMEX3>
Subject: Meeting
Hi Prof,
tut mir leid, dass ich es am Montag nicht geschafft hab. Wie wärs am <TIMEX3 tid="t2" type="DATE" value="2024-04-19">Freitag</TIMEX3>? Uhrzeit ist mir egal – vielleicht um <TIMEX3 tid="t3" type="TIME" value="T14:00:00">2</TIMEX3>?
Danke!
<Student>

Speaker icon, download icon, refresh icon, reply icon

17. April 2024

Prof. Dr. Jannik Strötgen @ H-KA

70

Further-Watching Material



Introduction to Information Extraction

https://youtu.be/j_ICWJctAfY?list=PLoROMvodv4rOFZnDyrlW3-nI7tMLtmijZ

Watch
me!



Evaluation of Named Entity Extraction

<https://youtu.be/uPU2oZM15I8?list=PLoROMvodv4rOFZnDyrlW3-nI7tMLtmijZ>

Sequence Models for Named Entity Extraction

<https://youtu.be/766a76-HMP4?list=PLoROMvodv4rOFZnDyrlW3-nI7tMLtmijZ>

Outline: End of Part I



1. Foundations and Pre-processing
 2. Part-of-speech Tagging
 3. Parsing
 4. Named Entity Recognition and Linking
 5. Similarity and Search
 6. Language Models: Static Word Embeddings
 7. Contextual Language Models
 8. Text Mining (Classification, Clustering, and Topic Models)
 9. Opinion Mining and Sentiment Analysis
 10. Relation Extraction and Question Answering
 11. Applications in Document Analysis
- “Linguistic” basics
and pre-processing
- Vector representations
of natural language

Thank you for your attention!
Questions?