NLP

- Basic concepts
- Tools

NLP: Basic Concepts

- Interactions between
  - COMPUTERS
    - Unstructured information
    - Programming Languages (c++, Java...)
  - Human (natural) language
    - Ambiguities, missing information...
    - Sentiments, e.g. irony..
    - Fluent dialogs
- Inter-disciplinary subject
  - AI, language, linguistics, statistics, CS,...

What can NLP do for me?

- Cleanup, tokenization
  - Clean the text and find the components (and functions) of each sentence
- Stemming
  - Find the root
- Lemmatization
  - Find the lemma

What can NLP do for me?

- Part-of-speach tagging
- Query suggestions/expansion
- Information extraction
- Entity recognition
- Sentiment analysis
- Disambiguation
- Text similarty

What can NLP do for me?

- Machine translation
- Language recognition
- Summarization
- Speech recognition/processing
- Question answering
- Text to speech
- …

Language recognition
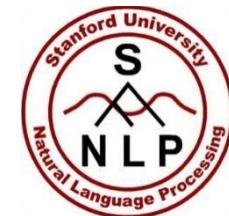
- Apache Tika
  - API for extracting metadata from (almost) any format
  - Useful for search engine indexing, content analysis, translation,…


```
LanguageIdentifier identifier = new LanguageIdentifier("This is my text");
String language = identifier.getLanguage();
```

Standford

- http://nlp.stanford.edu/software

  - Main NLP tools (in Java) for English, Spanish, Chinese, and German(*)

    - tokenization, part-of-speech tagging, named entity recognition, parsing, and coreference, classifier (ML).

    - open source, **licensed under the GNU General Public License**

    (*) Most tools

Stanford CoreNLP

Output format: [Visualise ▾]
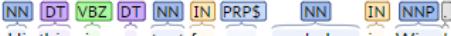
Please enter your text here:

```
Hi this is a test for our workshop in Wien!
```

[Senden] [Clear]

Part-of-Speech:

1 | NN DT VBZ DT NN IN PRP$ NN IN NNP
Hi this is a test for our workshop in Wien!
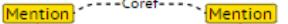
Named Entity Recognition:

1 | Hi this is a test for our workshop in [Pers] Wien!

Coreference:
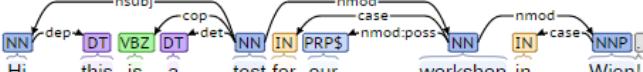
1 | [Mention] Hi this ----Coref---- [Mention] is a test for our workshop in Wien!
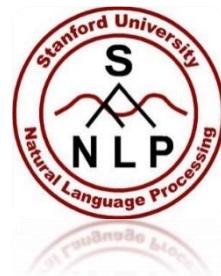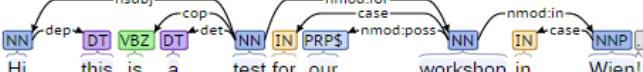
Basic Dependencies:

1 | Hi this is a test for our workshop in Wien!

Enhanced Dependencies:

1 | Hi this is a test for our workshop in Wien!

Visualisation provided using the brat visualisation/annotation software.
Copyright © 2015, Stanford University, All Rights Reserved.

Stanford Named Entity Tagger

Classifier: german.hgc_175m_600.crf.ser.gz ▾

Output Format: highlighted ▾

Preserve Spacing: yes ▾

Please enter your text here:
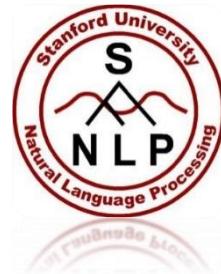
```
Am 9. Juni 2015 lädt die WU zu einer Gesprächsrunde
über die "Affäre Borodajkewycz" ein, die sich heuer
zum 50. Mal jährt. Damit leistet sie einen Beitrag zur
kritischen Auseinandersetzung mit einer der dunkelsten
Episoden der österreichischen Nachkriegsgeschichte und
der eigenen Universitätsgeschichte. Die Zeitzeugen
```

[Senden] [Clear]

Am 9. Juni 2015 lädt die `WU` zu einer Gesprächsrunde über die "Affäre `Borodajkewycz`" ein, die sich heuer zum 50. Mal jährt. Damit leistet sie einen Beitrag zur kritischen Auseinandersetzung mit einer der dunkelsten Episoden der österreichischen Nachkriegsgeschichte und der eigenen Universitätsgeschichte. Die Zeitzeugen Bundespräsident `Heinz Fischer`, Bundesminister a.D. `Ferdinand Lacina` und `Elizabeth T. Spira` werden im Gespräch mit Rektor `Christoph Badelt` und Wirtschaftshistoriker `Peter Berger` über ihr Erleben der Affäre `Borodajkewycz` berichten, das Zeitgeschehen analysieren und dazu Stellung nehmen. Melden Sie sich gleich an!

Potential tags:
`I-MISC`
`B-LOC`
`I-PER`
`I-LOC`
`B-MISC`
`I-ORG`
`B-ORG`

NLTK

- Python library supporting multiple NLP tasks
  - classification, tokenization, stemming, tagging, parsing, semantic reasoning,…
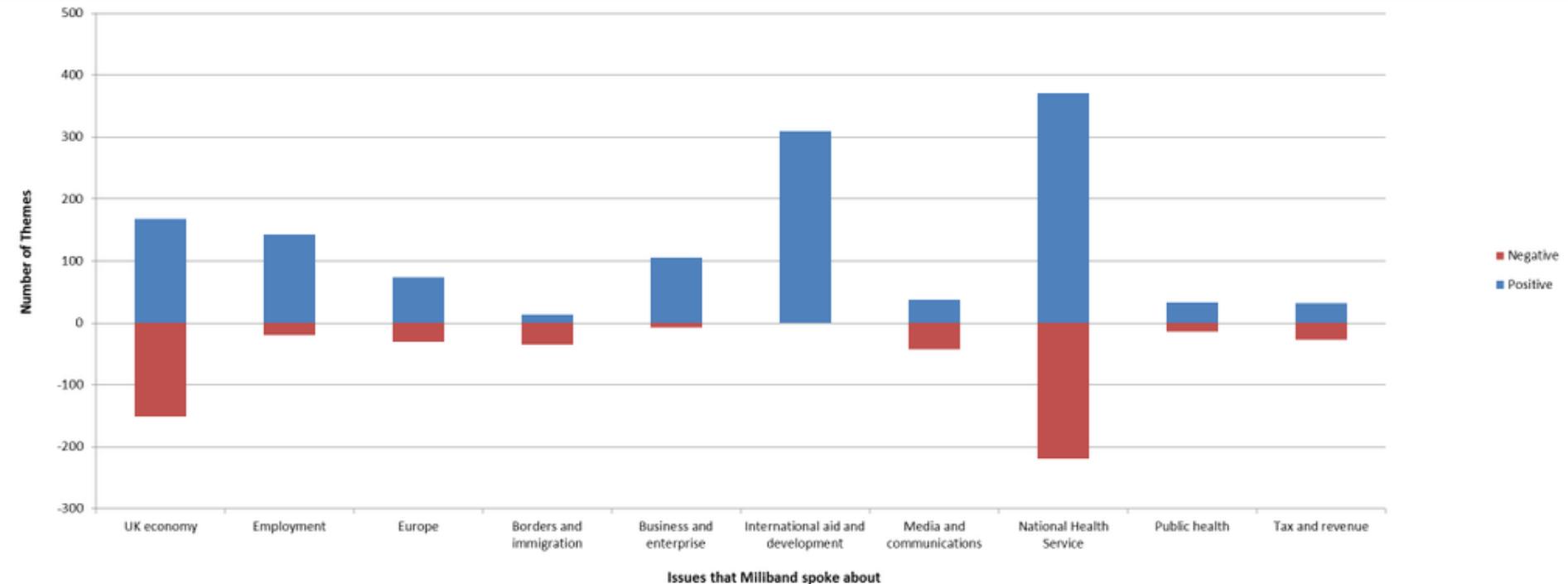
```
>>> import nltk
>>> sentence = """At eight o'clock on Thursday morning
... Arthur didn't feel very good."""
>>> tokens = nltk.word_tokenize(sentence)
>>> tokens
['At', 'eight', "o'clock", 'on', 'Thursday', 'morning',
'Arthur', 'did', "n't", 'feel', 'very', 'good', '.']
>>> tagged = nltk.pos_tag(tokens)
>>> tagged[0:6]
[('At', 'IN'), ('eight', 'CD'), ("o'clock", 'JJ'), ('on', 'IN'),
('Thursday', 'NNP'), ('morning', 'NN')]
```

Others

- R: http://cran.rstudio.org/web/packages/NLP/
  - Based on functionalities from Standford and Apache OpenNLP
- Apache OpenNLP: https://opennlp.apache.org/
  - tokenization, sentence segmentation, part-of-speech tagging, named entity extraction, chunking, parsing, and coreference resolution

Sentimental analysis

- Gate: https://gate.ac.uk/
  - Annotate and analyse corpus
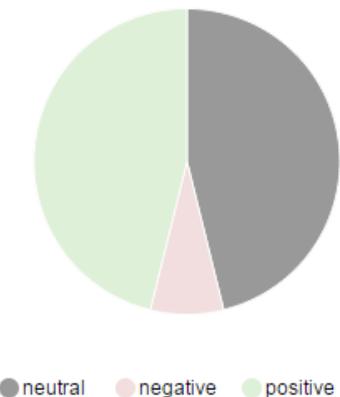
Sentimental analysis

- Plenty of (freemium) tools and APIs
  - https://www.metamind.io/
  - http://www.socialmention.com/
  - http://www.alchemyapi.com/
  - http://www.clarabridge.com/
  - http://www.lymbix.com/

MetaMind

positive

- Now: @CEMS_alliance event with @BCG at beautiful @WU_Vienna. Big thanks to Sabine Baresch (Head of #HR Austria) & @CEMSClubVienna!
- RT @cems_alliance: @Indogermanican @BCG @wu_vienna @CEMSClubVienna Hope it turns out to be a great time!

neutral

- RT @Prison_Health: Effect of Incentives & Meta-Incentives on the Evolution of Cooperation http://t.co/36UXLYZPYW HT @wu_vienna
- WU Campus is made of a lot of different buildings, which one is your favorite ?! #studyabroad #vienna @wu_vienna https://t.co/4u5xaKUwDl

negative

- RT @anna_narloch: You don't want to be embarrased by your #investment #ReinhardMillner #npotag#SocialImpactBond @npokompetenz @wu_vienna

○ neutral ○ negative ○ positive

@wu_vienna   **Classify**

socialmention*

european union

Search

16%
strength

4:1
sentiment

71%
passion

1%
reach

2 minutes avg. per mention

last mention 4 minutes ago

2 unique authors

0 retweets

Sentiment

positive   ▇   4
neutral   ▇▇▇▇▇▇▇▇   27
negative | 1

Top Keywords

european   ▇▇▇▇▇▇   43
union   ▇▇▇▇   25
british   ▇▇   14
trial   ▇▇   12
police   ▇▇   12
rights   ▇   9
human   ▇   9
minutes   ▇   9
vote   ▇   9
greece   ▇   8

Top Users

Amanda Joyal   ▇▇▇▇▇▇   6
Don Quijones   ▇   1

Top Hashtags

Mentions about european union

Sort By: Date ▾   Results: Anytime ▾   Results 1 - 15 of 32 mentions.

European trial over man mistakenly killed by British police

PARIS (AP) — The British government is on trial at the European Court of Human Rights over the death of a 27-year-old electrician, shot by police who thought [R...
kron4.com/2015/06/10/european-trial-over-man-mistakenly-killed-by-british-police/
4 minutes ago - by Amanda Joyal on wordpress

European trial over man mistakenly killed by British police

PARIS (AP) — The British government is on trial at the European Court of Human Rights over the death of a 27-year-old electrician, shot by police who thought [R...
klfy.com/2015/06/10/european-trial-over-man-mistakenly-killed-by-british-police/
6 minutes ago - by Amanda Joyal on wordpress

European trial over man mistakenly killed by British police

PARIS (AP) — The British government is on trial at the European Court of Human Rights over the death of a 27-year-old electrician, shot by police who thought [R...
kwqc.com/2015/06/10/european-trial-over-man-mistakenly-killed-by-british-police/
6 minutes ago - by Amanda Joyal on wordpress

Proposals for a referendum on the UK's membership of European Union have ...

The vote follows the first debate on the EU Referendum Bill and means the legislation move to the next of the process through Parliament. The Conservatives and Labou...
www.gloucestercitizen.co.uk/Proposals-referendum-UK-s-membership-European/story-26670087-detail/story.html
8 minutes ago - on google_news

European trial over man mistakenly killed by British police

PARIS (AP) — The British government is on trial at the European Court of Human Rights over the death of a 27-year-old electrician, shot by police who thought [R...
wavy.com/2015/06/10/european-trial-over-man-mistakenly-killed-by-british-police/
8 minutes ago - by Amanda Joyal on wordpress