

# Mittwoch: Parser und Tagger

- 7 Parser und Tagger
  - TreeTagger
  - MaltParser
  - Stanford Parser

# Parser und Tagger

- 7 Parser und Tagger
  - TreeTagger
  - MaltParser
  - Stanford Parser

# TreeTagger

- Statistischer Part-of-Speech-Tagger
- Frei verfügbar
- `http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/`
- Viele Erweiterungen
- Auf unseren Servern unter `/resources/processors/tagger/tree-tagger`
- Trainierte Modelle für Englisch, Deutsch, Französisch, Niederländisch, Griechisch, Italienisch, ...

# Verzeichnisinhalt

- `bin` – Die eigentlichen Programme
- `cmd` – Shell-Skripte mit eingebautem Preprocessing
- `doc` – Papers
- `lib` – Gespeicherte Modelle

# Englisches Tagset

- NNS Noun, plural
- NN Noun, singular or mass
- DT Determiner
- JJ Adjective
- MD Modal verb
- VBP Present tense verb

Vollständige Liste:

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf>

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - token Gibt das Token aus
  - lemma Gibt das Lemma aus
  - prob Gibt die Wahrscheinlichkeit aus
  - lex f Liest ein zusätzliches Lexikon aus Datei f

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - token Gibt das Token aus
  - lemma Gibt das Lemma aus
  - prob Gibt die Wahrscheinlichkeit aus
  - lex f Liest ein zusätzliches Lexikon aus Datei f

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Benutzung

- TreeTagger stellt Voreinstellungen (in Skripten) für die verschiedenen Sprachen bereit:  
`tree-tagger-english` , `tree-tagger-german` , ...
- Um Feineinstellungen vorzunehmen, muss `tree-tagger` direkt aufgerufen werden: `tree-tagger`
- Einstellungen:  
`tree-tagger [OPTIONS] <parameter file> [<input file> [<output file>]]`
- Options:
  - `token` Gibt das Token aus
  - `lemma` Gibt das Lemma aus
  - `prob` Gibt die Wahrscheinlichkeit aus
  - `lex f` Liest ein zusätzliches Lexikon aus Datei `f`

# Ein- und Ausgabe

`tree-tagger-(english|german|...)`

Der Text muss praktisch nicht vorverarbeitet sein.

`tree-tagger`

Ein Token pro Zeile.

`tree-tagger-german: Ausgabe`

```
reading parameters ...
tagging ...
finished.
Die      ART      d
Sonne   NN       Sonne
scheint VVFIN    scheinen
.        $.      .
```

## *Übung 8*