

Computerlexikographie-Tutorium 27.06.2008

- Themen für heute:
 - Statistische Angaben in Informationssystemen
 - Wortwarte
 - Wortschatz-Portal
 - CCDB
 - Wiederholung I (Kapitel 8)

Wortwarte

- Seit September 2000 werden jeden Tag **Neulexeme** aus Zeitungsartikeln gesammelt.
 - Neologismen: Neulexeme oder Neubedeutungen
- Link: <http://www.sfs.nphil.uni-tuebingen.de/~lothar/nw/>
 - **Frequenzangaben** zu den Wörtern: Wie oft kam das Wort ab der Registrierung als Neulexem in den betrachteten Texten vor.
 - Grundlage für die eventuelle Aufnahme des Lexems ins Wörterbuch
 - **Konkordanz** (Beleg)
- Folien mit Beispielen und mit Erklärung der Prinzipien der Wortauswahl:
 - <http://www.sfs.nphil.uni-tuebingen.de/~lothar/nw/Projekt/NWTFS.pdf>
 - <http://www.sfs.nphil.uni-tuebingen.de/~lothar/nw/Projekt/slidesDO.pdf>
 - <http://www.sfs.nphil.uni-tuebingen.de/~lothar/nw/Projekt/slidesMS.pdf>

Wortschatz-Portal

- 6 Millionen Wörter, 15 Millionen Sätze
- Link: <http://wortschatz.uni-leipzig.de/>
 - Suche (<http://wortschatz.uni-leipzig.de/abfrage/>)
 - **Vorkommenshäufigkeit**
 - morphosyntaktische Informationen, grammatische und lexikalisch-semantische Relationen
 - **Kookkurrenzen**
 - **Kollokationsgraph**
 - Wörter des Tages (<http://wortschatz.uni-leipzig.de/wort-des-tages>)
 - **Korpusbasierte** einsprachige Wörterbücher (23 Sprachen) (<http://corpora.informatik.uni-leipzig.de/>)
 - Wörterbuch Deutsch – Englisch mit **Frequenzangaben** (<http://dict.uni-leipzig.de/>)

CCDB

- Konkurrenzdatenbank vom IDS-Mannheim
- In der CCDB wurden anhand von einem Korpus mit über 2,2 Milliarden laufenden Wörtern mehr als 200000 Wörter aufbereitet.
- Link: <http://corpora.ids-mannheim.de/ccdb/>
 - **Konkordanzen** (Belege)
 - **Nahe Synonyme** und **semantisch ähnliche Wörter** werden graphisch dargestellt.
 - Topographisches Profil mit SOM (self organizing map): semantische Nähe
 - **Konkurrenzpartner**
 - alphabetisch / nach Kohäsionsstärke / nach Wortart geordnet