

Übung 12: spaCy

1. Starten Sie eine Python Interpreter-Session (`$ python` oder `$ ipython`).
2. Importieren Sie die spaCy-Bibliothek.
3. Laden Sie das spaCy-Modell für Englisch.
4. In dieser Aufgabe wollen wir die Performanz von spaCys Tagger und dem Standard-Tagger von NLTK vergleichen. Hierzu benutzen wir den in NLTK erhaltenen Brown-Korpus.
 - a) Benutzen Sie NLTK, um den Brown-Korpus zu laden.
 - b) Extrahieren Sie alle Wörter aus dem Brown-Korpus.
 - c) Wenden Sie NLTKs bevorzugten Tagger auf die Liste der Wörter an. Auf diesen Tagger können Sie mit `nltk.pos_tag` zugreifen.
 - d) Wenden Sie spaCys Tagger auf die Wörter an.
 - e) Extrahieren für beide Ausgaben eine Liste der zugewiesenen Tags (in spaCy benutzen Sie hierfür das Attribut `tag_`).
 - f) Extrahieren Sie die Wörter mit den annotierten Tags im Brown-Korpus, indem Sie die `tagged_words`-Methode des Korpus benutzen.
 - g) Erstellen Sie eine Liste der annotierten Tags.
 - h) Berechnen Sie mit `nltk.accuracy` die Accuracy der Ausgaben, indem Sie diese mit den annotierten Tags vergleichen.