

1 Chomsky-Normalform (2 + 2 Punkte)

Implementiere eine Funktion `cnf(rules)`, die eine Grammatik als Eingabe nimmt und eine äquivalente Grammatik zurückgibt, die nur Regeln enthält, deren rechte Seite aus ein oder zwei Nicht-terminalsymbolen besteht.

Die Eingabegrammatik soll als Liste von Regeln repräsentiert werden, und Regeln als 2-stelliges Tupel, dessen erstes Element die linke Seite (ein String) und dessen zweites Element die rechte Seite (eine Liste von Strings) der Regel darstellen. Der Einfachheit halber kannst Du davon ausgehen, dass das eigentliche Lexikon (Regeln wie `DET → der`) vom Rest der Grammatik getrennt wurde, in den Regeln selbst also keine Terminal-Symbole vorkommen.

Bonus: Erweitere die Implementierung so, dass die Ausgabegrammatik keine Kettenregeln mehr enthält.

2 CYK-Erkenner (3 + 2 Punkte)

Implementiere den CYK-Algorithmus als Erkenner.

Falls Du Aufgabe 1 implementiert hast, soll der Erkenner unbeschränkte Grammatiken als Eingabe nehmen, falls nicht soll er eine Ausnahme werfen, falls die Grammatik nicht in CNF vorliegt. Die Grammatik soll auf jeden Fall im gleichen Format angegeben werden wie in Aufgabe 1. Implementiere eine Methode `recognize_cyk(sentence, rules, lexicon)`, die als Eingabe einen Satz, eine Grammatik (Regeln) und ein Lexikon nimmt und `True` zurückgibt, falls der Eingabesatz in der Sprache der Grammatik enthalten ist (sonst `False`).

Bonus: Erweitere die Implementierung zu einem Parser. Der Parser soll alle Ableitungsbäume für den Eingabesatz als Ausgabe liefern.

3 Earley Erkenner (2 Punkte)

Ändere die Beispielimplementierung des Earley-Erkenners aus der Vorlesung so, dass statt einer Liste effizientere Datenstrukturen für die Regeln der Grammatik und die Chart verwendet wird. Die Datenstruktur soll den direkten Zugriff auf passende Regeln (bei `predict`) bzw. Kanten in der Chart über (bei `complete`) unterstützen (es soll also nicht mehr per for-Schleife nach passenden Regeln bzw. Kanten gesucht werden müssen).

Abgabe bis Donnerstag, 2011-06-16, 08:30 Uhr