

Einführung in die statistische Sprachverarbeitung

Detlef Prescher

4. Juli 2007

Hausaufgabe 10

(i) Wie ist SMT (engl. für Statistical Machine Translation) definiert? Gib eine Formel an und interpretiere sie.

(ii) Wie sieht das Noisy-Channel-Modell für SMT aus? Erläutere seine beiden Komponenten. Warum benutzt man das Noisy-Channel-Modell überhaupt? Gib ein motivierendes Beispiel an (aber nicht das aus der Vorlesung!).

(iii) Was ist Wort-zu-Wort-Alignierung? Gib drei Beispiele an.

(iv) Lies das "SMT Tutorial Workbook" von Kevin Knight. Was sind "spurious words" und wie werden sie in den IBM-Modellen behandelt?

(v) Benutze die Terminologie des EM-Algorithmus, um das Training eines SMT-Modells zu beschreiben. Frage also: was sind die unvollständigen und vollständigen Daten, wie sieht die symbolische Analysekomponente aus, wie der M-Schritt, u.s.w.?

Abgabetermin: Montag, 9. Juli

Besprechung: Mittwoch, 11. Juli