

Einführung in die statistische Sprachverarbeitung

Detlef Prescher

19. Juli 2007

Hausaufgabe 12

(i) Stell Dir einen unfähigen Computerlinguisten vor, der Evaluierung für überflüssig hält. Mit welchen Argumenten würdest Du ihn davon zu überzeugen versuchen, dass man auf Evaluierung unter keinen Umständen verzichten darf?

(ii) Wozu benötigt man TrainSet, DevSet und TestSet bei der System-Entwicklung und Evaluierung? Skizziere, wie man einen unfähigen Forscher erlarven könnte, der in seinen umfangreichen Experimenten auf die Benutzung eines DevSets verzichtete und zu Unrecht "hervorragende Evaluierungsergebnisse" publizierte...

(iii) Wie sind accuracy, precision, recall und f-score definiert?

(iv) Konstruiere ein System, das hervorragende Precision-, aber schlechte Recall-Werte produziert. Konstruiere anschliessend ein System, das hervorragende Recall-, aber schlechte Precision-Werte hat. Kannst Du auch ein Ad-hoc-System angeben, das hervorragende F-scores produziert?

(v) Schau Dir einen Band der ACL-Konferenz an und zähle in wievielen Papieren nicht evaluiert wird.

Abgabetermin: Montag, 23. Juli 2007

Besprechung: nach Bedarf