

# Tutorium

## Aufgaben

Einführung in die CL

28. November 2008

- 1 REs
- 2 EAs
- 3 FSTs
- 4 MED
- 5 N-Gramme
  - Aufgabe
  - Lösung
- 6 HMM
  - Aufgabe
  - Lösung
- 7 Earley
  - Aufgabe
  - Lösung

- 1 REs
- 2 EAs
- 3 FSTs
- 4 MED
- 5 N-Gramme
  - Aufgabe
  - Lösung
- 6 HMM
  - Aufgabe
  - Lösung
- 7 Earley
  - Aufgabe
  - Lösung

# Aufgabe 1

Schreibe einen RE für:

- Am Anfang ein e, dann entweder ein a, ein b oder ein c, dann mindestens ein g, danach zwei beliebige Zeichen exklusive B, am Ende ein e.
- Einen oder keinen Großbuchstaben, a oder ein u, drei Zahlen, ein Punkt

- 1 REs
- 2 EAs
- 3 FSTs
- 4 MED
- 5 N-Gramme
  - Aufgabe
  - Lösung
- 6 HMM
  - Aufgabe
  - Lösung
- 7 Earley
  - Aufgabe
  - Lösung

## Aufgabe 2

Gib jeweils einen EA für die folgenden RE an:

(a)  $/^e[abc]g+[^B]{2}e\$/$

(b)  $/[A-Z]?[au][0-9]{3}\./$

Sind die Automaten deterministisch? Weshalb (nicht)?

# Aufgabe 3

Zeichne einen FST, der ein Nomen+Kasus einliest und das entsprechend deklinierte Nomen ausgibt. Dabei sollen nur Nomina beachtet werden, die auf folgendes Schema zutreffen:

Bsp: Hund+Nom → Hund, Hund+Akk → Hund, Hund+Dat → Hund(e), Hund+Gen → Hundes

# Aufgabe 4

Wende den Minimum Edit Distance-Algorithmus an, um die Editierkosten für die Umwandlung von *kraus* in *hans* zu berechnen. (Kosten: Löschen, Einfügen = 1, Ersetzen = 2, Match = 0).

	#	k	r	a	u	s
#	0	1	2	3	4	5
h	1					
a	2					
n	3					
s	4					

Gib eine minimale Editiersequenz an.

# Aufgabe 5

**Gegeben:** Bigrammtabelle:

	DT	NNS	JJ	RB
DT	-	55	40	7
NNS	5	-	9	9
JJ	10	45	-	6
RB	7	15	32	-

Unigramme:

DT	NNS	JJ	RB
150	200	100	90

$$P(DT | \langle s \rangle) = 0.45$$

Berechne die **Tag**-Übergangswahrscheinlichkeiten:

- 1  $p(DT JJ NNS)$
- 2  $p(DT RB NNS)$

# Aufgabe 5

- ①  $p(DT JJ NNS)$
- $$\begin{aligned} &= p(DT | \langle s \rangle) \cdot p(JJ | DT) \cdot p(NNS | JJ) \\ &= 0.45 \cdot \frac{\#(DT JJ)}{\#(DT)} \cdot \frac{\#(JJ NNS)}{\#(JJ)} \\ &= 0.45 \cdot \frac{40}{150} \cdot \frac{45}{100} = 0.054 \end{aligned}$$
- ②  $p(DT RB NNS)$
- $$\begin{aligned} &= p(DT | \langle s \rangle) \cdot p(RB | DT) \cdot p(NNS | RB) \\ &= 0.45 \cdot \frac{\#(DT RB)}{\#(DT)} \cdot \frac{\#(RB NNS)}{\#(RB)} \\ &= 0.45 \cdot \frac{7}{150} \cdot \frac{15}{90} = 0.0035 \end{aligned}$$

# Aufgabe 6

## Hidden-Markov-Model

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} p(t_1^n | w_1^n) \approx \operatorname{argmax}_{t_1^n} \prod_{i=1}^n p(w_i | t_i) p(t_i | t_{i-1})$$

Gegeben: folgende Emittierungswahrscheinlichkeiten

$$\begin{aligned} p(\text{some} | DT) &= 0.02 & p(\text{mopeds} | NNS) &= 0.002 \\ p(\text{fast} | JJ) &= 0.03 & p(\text{fast} | RB) &= 0.025 \end{aligned}$$

Berechne die Wahrscheinlichkeit  
 $p(DT \ RB \ NNS | \text{some fast mopeds})$ .

# Aufgabe 6

$$\begin{aligned} & p(DT \ RB \ NNS | \text{some fast mopeds}) \\ &= p(\text{some} | DT) \cdot p(DT | \langle s \rangle) \cdot \\ & \quad p(\text{fast} | RB) \cdot (RB | DT) \cdot \\ & \quad p(\text{mopeds} | NNS) \cdot p(NNS | RB) \\ &= 0.02 \cdot 0.45 \cdot 0.025 \cdot \frac{7}{150} \cdot 0.002 \cdot \frac{15}{90} \\ &= 3.5 \cdot 10^{-9} \end{aligned}$$

# Aufgabe 7

Wende den Earley-Parser auf den Satz *die katze sieht den kleinen hund* an.

## Regeln:

S  $\rightarrow$  NP VP

VP  $\rightarrow$  V NP

VP  $\rightarrow$  V NP PP

NP  $\rightarrow$  Det N

NP  $\rightarrow$  Det Adj N

PP  $\rightarrow$  Prep NP

## Lexikon:

V  $\Rightarrow$  sieht | mag

N  $\Rightarrow$  katze | hund | baum

Prep  $\Rightarrow$  auf

Det  $\Rightarrow$  dem | den | die

Adj  $\Rightarrow$  kleinen

# Aufgabe 7

Eingabe: <sub>0</sub> die <sub>1</sub> katze <sub>2</sub> sieht <sub>3</sub> den <sub>4</sub> kleinen <sub>5</sub> hund <sub>6</sub>

1	0 0	S	→	●NP VP	Initialisierung
2	0 0	NP	→	●Det N	Initialisierung:Predict/Expand 1
3	0 0	NP	→	●Det Adj N	Initialisierung:Predict/Expand 1
4	0 0	Det	→	●dem	Initialisierung:Predict/Expand 2,3
5	0 0	Det	→	●den	Initialisierung:Predict/Expand 2,3
6	0 0	Det	→	●die	Initialisierung:Predict/Expand 2,3
7	0 1	Det	→	die●	Shift 6
8	0 1	NP	→	Det ● N	Reduce 2 mit 7
9	0 1	NP	→	Det ● Adj N	Reduce 2 mit 8
10	1 1	N	→	●katze	Predict/Expand 8
11	1 1	N	→	●hund	Predict/Expand 8
12	1 1	N	→	●baum	Predict/Expand 8
13	1 1	Adj	→	●kleinen	Predict/Expand 9

## Lösung

14	1 2	N	→	<i>katze</i> ●	Shift 10
15	0 2	NP	→	<i>Det N</i> ●	Reduce 8 mit 14
16	0 2	S	→	<i>NP ● VP</i>	Reduce 1 mit 15
17	2 2	VP	→	● <i>V NP</i>	Predict/Expand 16
18	2 2	VP	→	● <i>V NP PP</i>	Predict/Expand 16
19	2 2	V	→	● <i>sieht</i>	Predict/Expand 17,18
20	2 2	V	→	● <i>mag</i>	Predict/Expand 17,18
21	2 3	V	→	<i>sieht</i> ●	Shift 19
22	2 3	VP	→	<i>V ● NP</i>	Reduce 17 mit 21
23	2 3	VP	→	<i>V ● NP PP</i>	Reduce 18 mit 21
24	3 3	NP	→	● <i>Det N</i>	Predict/Expand 22,23
25	3 3	NP	→	● <i>Det Adj N</i>	Predict/Expand 22,23
26	3 3	Det	→	● <i>dem</i>	Predict/Expand 24,25
27	3 3	Det	→	● <i>den</i>	Predict/Expand 24,25
28	3 3	Det	→	● <i>die</i>	Predict/Expand 24,25
29	3 4	Det	→	<i>den</i> ●	Shift 27
30	3 4	NP	→	<i>Det ● N</i>	Reduce 24 mit 29

31	3 4	NP	→	<i>Det • Adj N</i>	Reduce 25 mit 29
32	4 4	N	→	<i>•katze</i>	Predict/Expand 30
33	4 4	N	→	<i>•hund</i>	Predict/Expand 30
34	4 4	N	→	<i>•baum</i>	Predict/Expand 30
35	4 4	Adj	→	<i>•kleinen</i>	Predict/Expand 31
36	4 5	Adj	→	<i>kleinen•</i>	Shift 35
37	3 5	NP	→	<i>Det Adj • N</i>	Reduce 25 mit 36
38	5 5	N	→	<i>•katze</i>	Predict/Expand 37
39	5 5	N	→	<i>•hund</i>	Predict/Expand 37
40	5 5	N	→	<i>•baum</i>	Predict/Expand 37
41	5 6	N	→	<i>hund•</i>	Shift 39
42	3 6	NP	→	<i>Det Adj N•</i>	Reduce 37 mit 41
43	2 6	VP	→	<i>V NP•</i>	Reduce 22 mit 42
44	2 6	VP	→	<i>V NP • PP</i>	Reduce 23 mit 42
45	0 6	S	→	<i>NP VP•</i>	Reduce 16 mit 43
46	6 6	PP	→	<i>•Prep NP</i>	Predict/Expand 44
47	6 6	Prep	→	<i>•auf</i>	Predict/Expand 46