

Løsningsforslag - Eksamen V14

Oppgave 1)

a)

1. $(\backslash w+.\backslash w+.)\backslash s$
2. $\backslash d+,\backslash d+$
3. $(\text{http://www.})(\backslash w+)|(\backslash w+.\backslash w+))(\text{.no}((\backslash w^*)^*)?)$

b)

$\backslash.\backslash n$

d)

Q	\w	\.	\s
q0	q1	∅	∅
q1	q1	q2	∅
q2	q3	∅	∅
q3	q3	q4	∅
q4	∅	∅	q5
q5	∅	∅	∅

Oppgave 2)

a)

Åpne ordklasser; ordklasser hvor vi kan tilføye nye ord. \n Lukkede ordklasser; hvor dette (over) oftest ikke er mulig.

Åpne: Adjektiv, adverb, verb, substantiv, interjeksjon \n Lukkede:

Konjunksjon, preposisjon, subjunksjon, pronomen

b)

De åpne ordklassene vil være vanskeligst, ettersom modellen vil slite når den ser helt nye ord.

c)

$$P(t_n|t_{n-1}) = C(t_{n-1}, t_n) / C(t_{n-1}) \quad P(w_i|t_i) = C(t_i, w_i) / C(t_i)$$

$$P(\text{mannen}|N) * P(N|<s>) * P(\text{likte}|V) * P(V|N) * P(\text{en}|D) * P(D|V) * \\ P(\text{kakebit}|N) * P(N|D) * P(</s>|N)$$

- $P(\text{mannen}|N) = 1/3$
- $P(N|<s>) = 0$ (Feil i oppgaven)
- $P(\text{likte}|V) = 1/2$
- $P(V|N) = 1/3$
- $P(\text{en}|D) = 2/3$
- $P(D|V) = 1/2$
- $P(\text{kakebit}|N) = 1/3$
- $P(N|D) = 1/3$
- $P(</s>|N) = 1/3$

d)

Modellen vil tilordne setningen sannsynligheten 0.

Vi unngår slike problemstillinger ved hjelp av smoothing-metoder, hvor vi tildeler manglende forekomster ("missing matter") kunstige sannsynlighetsverdier.

Oppgave 3)

b)

En kontekstfri grammatikk inneholder:

- en endelig mengde ikke-terminaler,
- en endelig mengde terminaler,
- en endelig mengde med produksjonsregler,
- en startvariabel, som må være et element av mengden ikke-terminaler.

Vi ender da opp med:

- $N = \{S, NP, VP, Nom, CoordP, D, Adj, N, V, Coord, Pro\}$
- $\Sigma = \{"den", "en", "gamle", "liten", "mannen", "kaffetår", "kakebit", "smilte", "likte", "og", "han"\}$
- $R = (nedenfor)$
- $S = S$

Følgende regler, mengden R, kan utledes:

```

S -> NP VP
NP -> D Nom | Pro | NP CoordP
VP -> V | V NP
Nom -> Adj Nom | N
CoordP -> Coord NP
D -> "den" | "en"
Adj -> "gamle" | "liten"
N -> "mannen" | "kaffetår" | "kakebit"
V -> "smilte" | "likte"
Coord -> "og"
Pro -> "han"

```

c) Ja, Nom og NP har rekursive høyresider.

d) Følgende regler legges til den eksisterende CFG'en:

```

CoordVP -> Coord V
VP -> VP CoordVP
V -> "blunket" | "tok"

```

e)

(S (NP (Pro han)) (VP (VP (V tok) (NP (D en) (Nom (N kakebit)))) (CoordVP (Coord og) (V smilte))))

f)

i) og ii) forutsatt at vi benytter oss av den utvidede grammatikken.

Oppgave 4)

a) [Side 7](#)

b)

Setningen er sann dersom individet som NP'en denoterer ("Den gamle mannen") er medlem av mengden av smilende (VP) individer.

(Kapittel 3, i Fromkin et al.)

c)

i) Her representerer vi et dokument eller en tekst som mengden av ordene som forekommer, og hvor vi ser helt bort ifra setningsstruktur.

den, mannen, smilte: gamle

ii) Trekkvektor: [den, D, mannen, N, smilte, V]