OPPSUMMERING

Et sammendrag av (nesten) alt vi har snakket om i IN1140..

MORFOLOGI

Hvordan ord er bygd opp, bøyes og dannes, og endrer ordklasse

- I språkteknologi: tokens og typer
- *Innholdsord* vs. *Funksjonsord*
- *Morfem*: Minste meningsbærende enheten av et ord. «Betal» + «-ing»
- *Leksem*: Alle ord som er former av et bestemt ord. Hus – huset – hus – husene
- *Stamme*: grunnelementet av et ord. Fjerner alle bøyningsaffikser. Betalingen
- *Rot*: «kjernen» i et ord. Fjerner alle affikser. Uvennlig
- *Avledning*: Betal - «Betal» + «-ing»
- *Bøyning*: betaler - betalte

EKSAMEN 2019

6 Ordklasser (10 poeng)

Her skal vi jobbe med følgende setning:

Underholdende og sofistikert . Herman tok Norge med storm .

Gitt ordklassene i Tabell 1 under, tildel ordklasser til alle ordene i setningen. Du må velge ett alternativ for hvert ord.

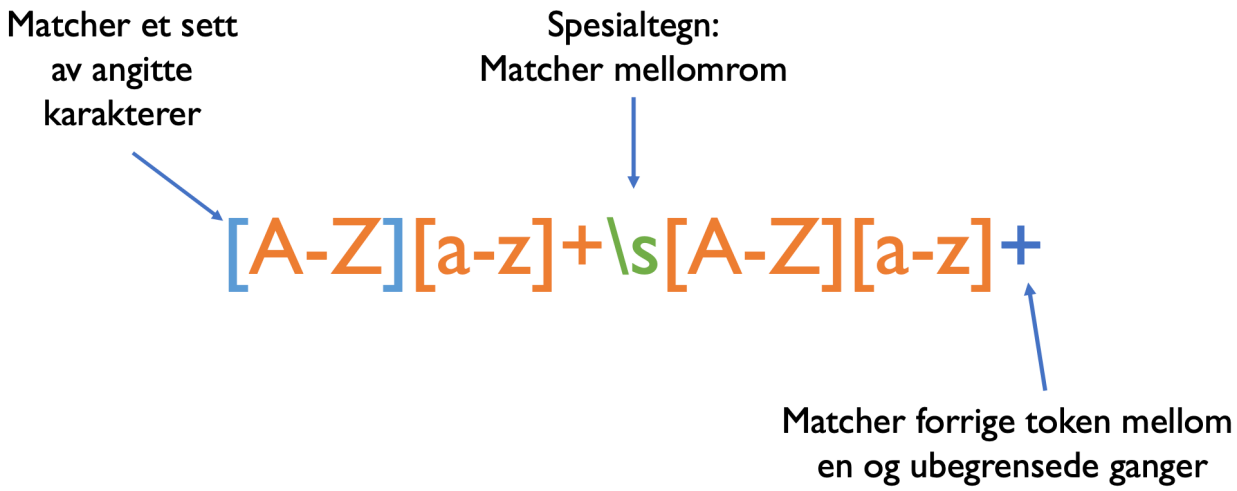
NOUN	Substantiv
ADJ	Adjektiv
VERB	Verb
PROPN	Egennavn
PREP	Preposisjon
CONJ	Konjunksjon
SUBJ	Subjunksjon
ADV	Adverb
DET	Determinativ
PUNCT	Tegnsetting

Finn de som passer sammen

	ADJ	PREP	PROPN	PUNKT	VERB	SUBJ	ADV	CONJ	N
sofistikert	○	○	○	○	○	○	○	○	
tok	○	○	○	○	○	○	○	○	
Underholdende	○	○	○	○	○	○	○	○	
.	○	○	○	○	○	○	○	○	
Herman	○	○	○	○	○	○	○	○	
Norge	○	○	○	○	○	○	○	○	
og	○	○	○	○	○	○	○	○	
storm	○	○	○	○	○	○	○	○	
med	○	○	○	○	○	○	○	○	
.	○	○	○	○	○	○	○	○	

REGULÆRE UTTRYKK

En sekvens av karakterer som danner et søkbart mønster. Kan brukes til å ekstrahere ut det man ønsker i en tekst.



[hallo!]

«hallo! Hvordan går det?»

[«h» «a» «l» «l» «o» «!» «h» «o» «a»]

(hallo!)

«hallo! Hvordan går det?»

[«hallo!»]

KONTEEKSAMEN 2017

1 Regulært uttrykk for URL'er (2 poeng)

Hvilken av følgende URL'er dekkes **ikke** av det regulære uttrykket

```
(www\.)?[a-zA-Z0-9]+\.[a-z]{2}([a-zA-Z0-9\+\.\?]+)?
```

- RegExr.com?2rjl6
- www.ox.ac.uk
- www.uio.no
- www.aftenposten.no/verden

SPRÅKMODELLER

En modell som beregner sannsynligheten for en sekvens av ord

$$P(A|B) = \frac{P(A \cap B)}{\mathrm{P(B)}} \longrightarrow P(\textit{første ord}| < s >) = \frac{P(\textit{første ord}| < s >)}{P(< s >)}$$

For å beregne sannsynligheten for resten av setningen, ganger vi sannsynligheten for hvert bigram sammen:

$$\frac{P(\textit{første ord}|< s >)}{P(<s>)} * \frac{P(\textit{neste ord}|\textit{forrige ord})}{P(\textit{forrige ord})} * \frac{P(< \backslash s >|\textit{siste ord})}{P(\textit{siste ord})}$$

Denne metoden kalles maximum likelihood estimation (MLE)

EKSAMEN 2019

4 Trigram (5 poeng)

Hvor mange trigram forekommer i teksten under?

<s> Bjelleklang bjelleklang over skog og hei </s>
<s> Hør på bjellens muntre klang når Blakken drar i vei </s>

Velg ett alternativ

- 14
- 10
- 16
- 12

5 Estimering av sannsynlighet (5 poeng)

Ta for deg ett av trigrammene fra forrige oppgave og vis hvordan det kan brukes til å beregne sannsynligheten for et ord gitt de to foregående ordene i en trigrammodell.

Skriv ditt svar her...

KONTEKSTFRIE GRAMMATIKKER

Et sett frasestrukturregler som fanger konstituentstatus og rekkefølge

S → NP VP

NP → N PP | N

VP → VP PP | V NP

PP → P NP

Frasale kategorier

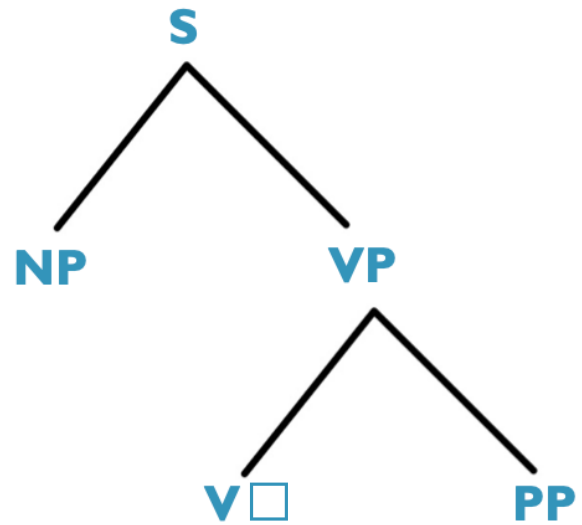
N → pensjonisten | inntrengeren | gevær

P → med

V → jager

Leksikale kategorier

Kan visualiseres med syntaktiske trær



EKSAMEN 2017

Her skal vi jobbe med følgende setning:

`Jeg ser mannen med kikkerten.`

Gitt ordklassene i Tabell 1 under, tildel ordklasser til alle ordene i setningen. Du må velge ett alternativ for hvert tilfelle.

CC	konjunksjon
DET	determinativ
JJ	adjektiv
NN	substantiv
PR	preposisjon
PO	pronomen
RB	adverb
SB	subjunksjon
VB	verb

Setningen i vårt forrige eksempel er strukturelt flertydig. Definer en kontekstfri grammatikk med regler som kan vise ulike analyser av denne setningen. Altså:

`Jeg ser mannen med kikkerten.`

Ta også stilling til om grammatikken din er rekursiv. Begrunn svaret ditt.

NAIVE BAYES

En probabilistisk modell som har som mål å predikere den mest sannsynlige klassen for et dokument

- Naiv: antar at alle trekk er uavhengige

Prior-sannsynlighet

$$P(c) = \frac{Antall\ dokumenter\ med\ klassen}{Totalt\ antall\ dokumenter}$$

Likelihood-sannsynlighet

$$P(w_i|c) = \frac{Antall\ forekomster\ av\ ordet\ i\ klassen}{Antall\ forekomster\ av\ alle\ ord\ i\ klassen}$$

- *Add-one-smoothing*: Legger til 1 i telleren og lengden av vokabularet (for hele treningssettet) i nevneren
- *Out-of-vocabulary-words*: Ignorer ord som ikke forekommer i treningssettet i det hele tatt

EKSAMEN 2019

12 Naive Bayes klassifisering (10 poeng)

I denne oppgaven har vi et lite utvalg ord fra film-anmeldelser som hører til klassene *positiv* eller *negativ*.

1. god, fantastisk, morsom (*POS)*
2. teit, morsom, gøy *(POS)*
3. dårlig, kjedelig, morsom (*NEG)*
4. dårlig, kjedelig *(NEG*)
5. dårlig, teit, kjedelig *(NEG)*

Gitt et nytt test-dokumentet D som inneholder følgende ord: **god**, **teit**, **fantastisk**, bruk Naive Bayes-formelen under til å klassifisere test-dokumentet D.

$$\hat{b} = argmax_{b \in B} \; P(b) \prod_{j=1}^{n} P(v_j|b)$$

Her skal du:

1. Regne ut sannsynlighetene for de forskjellige ordene. Du trenger bare å regne ut for ordene i test-dokumentet. **Ikke** bruk glatting.
2. Regne ut hvilken verdi som er størst. Blir dokumentet klassifisert som *positiv* eller *negativ*?

Skriv ditt svar her...

SEMANTIKK

Kunnskap om betydning- hva betyr ord og setninger?

LEKSIKALE RELASJONER

Beskriver et ords betydning ved å beskrive hvordan det forholder seg til andre ords betydning

Homonymi, polysemi, meronymi, antonymi, hyponymi, synonymi

SEMANTISKE RELASJONER

Aspekt ved setningsbetydning: Hvilke roller de forskjellige deltagerne har

Agent, patient, experiencer, instrument, theme, goal, source, beneficiary

SEMANTIKK I SPRÅKTEKNOLOGI

Hva slags oppgaver kan løses innen semantikk med språkteknologi?

- Ord: Words Sense Disambiguation (WSD)
  - Naive Bayes
- Fraser: Named Entity Recognition (NER)
  - BIO-klassifisering
- Setninger: Semantic Role Labeling (SRL)
  - Syntaktisk analyse- klassifiserer konstituenter i et syntaktisk tre

EKSAMEN 2017

Gitt følgende setninger:
"**Markus** ryddet lekene. Plutselig, kastet Nora **brannbilen** på ham og skadet ham. **Edna** så hva som skjedde. Hun ropte på mor som måtte komme og rense såret med et **antibakterielt middel**."
Angi de semantiske rollene for ordene i tabellen under. Du må velge ett alternativ for hvert tilfelle.

Finn de som passer sammen

	INSTRUMENT	BENEFICIARY	PATIENT	EXPERIENCER	THEME	AGENT
Edna	○	○	○	○	○	○
Markus	○	○	○	○	○	○
lekene	○	○	○	○	○	○
et antibakterielt middel	○	○	○	○	○	○
brannbilen	○	○	○	○	○	○

EKSAMEN 2019

13 Leksikale relasjoner

Hvilken semantisk relasjon holder mellom følgende ord-par? NB! Her får du poeng for riktig svar, men ikke negative poeng for feil svar.

	synonymi	meronymi	hyponymi	antonymi
fot -- tå	○	○	○	○
pen -- vakker	○	○	○	○
singel -- gift	○	○	○	○
sommerfugl -- insekt	○	○	○	○
demokrati -- folkestyre	○	○	○	○
inne -- ute	○	○	○	○