

IN1140: Introduksjon til språkteknologi

Eksamen H2017

Universitetet i Oslo



1. Regulære uttrykk (2 poeng)



Hvilken av de følgende epost-adressene kan ikke gjenkjennes med det følgende regulære uttrykket:

```
[aA-zZ0-9\._%+~]+@[aA-zZ0-9\.-]+\. [aA-zZ]{1,2}
```

Velg ett alternativ

1. Ola.Norman@epost.com
2. olanorman@meg.epost.no
3. OlaNorman@meg.co
4. Ola.Norman@epost.no

2. Regulære uttrykk for verb (6 poeng)

Skriv ett regulært uttrykk som gjenkjenner formene imperativ, infinitiv, presens, og preteritum av følgende norske verb: spise, kjøpe, tenke, løpe.

3. Ordklasser (5 poeng)



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

Table: 1

4. Grammatikk og strukturell flertydighet (10 poeng)



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.

5. Oppgave 5 Leksikalsk semantikk/relasjoner (7 poeng)



1. Forklar forskjellen mellom homonymi og polysemi. Gi eksempler på begge.
2. Forklar og gi eksempler på relasjonen hyponnymi.
3. Forklar og gi eksempler på relasjonen meronymi.

6. Komposisjonalitet (5 poeng)



Forklar med noen få setninger hva vi mener med at setningssemantikk er komposisjonell (compositional)?

7. Semantiske roller (5 poeng)



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	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Markus	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
lekene	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
et antibakterielt middel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
brannbilen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. Maximum likelihood estimering (MLE) (5 poeng)



For å estimere sannsynligheter i n-gram språkmodeller har vi benyttet oss av såkalt maximum likelihood estimering (MLE). Vis formlene for hvordan vi fra et korpus estimerer sannsynligheten for et ord w_i gitt det foregående ordet w_{i-1} i en sekvens, altså $P(w_i|w_{i-1})$.

9. Glatting (smoothing) (5 poeng)



Å bruke MLE (maximum likelihood estimation) alene kan by på problemer. For språkmodeller utvider vi gjerne MLE med såkalt glatting (smoothing). Forklar kort hvorfor MLE alene byr på problemer. Forklar også generelt hva glatting er, og hvorfor vi bruker det.

10. Metodologiske paradigmer (5 poeng)



I språkteknologi kan vi grovt skille mellom regelbaserte metoder og empiriske metoder basert på maskinlæring. Forklar med noen få setninger hva som karakteriserer forskjellene.

11. Lingvistiske nivåer (5 poeng)



Hvilke nivåer av analyse opererer vi med i lingvistikk? Forklar også helt kort (stikkord er nok) hva som er fokuset for analyse på hvert nivå.

12. Named Entity Recognition (5 poeng)



Named Entity Recognition (NER) er automatisk identifisering og kategorisering av egennavn. Forklar kort hvorfor oppslag i en navneliste alene er ikke den beste måten å gjøre NER på. Du kan bruke følgende setning for inspirasjon:

Flyet lander på John F. Kennedy i kveld.

13. Språkmodeller (5 poeng)



Hva er en språkmodell? Forklar med noen få setninger. Nevn også noen eksempler på anvendelser av slike modeller.