

**i Info****Eksamen i IN1140****Introduksjon til språkteknologi**

27. november kl. 14:30 (4 timer)

**Tillatte hjelpemidler**

Ingen

**1 Regulære uttrykk for ord (2 poeng)**Hvilket av alternativene kan **ikke** gjenkjennes med det følgende regulære uttrykket: $\sim \dots j \dots t \dots \{1,3\} \$$ **Velg ett alternativ**

- Adjunction
- Adjusting
- Majestic
- Adjective

---

Maks poeng: 2**2 Regulære uttrykk for enkle aritmetiske uttrykk (6 poeng)**

Skriv ett regulært uttrykk som gjenkjenner følgende klasse av strenger :

- En streng starter med "=" etterfulgt av et aritmetisk uttrykk som bruker heltal, addisjon, subtraksjon, og/eller multiplikasjon. Anta her at heltallet inneholder maksimum fire siffer, og at strenger inneholder maksimum tre aritmetiske operasjoner.

For eksempel strengene " $= 2*3+8-1$ " og " $= 2+1000+120+9$ " skal gjenkjennes av dit regulære uttrykk.**Skriv ditt svar her**

---

Maks poeng: 6

### 3 Tokens og types (4 poeng)

Gitt følgende setning (dikt fra André Bjerke):

**I landet Miramarmora var Farao på ferie hos farmora og mormora . En morgen klatret mormora til Farao i furua , og s**

1. Forklar forskjellen mellom begrepene tokens og types.
2. Gi antall tokens og types i setningen.

Skriv ditt svar her...

---

Maks poeng: 4

### 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

---

Maks poeng: 5

### 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...

---

Maks poeng: 5

**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	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
tok	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Underholdende	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Herman	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Norge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
og	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
storm	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
med	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 10

**7 Ordklassetagging (6 poeng)**

Gitt at vi har trent en ordklassetagger. Det finnes to stragier for å evaluere en slik modell: Ekstrinsisk og intrinsisk evaluering.

Forklar forskjellen mellom disse to strategiene.

**Skriv ditt svar her...**

Maks poeng: 6

**8 Grammatikk (10 poeng)**

Anta følgende kontekstfrie grammatikk:

S -> NP VP  
 NP -> NP PP  
 VP -> V2 NP  
 VP -> V1  
 PP -> P NP  
 NP -> *Rudolf, Nissen, nissen, gavene, sleden, grøten*  
 V1 -> *sov, danset*  
 V2 -> *trakk, spiste*  
 P -> *med*

Grammatikken over er ikke en komplett grammatikk for norsk. For hver av setningene under, ta stilling til om de gis en analyse av grammatikken.

	JA	NEI
Rudolf spiste med nissen	<input type="radio"/>	<input type="radio"/>
Nissen spiste gavene	<input type="radio"/>	<input type="radio"/>
Nissen danset med Rudolf	<input type="radio"/>	<input type="radio"/>
Nissen med gavene sov	<input type="radio"/>	<input type="radio"/>
Rudolf danset	<input type="radio"/>	<input type="radio"/>

Maks poeng: 10

**9 Rekursjon (5 poeng)**

Er grammatikken i forrige oppgave rekursiv? Hvis ja, forklar hvorfor. Hvis nei, gi et eksempel på en rekursiv regel.

Skriv ditt svar her...

Maks poeng: 5

**10 Konstituentter (5 poeng)**

Vi ønsker å avgjøre hvorvidt *sleden* er en konstituent i setningen *Rudolf trakk sleden*. Beskriv minst to konstituenttester og vis hvordan de kan brukes for å hjelpe oss med å avgjøre om *sleden* er en konstituent.

Skriv ditt svar her...

Maks poeng: 5

**11 Utvidelse av grammatikk (5 poeng)**

Du skal nå utvide grammatikken slik at den kan håndtere setninger som *Rudolf tror at nissen sov* og *Rudolf tror at nissen spiste grøten*. Hvilke regler må du legge til den opprinnelige grammatikken for å kunne analysere disse setningene?

Skriv ditt svar her...

Maks poeng: 5

## 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 ( <i>POS</i> )<br>2. teit, morsom, gøy ( <i>POS</i> )<br>3. dårlig, kjedelig, morsom ( <i>NEG</i> )<br>4. dårlig, kjedelig ( <i>NEG</i> )<br>5. dårlig, teit, kjedelig ( <i>NEG</i> ) |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

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

$$\hat{b} = \underset{b \in B}{\operatorname{argmax}} P(b) \prod_{j=1}^n P(v_j|b)$$

Her skal du:

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

Skriv ditt svar her...

Maks poeng: 10

## 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	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
pen -- vakker	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
singel -- gift	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sommerfugl -- insekt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
demokrati -- folkestyre	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
inne -- ute	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 3

## 14 Semantiske roller (8 poeng)

Gi en kort beskrivelse av hva som kjennetegner de følgende semantiske rollene og illustrer svaret ditt med minst ett eksempel for hver rolle.

- EXPERIENCER
- AGENT
- PATIENT
- INSTRUMENT

Skriv ditt svar her...

Maks poeng: 8

## 15 Named Entity Recognition (11 poeng)

Den vanligste måten å løse Named Entity Recognition på er ved ord-for-ord klassifisering, så kalt BIO klassifisering.

Anta følgende setning:

**Thorbjørn Egner ble født 12.12.1912 i Norge på Kampen i Oslo**

Angi riktig BIO klassifisering med riktig kategori for hvert ord i setningen. Her opererer vi med følgende kategori: PER (person), ORG (organisasjon), LOC (location), DT (dato), GPE (geopolitical entity).

Finn de som passer sammen

	O	B_PER	I_PER	B_ORG	I_ORG	B_LOC	I_LOC	B_DT	I_DT
Thorbjørn	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Egner	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ble	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
født	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12.12.1912	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Norge	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
på	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Kampen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Oslo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 11

## 16 Dialogsystemer (2 poeng)

Vi har sett på to hovedeksempler av dialogsystemer: oppgaveorienterte dialogagenter og chatbots.

For å kunne utvikle slike systemer må vi først forstå hvilke egenskaper av menneskelige samtaler kan være utfordrende å automatisere.

Gi eksempler på to slike egenskaper, og forklar hvorfor de er utfordrende.

Skriv ditt svar her...

Maks poeng: 2

