

i Informasjon**Kotteeksamen i IN1140, høst 2017****Tid:**

23.januar klokken 14:30 (4 timer)

Hjelpemidler:

Ingen.

Før du begynner:

Vi anbefaler å lese gjennom hele oppgaveteksten før du begynner. Hvis du føler du mangler informasjon for å løse en oppgave, gjør dine egne antakelser og redegjør for dem. Ca 30–90 minutter etter påbegynt eksamen vil det bli anledning til å spørre en faglærer om eventuelle oppklaringer til oppgaveteksten. Du kan svare på enten norsk eller engelsk.

I dette oppgavesettet har du mulighet til å svare med digital håndtegning (oppgave 6). Du bruker skisseark du får utdelt. Det er anledning til å bruke flere ark per oppgave. Se instruksjon for utfylling av skisseark på pult.

Det er IKKE anledning til å bruke digital håndtegning på andre oppgaver enn oppgave 6. Det blir IKKE gitt ekstratid for å fylle ut informasjonsboksene på skisseark (engangskoder, kand.nr. o.l.).

1 Regulært uttrykk for URL'er (2 poeng)

Hvilken av følgende URL'er dekkles **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

Maks poeng: 2

2 Regulært uttrykk for forkortelser (5 poeng)

Skriv et regulært uttrykk som kjenner igjen forkortelser av typen eksemplifisert under. Uttrykket skal være generelt nok til å også gjenkjenne andre forkortelser som ikke forekommer i listen under men som er oppbygd på samme måte.

Eksempler:

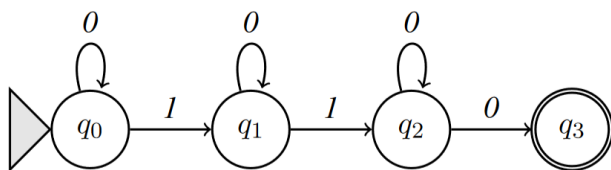
Dr.
Jr.
F.eks.
f.eks.
o.l.
U.S.A.
D.C.

Skriv ditt svar her

Maks poeng: 5

3 Endelige tilstandsmaskiner (10 poeng)

Ta utgangspunkt i følgende endelige tilstandsmaskinen:



1. Angi transisjonstabellen for maskinen.
2. Er maskinen deterministisk? Hvorfor/hvorfor ikke? Begrunn svaret ditt.

Skriv ditt svar her

Maks poeng: 10

4 Ordklasser (10 poeng)

Ta for deg setningen under og tildel ordklasser til alle ordene i setningen. Du skal benytte deg av ordklassene i vedlagte oversikt.

Politiet ber om tips etter et våpentyveri i Oslo sentrum

Finn de som passer sammen

	preposisjon	determinativ	adverb	substantiv	verb	konjunksjon
om	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
etter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tips	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Oslo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
sentrum	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
et	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
våpentyveri	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Politiet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ber	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Maks poeng: 10

5 Strukturell flertydighet (15 poeng)

Anta følgende kontekstfrie grammatikk for et fragment av engelsk:

$$S \rightarrow NP VPNP \rightarrow D NomNom \rightarrow Nom PP \mid NVP \rightarrow V \mid V NP \mid VP PPPP \rightarrow P NPN \rightarrow man \mid bartender \mid cocktail \mid umbrella$$

Hvor mange analyser tildeler grammatikken over til følgende setninger:

	3	1	2	0
the bartender danced with the chef	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the man danced	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the man with the umbrella danced	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the bartender mixed the cocktail with the umbrella	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
the woman at the bar sang	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 15

6 Syntaktiske trær (10 poeng)

I denne oppgaven skal du svare med digital håndtegning. Bruk eget skisseark (utdelt). Se instruksjon for utfylling av skisseark på pult.

Angi det/de syntaktisk(e) treet/træene grammatikken tildeler følgende setning. Du skal her tegne det/de syntaktisk(e) treet/træene for hånd.

the bartender mixed the cocktail with the umbrella.

Bruk følgende grammatikk:

$$S \rightarrow NP VPNP \rightarrow D NomNom \rightarrow Nom PP \mid NVP \rightarrow V \mid V NP \mid VP PPPP \rightarrow P NPN \rightarrow man \mid bartender \mid cocktail \mid umbrella$$

Maks poeng: 10

7 Utvidelse av grammatikken (6 poeng)

Vi ønsker å tillate adjektiver som modifierer substantiver i grammatikken vår. For eksempel:

1. *the angry bartender by the bar mixed the green cocktail*

2. *the short angry man danced*

Hvordan kan vi utvide grammatikken under slik at den kan analysere slike setninger? Merk at i hvert fall en av de nye regelenes skal være rekursiv, dvs at den skal tillate et ubegrenset antall adjektiver (*the short angry strange unpleasant weird little man*).

$$S \rightarrow NP VPNP \rightarrow D NomNom \rightarrow Nom PP \mid NVP \rightarrow V \mid V NP \mid VP PPPP \rightarrow P NPN \rightarrow man \mid bartender \mid cocktail \mid umbrella$$

Skriv ditt svar her

Maks poeng: 6

8 Leksikale relasjoner (5 poeng)

Hvilken semantisk relasjon holder mellom følgende ord-par:

Finn de som passer sammen

	Synonymi	Homonymi	Antonymi	Hyponymi	Meronymi
keep - keep	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
red - green	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
mother - woman	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
screen - phone	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
lawyer - attorney	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 5

9 Semantiske roller (7 poeng)

1. Forklar kort hva en semantisk rolle er og hvordan semantiske roller brukes til å beskrive setningsbetydning.
2. Gi en kort beskrivelse av hva som kjennetegner følgende roller:
 - a. AGENT
 - b. THEME
 - c. PATIENT
 - d. INSTRUMENT

Skriv ditt svar her

Maks poeng: 7

10 Navngjenkjenning (5 poeng)

1. Hva er navngjenkjenning (Named Entity Recognition -- NER) ?
2. Gi noen eksempler på hvordan flertydighet kan være en utfordring for en NER oppgave.

Skriv ditt svar her

Maks poeng: 5

11 Informasjonsgjenfinning (5 poeng)

Ikke alle ord i en tekstkorpus er like informative. Forklar hvordan vi kan vekte ord slik at informative ord teller mer og ikke-informative ord teller mindre.

Skriv ditt svar her

Maks poeng: 5

12 Evaluering (5 poeng)

Generelt finnes det to strategier for å evaluere en trent modell (feks. PoS-tagger, språkmodell, ...): Ekstrinsisk evaluering og intrinsisk evaluering.
Forklar forskjellen mellom disse to strategier.

Skriv ditt svar her

Maks poeng: 5

i Introduksjon

Her skal vi tenke oss at vi ønsker å utvikle et system som automatisk skal skille mellom to forskjellige betydninger av substantivet "mål". Vi antar at dette substantivet har to hovedbetydninger: en konkret betydning (**CONCR**) og en abstrakt betydning (**ABSTR**). Vedlagt finner du et lite korpus som er annotert med betydningsinformasjon.

Vi ønsker å beregne den mest sannsynlige betydningen **b** (fra en mengde mulige betydninger **B**) for en forekomst av ordet "mål", gitt en n-dimensjonal trekkvektor **v** som beskriver forekomsten. Vi bestemmer oss for å løse denne oppgaven ved Naive Bayes-klassifisering, angitt ved vedlagte formel.

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

Korpus:

Ballen suste rett i mål. (CONCR)

Mitt mål i livet er å bli forfatter. (ABSTR)

Jeg fikk inn tre mål i den kampen. (CONCR)

Jeg kommer til å nå mitt mål. (ABSTR)

13 Sannsynlighet (5 poeng)

Hva er (prior) sannsynligheten for betydningen **CONCR** $\in B$ i korpuset, altså **P(CONCR)**?

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

Korpus:

Ballen suste rett i mål. (CONCR)

Mitt mål i livet er å bli forfatter. (ABSTR)

Jeg fikk inn tre mål i den kampen. (CONCR)

Jeg kommer til å nå mitt mål. (ABSTR)

Skriv ditt svar her

Maks poeng: 5

14 Sannsynlighet for individuelle trekk (5 poeng)

Hvordan beregner vi sannsynligheten for et individuelt trekk, gitt en betydning $P(v_j|b)$? Anta en bag-of-words representasjon av forekomstene i korpuset, slik at ordet "mitt" er et av trekkene.. Hva blir da $P(\text{mitt}|\text{ABSTR})$?

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

Korpus:

Ballen suste rett i mål. (CONCR)

Mitt mål i livet er å bli forfatter. (ABSTR)

Jeg fikk inn tre mål i den kampen. (CONCR)

Jeg kommer til å nå mitt mål. (ABSTR)

Skriv ditt svar her

Maks poeng: 5

15 Naive Bayes-formelen (5 poeng)

Forklar hvordan Naive Bayes-formelen er utledet fra den opprinnelige sannsynligheten $P(b|v)$ og beskriv hvilke antagelser som er gjort.

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

Korpus:

Ballen suste rett i mål. (CONCR)

Mitt mål i livet er å bli forfatter. (ABSTR)

Jeg fikk inn tre mål i den kampen. (CONCR)

Jeg kommer til å nå mitt mål. (ABSTR)

Skriv ditt svar her

Maks poeng: 5

