

i Informasjon

Skriftlig hjemmeeksamen i IN1140 2021 HØST

Varighet: 29 november kl 09:00 til 29 november kl 13:00.

Det er viktig at du leser denne forsiden nøye før du starter.

Generell informasjon:

- Viktige beskjeder under eksamen blir gitt direkte fra faglærer på emnets semesterside. Det er derfor viktig at du sjekker emnets semesterside jevnlig.
- Besvarelsen din skal reflektere ditt eget, selvstendige arbeid og skal være et resultat av din egen læring og arbeidsinnsats.
- Alle hjelpemidler er tillatt ved skriftlig hjemmeeksamen. Dersom du gjengir tekst fra bøker, nettartikler eller lignende, så må det henvises til disse kildene i besvarelsen for å unngå mistanke om ulovlig tekstlikhet. Dette gjelder også dersom du oversetter tekst fra andre språk.
- Du er selv ansvarlig for å sørge for at eksamensbesvarelsen din ikke er tilgjengelig for andre under eksamenstiden, hverken fysisk eller digitalt.
- Husk at besvarelsen skal være anonym, du skal ikke oppgi hverken ditt eller medstudenters navn.
- Om du vil trekke deg fra eksamen, trykk på hamburgermeny oppe til høyre i Inspera og velg "Jeg vil trekke meg".

Viktig informasjon:

Om du bruker nettleseren Google Chrome og har lagt til utvidelsen "Tex All The Things" (f.eks. for å bruke Padlet), må du skru av denne utvidelsen under eksamen. Om utvidelsen er slått på, vil ikke oppgaveteksten vises riktig.

Samarbeid under eksamen:

Det er ikke tillatt å samarbeide eller kommunisere med andre under eksamen. Samarbeid og kommunikasjon vil bli betraktet som forsøk på fusk. Det blir gjort plagiatskontroll av alle innleverte eksamener der tekstlikhet mellom besvarelser blir sjekket. Om du bruker notater som er utarbeidet i samarbeid med andre før eksamen, kan dette gi treff i en plagiatsjekk. Slik tekstlikhet kan bli betraktet som lav selvstendighet eller forsøk på fusk. Unngå derfor klipp/lim fra notater laget i samarbeid med andre.

Fusk:

Les om [hva som regnes som fusk på UiOs nettsider](#).

Kontaktinfo:

[Brukerstøtte eksamen](#)

1 Telefonnummer

Anta følgende regulære uttrykk:

regex = "`((\+47|0047)\s)?([0-9][0-9]\s){3}([0-9][0-9]){1}`"

Hvilken av følgende påstander om uttrykket er **usann**? Begrunn svaret ditt.

1. Det regulære uttrykket gjenkjenner kun norske telefonnumre som begynner med +47 eller 0047
2. Det regulære uttrykket gjenkjenner norske telefonnumre som kan begynne med +47 eller 0047

Skriv ditt svar her

Maks poeng: 4

2 Postadresser

Skriv et regulært uttrykk som sjekker om en postadresse er gyldig i Norge. Merk at for at posten skal kunne levere brevene/pakkene, må adressen skrives på ett av følgende formater:

Fornavn Etternavn
Gatenavn Husnummer
Postkode By

Fornavn Etternavn
P.O. BOX Nummer
Postkode By

Fornavn Etternavn
Postboks Nummer
Postkode By

Merk følgende formateringer:

- Husnummer kan være tall mellom 01 og 99.
- Nummer kan være tall mellom 0001 og 9999.
- Postkode kan være tall mellom 0001 og 5999.
- Det skal være kun ett fornavn, og kun ett etternavn.
- Det er mellomrom mellom hver element, utenom etter Etternavn, Husnummer, og Nummer, der skal det være ny linje.
- Uttrykket ditt må kunne håndtere de norske bokstavene æ, ø, å, Æ, Ø, og Å.

Eksempelvis skal følgende strenger gjenkjennes av ditt regulære uttrykk:

Per Post
Storgata 15
0155 Oslo

Bedriften AS
P.O. BOX 9999
3705 Skien

Julius Box
Postboks 6890
2712 Brandbu

Skriv ditt svar her

Maks poeng: 15

3 Tokenisering

1. Beskriv kort hva **tokenisering** er og hvordan det typisk inngår i et språkteknologisk system. Hvilke andre språkteknologiske oppgaver er avhengig av tokenisering? Gi minst ett eksempel.
2. Hvordan kan output fra en tokeniserer se ut for denne teksten? Skriv en tokenisert versjon av teksten under.
3. Tenk deg at du skal lage en regelbasert tokeniserer for teksten under. Hvilke utfordringer ser du som du må ta høyde for i tokenisereren din? Gi minst tre eksempler fra teksten.

Vi søker motiverte medarbeidere og "Team-players" til vår nye supersatsing! Du må kunne salg og være deg selv 110%. For mer informasjon om lønn, arbeidvilkår, osv., ta kontakt via vår hjemmeside <http://www.supersjansen.no>. Vi ser fram til å høre fra deg!

Skriv ditt svar her

Maks poeng: 15

4 Språkmodell

Anta at vi vil trene en bigram-modell på følgende korpus:

<s> jeg er rar </s>
 <s> rar er jeg </s>
 <s> rar er rar </s>

Besvar følgende spørsmål:

1. Hvor mange *unike* bigram forekommer i korpuset? Oppgi bigrammene i svaret ditt.
2. Vis hvordan du kan beregne sannsynligheten for hvert bigram fra korpuset.
3. Gitt bigram-sannsynlighetene du beregnet over, hva er sannsynligheten for følgende setning? Vis hvordan du kom fram til svaret ditt.

<s> jeg er jeg </s>

Skriv ditt svar her

Maks poeng: 15

5 Ordklasser

Her skal vi jobbe med følgende setning:

Når vinteren senker seg sakte over byen er det tid for et varmt bad

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
PRON	Pronomen
PREP	Preposisjon
CONJ	Konjunksjon
SUBJ	Subjunksjon
ADV	Adverb
DET	Determinativ

Finne de som passer sammen (her får du poeng for riktig svar, men ikke negative poeng for feil svar):

	NOUN	ADJ	VERB	PRON	PREP	CONJ	SUBJ	ADV	DET
Når	<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"/>
vinteren	<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"/>
senker	<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"/>
seg	<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"/>
sakte	<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"/>
over	<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"/>
byen	<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"/>
er	<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"/>
det	<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"/>
tid	<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"/>
for	<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"/>
et	<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"/>
varmt	<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"/>
bad	<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: 7

6 Ordklassetagging

Vi ønsker å lage et system for å automatiske merke opp en tekst med ordklasser. Som et utgangspunkt utvikler vi et system der vi slår opp i et leksikon der hvert ord er assosiert med én ordklasse. Systemet vårt viser seg dessverre å ikke fungere særlig bra.

- Hva er grunnen til det tror du? Illustrer svaret ditt med eksempler.
- Hva kan vi gjøre for å forbedre systemet vårt?

Skriv ditt svar her

Maks poeng: 8

7 Grammatikk

Trærne på bildet utgjør en liten *trebank*, et syntaktisk annotert korpus som vi skal benytte oss av i denne oppgaven.

1. Utled en frasestrukturgrammatikk fra de syntaktiske analysene i korpuset og angi grammatikken under.
2. Er grammatikken din rekursiv? Begrunn svaret ditt.

Skriv ditt svar her

Maks poeng: 6

8 Utvid grammatikken

Hvilke regler må du utvide grammatikken din med for at den skal kunne gi en analyse for følgende setninger?

- En katt sitter eller sover
- En hund elsker en mann og hater en katt

Skriv ditt svar her

Maks poeng: 5

9 Ugrammatiske analyser

Grammatikken kan gi opphav til noen ugrammatiske setninger. Diskutér påstanden kort og illustrér med et eksempel.

Skriv ditt svar her

Maks poeng: 4

10 Konstituentkriterier

Vi ønsker å avgjøre hvorvidt "en hund og en katt" er en konstituent i følgende setning:

- en hund og en katt fanger en fugl

Vis hvordan du kan bruke minst to konstituentkriterier for å avgjøre dette.

Skriv ditt svar her

Maks poeng: 5

11 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.

	antonymi	hyponymi	meronymi	synonymi
fot -- tå	<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"/>
inne -- ute	<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"/>
pen -- vakker	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 3

12 Semantiske roller

Angi semantisk rolle for de **uthevede** ordene (her får du poeng for riktig svar, men ikke negative poeng for feil svar):

	GOAL	SOURCE	INSTRUMENT	AGENT	THEME	EXPERIENCE
Jon rører i suppa	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jon kjører mot *Bergen*	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Han kommer fra *Oslo*	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jon hører en lyd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nøkkelen åpner døren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jon skjærer brød med *en kniv*	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Maks poeng: 3

13 Evaluering

En måte å evaluere dialogsystemer på er ved å be mennesker manuelt analysere og se gjennom to dialoger produsert av systemet, og velge den dialogen de synes var best.

Da vi evaluerte vår Naive Bayes sentimentklassifiserer, brukte vi imidlertid en annen måte å evaluere modellen vår på. Da hadde vi en såkalt "gullstandard", og vi sammenlignet klassene til klassifisereren vår med de virkelige sanne klassene i gullstandard.

1. Beskriv kort hvordan vi evaluerte maskinlæringsmodellen vår ved å bruke disse gullstandardklassene.
2. Å ha gullstandardannotasjoner for et dialogsystem er ikke alltid ønskelig. Om vi ønsker å ha et system som kan kommunisere med mennesker, og som kan håndtere flere typer samtaler trenger vi et system som kan "kopiere" menneskelige samtaler uten å følge en mal. Kan du tenke deg noen utfordringer innen menneske-menneske kommunikasjon som er viktige å kunne håndtere og som gjør det vanskelig å forholde seg til en gullstandard?

Skriv ditt svar her

Maks poeng: 10

Question 7
Attached



