

# UNIVERSITETET I OSLO

## Det matematisk-naturvitenskapelige fakultet

Eksamen i: INF1820 — Introduksjon til språk- og  
kommunikasjonsteknologi

Eksamensdag: 17. juni 2016

Tid for eksamen: 14.30 – 18.30

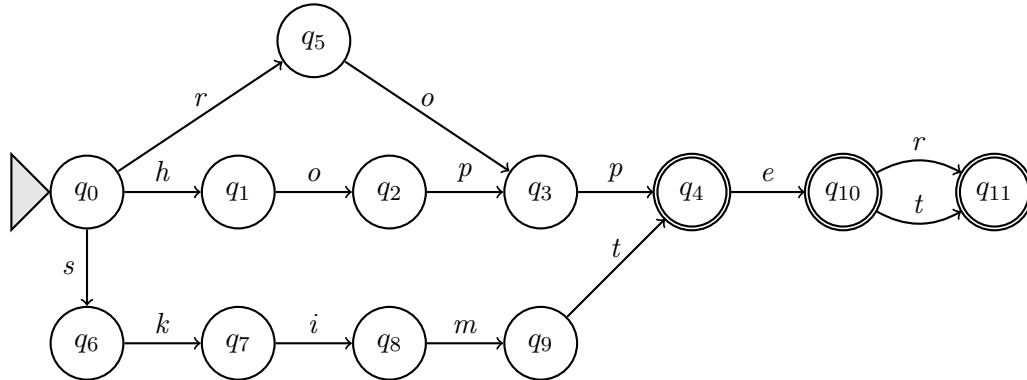
Oppgavesettet er på 3 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Ingen

Kontroller at oppgavesettet er komplett før  
du begynner å besvare spørsmålene.

## Oppgave 1 Tilstandsmaskiner og regulære uttrykk (vekt 25%)



1. I figuren over ser du en tilstandsautomat som gjenkjenner enkelte norske verbformer. Hvilke former av hvilke verb gjenkjenner automaten? Du trenger ikke liste opp alle formene, men beskriv alle mulighetene for automaten.
2. Skriv et regulært uttrykk som gjenkjenner de samme formene som automaten.
3. I faget har vi snakket en del om skillet mellom deterministiske og ikke-deterministiske automater. Hva kjennetegner en deterministisk automat?

## Oppgave 2 Morfologi (vekt 25%)

1. Vi skiller mellom såkalte åpne og lukkede ordklasser. Hva skiller de to typene ordklasse fra hverandre?
2. Av de ti ordklassene vi opererer med på norsk, hvilke regner vi som åpne?

$$P(\vec{t}|\vec{w}) = \prod_{i=1}^N P(t_i|t_{i-1})P(w_i|t_i)$$

3. Ligningen over viser sannsynligheten en HMM-modell tilordner en taggsekvens gitt en ordsekvens. Hvilke sannsynligheter trenger vi for å regne ut sannsynligheten for at setningen Babyen begynte å gå tidlig har taggsekvensen subst verb sbu verb adv?
4. Et viktig problem for HMM-modeller er ukjente ord. Beskriv kort hvilket problem ukjente ord medfører for en HMM-modell og hvordan vi løser det.

(Fortsettes på side 3.)

### Oppgave 3 Syntaks (vekt 25%)

Anta følgende grammatikk for et fragment av norsk:

$$\begin{array}{ll}
 S \rightarrow NP VP & D \rightarrow \text{en} \mid \text{et} \\
 NP \rightarrow D N & N \rightarrow \text{baby} \mid \text{vindu} \mid \text{vugge} \\
 NP \rightarrow NP PP & V \rightarrow \text{sover} \\
 VP \rightarrow V \mid V NP & P \rightarrow \text{i} \mid \text{ved} \\
 VP \rightarrow VP PP & \\
 PP \rightarrow P NP & 
 \end{array}$$

1. Hvor mange analyser tildeler grammatikken til setningene under?
  - (a) en baby sover i en vugge ved et vindu
  - (b) et barn i en vugge sover
2. Tegn treet eller trærne grammatikken tilordner setning (a).
3. Utvid grammatikken slik at den tillater vilkårlig mange adjektiver foran substantivet i substantivfrasen og godtar setninger som:
  - (a) en liten søt baby sover i en stor vugge
4. Hvor mange analyser tilordner den utvidede grammatikken din til setningen en rund liten søt baby sover?

### Oppgave 4 Semantikk (vekt 25%)

1. Vi sier at *Erna Solberg* og *statsministeren* har lik referanse, men forskjellig betydning. Hva mener vi med dette?
2. Forklar kort hva en semantisk rolle er, og hvordan vi bruker semantiske roller til å analysere setninger.
3. Analyser setningen Den søte babyen kaster grøt. ved hjelp av semantiske roller. Beskriv de forskjellige rollene du har tilordnet.
4. I informasjonsgjenfinning (Information Retrieval) brukes som regel den såkalte *vektorrommodellen* for å representere dokumenter. Beskriv kort hvordan dokumenter representeres i denne modellen.
5. I en vektorrommodell brukes ofte *tf-idf*-vekting for å forbedre modellen. Hvordan regner vi ut idf-vekten, og hva er hensikten med denne vektingen?