

INF1820 V2013 — Oppgave 1b

Regulære uttrykk og endelige tilstandsmaskiner

Innleveringsfrist, onsdag 25. februar

Lever inn svarene dine med Devilry (<https://devilry.ifi.uio.no>) i en fil som angir brukernavnet ditt, slik: `oblig1a_brukernavn.py`

En perfekt besvarelse på denne oppgaven er verdt 100 poeng.

1 Tostavelsesord i “Anna Karenina” (20 poeng)

I denne oppgaven skal vi se nærmere på en tekst, som i oblig1a, men ikke bare en liten bit. Denne gangen skal vi studere hele romanen. Last ned den engelske oversettelsen av *Anna Karenina* fra <http://www.gutenberg.org/ebooks/1399> (pass på at du laster ned versjonen som er ren tekst, ikke HTML-versjonen).

Aller først skal vi telle tostavelsesord i romanen, hvor vi sier at en stavelse er null eller fler konsonanter, minst én vokal, og så null eller fler konsonanter. Her skal vi ikke skille mellom liten og stor bokstav. Du kan enten håndtere dette i det regulære uttrykket, eller ved å gjøre alt til små bokstaver. Hvor mange slike ord er det?

Men denne definisjonen er ikke perfekt. Det finnes ord som ikke har to stavelser som dekkes av uttrykket, og det er ord som ikke dekkes av uttrykket som likevel har to stavelser. Finn eksempler på begge typene feil.

For å oppsummere:

1. Skriv et program som leser inn *Anna Karenina* og skriver ut antall tostavelsesord (ikke hele listen!)
2. Gi minst tre eksempler på ord som dekkes av uttrykket men ikke er tostavelsesord. Finner du et eksempel på et tostavelsesord som *ikke* dekkes av uttrykket?

2 Regulære uttrykk for datouttrykk (50 poeng)

Det finnes mange måter å angi dato'er på. For eksempel: “neste torsdag”, “mandag 13 september, 2010”, “02/07/2011”.

En ting man kan gjøre med et system som henter ut datouttrykk er å lage en applikasjon som identifiserer møtetidspunkter i e-postene dine og automatisk legger disse til i kalenderen. (Merk dog at en slik applikasjon vil måtte være svært pålitelig for å være nyttig.)

Skriv regulære uttrykk som dekker følgende typer datouttrykk:

- neste/forrige ukedag
- ukedag dag måned, år

Der **ukedag** er en av mandag, tirsdag, ..., søndag, **måned** er en av januar, februar, ..., desember, og **dag** er et tall mellom 1 og 31. (Du trenger ikke sjekke at du ikke matcher 31 februar og andre umulige datoer). **år** er et tall mellom 1900 og 2099. Test de regulære uttrykkene på minst 6 setninger som inneholder datouttrykk av hver type over og skriv ut resultatet.

3 Endelige tilstandsmaskiner (30 poeng)

I denne oppgaven skal vi bruke eksternt og fritt tilgjengelig programvare for å tegne FSA'er, nemlig JFLAP. Dette må dere laste ned selv fra <http://www.cs.duke.edu/csed/jflap/jflaptmp/>.

Før du besvarer oppgavene, eksperimenter litt med programmet og finn ut av følgende:

- hvordan legger man til tilstander?
- hvordan markerer man en tilstand som start-og slutttilstand?
- hvordan legger man inn kanter fra en tilstand til en annen?
- hvordan tester man maskinen på input?

Dersom du ikke får til dette, bør du se på en tutorial som ligger på JFLAPs nettside (<http://www.cs.duke.edu/csed/jflap/tutorial/>).

Tegn en endelig tilstandsmaskin i JFLAP som gjenkjenner følgende språk, der alfabetet er $\{a, b\}$. Alle tilstandsmaskinene skal være deterministiske.

1. $\{w | w \text{ inneholder minst sekvensen } ab\}$

2. $\{w \mid \text{hver oddetallsposisjon i } w \text{ er en } b\}$
3. $\{w \mid w \text{ begynner ikke med } aba\}$

Lagre de ferdige maskinene som GIF-filer (File \rightarrow Save Image As), og lever dem sammen med koden din som separate filer med navnene **3a.gif**, **3b.gif** og **3c.gif**.