INF 2080 – vår 2015

hrj – 2

Kap 1B

Regulære uttrykk (RE)

- Bygd opp fra
  - Symbolene i alfabetet
- Ved bruk av
  - Union
  - Konkatenering
  - Kleene stjerne

- Tolkes som mengder av ord – dvs språk – i alfabetet
- $ab^*(ba \cup bb)$ $(aaa)^*$

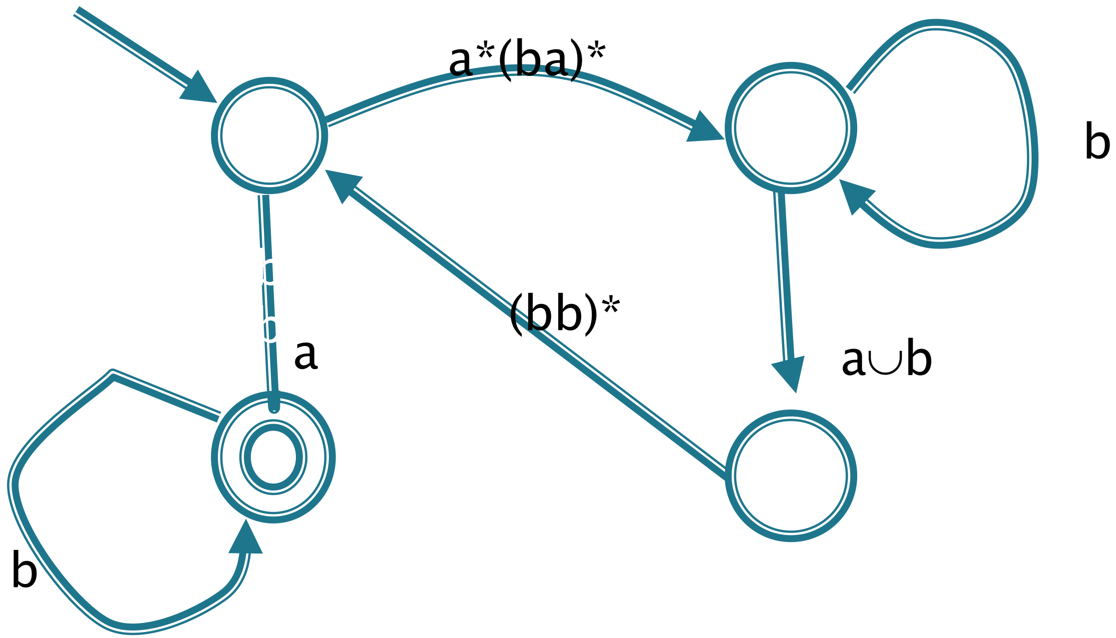
RE = Reg språk = DFA = NFA

- RE brukes i
  - Perl
  - grep
  - Søke etter uttrykk

- Lete etter mississippi i en fil

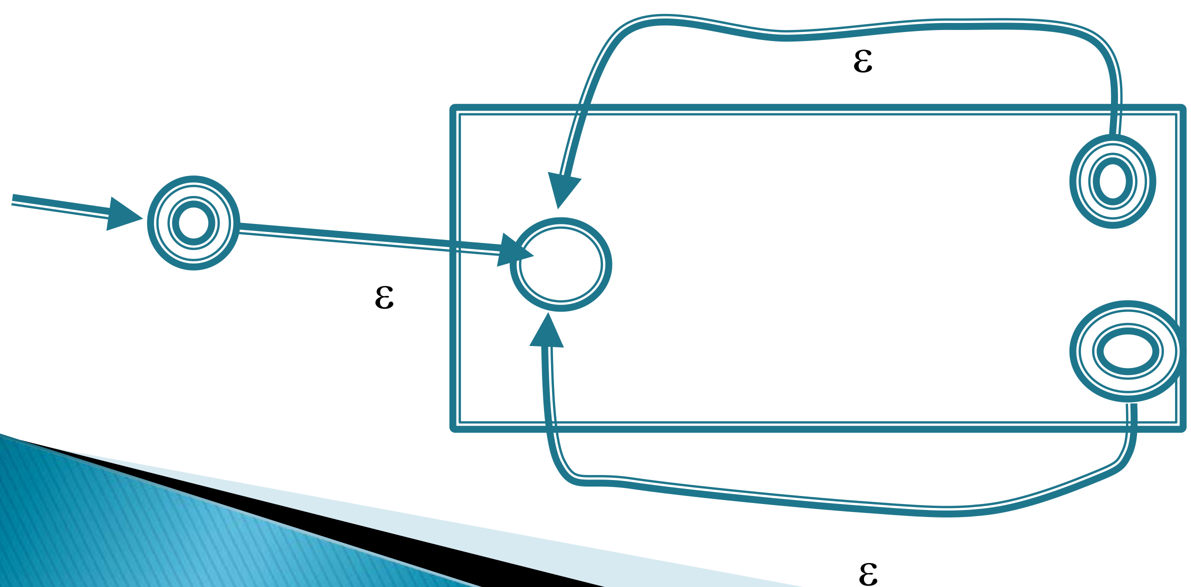
GNFA

- Ikke deterministisk automat med regulære uttrykk på pilene



RE ⊆ NFA

- Nok å vise at NFA'ene inneholder
  - Symbolene i alfabetet og det tomme språket
- Samt er lukket under
  - Union, konkatenering og Kleene stjerne
    - Enkel konstruksjon – pass på konstruksjonen av Kleene stjerne



NFA $\subseteq$ GNFA $\subseteq$ DFA

- Den første er opplagt
- Den andre vises ved å bruke makrotilstander

- Det er vanskeligere å vise den andre veien –at gitt en vilkårlig DFA så kan vi finne et regulært uttrykk som gir alle veier fra start til akseptering.

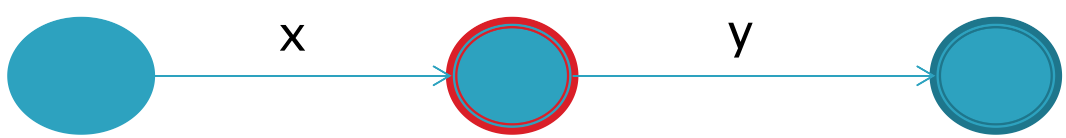
Idé bak DFA $\subseteq$ RE – 1

- Gitt en DFA M med regulært språk S
- Ved å bruke ε–transisjoner lager vi NFA N for språket S med enkel start og akseptering

Erstatter så alt inne i boksen med et regulært uttrykk. Dette uttrykket gir språket S.

Idé bak DFA $\subseteq$ RE – 2

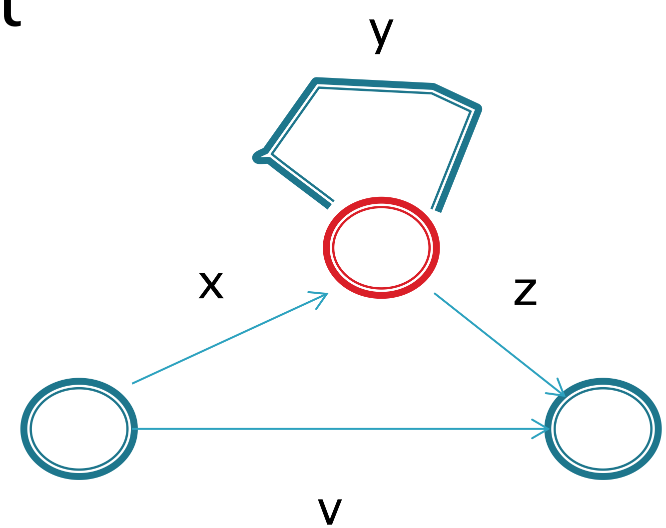
- Vi tar vekk en-og-en tilstand i boksen, og reparerer ved å bruke regulære uttrykk i transisjonene i stedet for tilstanden som er tatt vekk. Noen eksempler



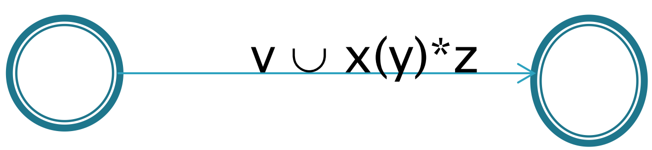
Erstattes med

Idé bak DFA $\subseteq$ RE – 3

- Erstatt



- med



DFA = NFA = RE

- Idéen bak likheten
- Flere detaljer når tilstanden vi tar vekk er forbundet med flere tilstander – dette overlates dere
- Observer at vi trenger akkurat union, konkatenering og Kleene stjerne i prosessen

- De regulære uttrykkene beskriver akkurat de veiene som går fra start til akseptering

Problemer med lange ord

- Gitt en DFA M med m tilstander
- Et ord satt inn i M gir opphav til en vei gjennom tilstandene – veien er like lang som ordet
- Om lengden er større enn antall tilstander, så må veien gå i en løkke.

- Dette er hva vi utnytter i pumpelemmaet – og som vi bruker til å vise at noen språk ikke er regulære.

Ord og veier i DFA'er

- Ethvert ord gir en entydig vei
- Om veien går i en løkke, så kan vi finne ord som gir veier som går i 0, 1 eller flere ganger i den samme løkka
- Gitt DFA M med m tilstander, da vil ord med lengde > m , gi vei med løkke
- Gitt akseptert ord x. Ved å gå rundt i løkke kan vi finne mange andre aksepterte ord.

Pumpelemmaet

- Om A er et regulært språk, så finnes det et tall p, slik at for ethvert ord s i A av lengde $\geq p$, så kan s deles opp i tre deler $s = xyz$ slik at
  - For hver $i \geq 0$, så er xy^iz med i A
  - $|y| > 0$
  - $|xy| \leq p$

Teknisk formulering. Idéen er den samme som foregående slides.

Parentesspråket P

- Alfabetet ( )
- Språket ε () (()) ()() (()()) ……
- P er ikke regulært

- Viser først at språket Q av alle ord $(^i)^i$ – først venstreparenteser, og så like mange høyre parenteser – ikke er regulært
- Dette følger av pumpelemmaet
- Men som vi viser på neste slide, da kan heller ikke P være regulær

Lukking under snitt

- De regulære språk er lukket under
  - Komplement, union, konkatenering, stjerne
- Observer $A \cap B = -(-A \cup -B)$
  - Regulær lukket under snitt
- Nå er (*)* regulært
- Men $Q = (^*)^* \cap P$ – slik at om P er regulær, så er også Q regulær. Motsigelse.
- Parentesspråket er ikke regulært