Uke 45: Endelige tilstandsmaskiner og regulære språk

Oskar Goldhahn

Høst 2024

Språk (repetisjon)

Git et alfabet A er et *språk* en delmengde av A^*. Dvs. en mengde av endelige strenger over A. Vi bruker tegnet Λ for den tomme strengen.

Språk (repetisjon)

Git et alfabet A er et *språk* en delmengde av A^*. Dvs. en mengde av endelige strenger over A. Vi bruker tegnet Λ for den tomme strengen.

Mange problemer i datavitenskapen kan formuleres som spørsmål på følgende form; Finn en effektiv metode som gitt en $x \in A^*$ avgjør om $x \in L$ eller $x \notin L$.

Eksempler

- Grafisomorfiproblemet
- Skille mellom krypteringer av 0 og 1

Operasjoner på språk

Union: Gitt to språk L og M er $L \cup M$ også et språk.

Operasjoner på språk

Union: Gitt to språk L og M er $L \cup M$ også et språk.

Konkatenering: Gitt to språk L og M kan vi konkatenere dem til LM, mengden med strenger der første del er en streng i L og andre er i M.

Eksempel $L = \{a, b, ab\}$, $M = \{a, b, ba\}$:

$$LM = \{aa, ab, aba, ba, bb, bba, abb, abba\}$$

Operasjoner på språk

Union: Gitt to språk L og M er $L \cup M$ også et språk.

Konkatenering: Gitt to språk L og M kan vi konkatenere dem til LM, mengden med strenger der første del er en streng i L og andre er i M.

Eksempel $L = \{a, b, ab\}$, $M = \{a, b, ba\}$:

$$LM = \{aa, ab, aba, ba, bb, bba, abb, abba\}$$

Obs: Vi glemmer hvilken del som kommer fra L og M.

Kleene-tillukning

Vi kan gjenta konkatenering av samme språk. Vi definerer det rekursivt $L^0 = \{\Lambda\}$ og $L^{n+1} = LL^n$. slik at $L^1 = L$, $L^2 = LL$ og $L^3 = LLL$.

Kleene-tillukning

Vi kan gjenta konkatenering av samme språk. Vi definerer det rekursivt $L^0 = \{\Lambda\}$ og $L^{n+1} = LL^n$. slik at $L^1 = L$, $L^2 = LL$ og $L^3 = LLL$.

Gitt et språk L så er *kleene-tillukningen* L^* mengden av alle konkateneringer av vilkårlig mange strenger fra L.

$$L^* = \bigcup_{i \in \mathbb{N}} L^i = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \ldots$$

Kleene-tillukning

Vi kan gjenta konkatenering av samme språk. Vi definerer det rekursivt $L^0 = \{\Lambda\}$ og $L^{n+1} = LL^n$. slik at $L^1 = L$, $L^2 = LL$ og $L^3 = LLL$.

Gitt et språk L så er *kleene-tillukningen* L^* mengden av alle konkateneringer av vilkårlig mange strenger fra L.

$$L^* = \bigcup_{i \in \mathbb{N}} L^i = L^0 \cup L^1 \cup L^2 \cup L^3 \cup \ldots$$

Eksempel med $L = \{a, ab\}$:

$$L^* = \{\Lambda, a, ab, aa, aab, aba, abab, aaa, aaab, aaba, abaa, abaab, aabab, ababa, ababab, \ldots\}$$

Regulære Språk

Defineres rekursivt. $\emptyset$ og $\{\Lambda\}$ er regulære språk og $\{a\}$ er et regulært språk for alle a i alfabetet. Gitt to språk L og M så er $L \cup M$, LM og L^* regulære språk.

Regulære Språk

Defineres rekursivt. $\emptyset$ og $\{\Lambda\}$ er regulære språk og $\{a\}$ er et regulært språk for alle a i alfabetet. Gitt to språk L og M så er $L \cup M$, LM og L^* regulære språk.

Eksempler med $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- Språk med partall
  $= (\{0\} \cup \{1\} \cup \ldots \cup \{9\})^*(\{0\} \cup \{2\} \cup \ldots \cup \{8\})$

Regulære Språk

Defineres rekursivt. $\emptyset$ og $\{\Lambda\}$ er regulære språk og $\{a\}$ er et regulært språk for alle a i alfabetet. Gitt to språk L og M så er $L \cup M$, LM og L^* regulære språk.

Eksempler med $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- Språk med partall
  $= (\{0\} \cup \{1\} \cup \ldots \cup \{9\})^*(\{0\} \cup \{2\} \cup \ldots \cup \{8\})$

Eksempler med $A = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \ldots, \mathsf{å}, \mathsf{@}, \text{"."}\}$.

- Språk med gyldige ntnu-epostadresser $= (\{\mathsf{a}\} \cup \ldots \cup \{\mathsf{å}\})^*\{.\}(\{\mathsf{a}\} \cup \ldots \cup \{\mathsf{å}\})^*\{\mathsf{@}\}\{\mathsf{n}\}\{\mathsf{t}\}\{\mathsf{n}\}\{\mathsf{u}\}\{.\}\{\mathsf{n}\}\{\mathsf{o}\}$

Regulære Språk

Defineres rekursivt. $\emptyset$ og $\{\Lambda\}$ er regulære språk og $\{a\}$ er et regulært språk for alle a i alfabetet. Gitt to språk L og M så er $L \cup M$, LM og L^* regulære språk.

Eksempler med $A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

- Språk med partall
  $= (\{0\} \cup \{1\} \cup \ldots \cup \{9\})^*(\{0\} \cup \{2\} \cup \ldots \cup \{8\})$

Eksempler med $A = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \ldots, \mathsf{å}, \mathsf{@}, \text{"."}\}$.

- Språk med gyldige ntnu-epostadresser $= (\{\mathsf{a}\} \cup \ldots \cup \{\mathsf{å}\})^*\{.\}(\{\mathsf{a}\} \cup \ldots \cup \{\mathsf{å}\})^*\{\mathsf{@}\}\{\mathsf{n}\}\{\mathsf{t}\}\{\mathsf{n}\}\{\mathsf{u}\}\{.\}\{\mathsf{n}\}\{\mathsf{o}\}$

I prasis er det mye lettere å bestemme medlemskap i et regulært språk enn i et vilkårlig språk

Regulære uttrykk

Regulære uttrykk er strenger som kan brukes til å beskrive regulære språk.

Regulære uttrykk

Regulære uttrykk er strenger som kan brukes til å beskrive regulære språk.

$\emptyset$ og λ er regulære uttrykk og alle $a \in A$ er regulære uttrykk. Gitt to regulære uttrykk R og S er (R), $R|S$, RS og R^* også regulære uttrykk.

$R|S$ korresponderer til union av språk, RS korresponderer til konkatenering og R^* korresponderer til Kleene-tillukning.

Eksempel:

- Uttrykket som korresponderer til språket med gyldige ntnu-epostadresser: $(a|b|c|\ldots|\text{å})^*.(a|b|c|\ldots|\text{å})^*.ntnu.no$
- PartaII: $(0|1|2|3|4|5|6|7|8|9)^*(0|2|4|6|8)$

Deterministiske Endelige Tilstandsmaskiner

Generisk beregning må modelleres med Turingmaskiner, men i spesialtilfeller kan man klare seg med mye enklere modeller.

Hvis man har endelig mange tilstander kan man bruke endelige tilstandsmaskiner.

- En heis kan bare være i endelig mange etasjer
- En dør er enten lukket, låst eller åpen

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander
- En endelig mengde med symboler

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander
- En endelig mengde med symboler
- En starttilstand som er en av tilstandene

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander
- En endelig mengde med symboler
- En starttilstand som er en av tilstandene
- En mengde med aksepterende tilstander som er en delmengde av tilstandene

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander
- En endelig mengde med symboler
- En starttilstand som er en av tilstandene
- En mengde med aksepterende tilstander som er en delmengde av tilstandene
- En overgangsfunksjon som tar en tilstand og et symbol og spytter ut en ny tilstand

Deterministiske Endelige Tilstandsmaskiner

En endelig tilstandsmaskin har:

- En endelig mengde med tilstander
- En endelig mengde med symboler
- En starttilstand som er en av tilstandene
- En mengde med aksepterende tilstander som er en delmengde av tilstandene
- En overgangsfunksjon som tar en tilstand og et symbol og spytter ut en ny tilstand

Vi kan mate strenger inn i tilstandsmaskinen, bruke overgangsfunksjonen til å bevege oss videre for hvert tegn og se om maskinen aksepterer. Dette definerer et språk av strenger som den aksepterer. Vi sier at tilstandsmaskinen *gjenkjenner* dette språket.

Tilstandsmaskiner og språk

Det viser seg at alle *regulære* språk kan gjenkjennes av endelige tilstandsmaskiner.

Tilstandsmaskiner og språk

Det viser seg at alle *regulære* språk kan gjenkjennes av endelige tilstandsmaskiner.

Noen språk kan ikke gjenkjennes av endelige tilstandsmaskiner

Eksempel:

- $\{1^n 0^n | n \in N\}$
- gyldige uttrykk med multiplikasjon, addisjon og parenteser.