Oversættere, ugeopgave 1

Anders Bjerg Pedersen (andersbp@me.com)

15. november 2009

Opgave 1

a)

Vi anvender følgende regulære udtryk i Emacs:

```
<title>.*</title>
```

b)

Denne opgave kan gøres nærmest vilkårlig kompleks, alt efter hvor mange specialtegn man vil tillade i en email. Lad os for nemheds skyld antage, at vi kun tillader alfanumeriske tegn, enkeltestående '.', '_', '-' og '∼' foran @'et, og kun alfanumeriske tegn samt enkeltstående '.' og '-' efter.

```
[a-zA-Z0-9\-\_\~]+\(\.[a-zA-Z0-9\-\_\~]+\)*
@ [a-zA-Z0-9]\([a-zA-Z0-9\-]*[a-zA-Z0-9]\)?
\(\.[a-zA-Z0-9]\([a-zA-Z0-9\-]*[a-zA-Z0-9]\)?\)+
```

c)

Igen, her kan kompleksiteten øges ved at tilføje specialtegn og protokoller til de lovlige karakterer. Vi vælger desuden at se bort fra muligheden for IP-adresser, portnumre, brugernavne og passwords (til brug i FTP-adresser). Antagelsen er nu, at et "link" består af en protokol (http, https, ftp eller ftps) samt "://" efterfulgt af en serie alfanumeriske eller specialtegn. Desuden kan vi ikke tage højde for eventuelle sproglige tegn såsom komma eller punktum, som kunne tænkes at komme lige efter et link. Vi finder nu links med følgende regex:

```
\(ht\|f\)tps?://[a-zA-Z0-9_!~*'()\-\/&\?\.=]+
```

I Emacs laver vi nu vores regex replace på følgende måde (`<RET>` er et tryk på returtasten):

```
M-x replace-regexp <RET>
\(ht\|f\)tps?://[a-zA-Z0-9_!~*'()\-\/&\?\.=]+ <RET>
<a href="\&">\&</a> <RET>
```

Opgave 2

Da der ingen ε-transitioner forekommer i denne NFA, er $\varepsilon\text{-}closure(\{M\}) = M$ i alle beregninger. Vi har $s_0 = \{A\}$, $s'_0 = \varepsilon\text{-}closure(\{A\}) = \{A\}$ og initierer med $S' = \{s'_0\}$. Vi mærker s'_0 og beregner $move(s'_0, 0)$ og $move(s'_0, 1)$:

$$\begin{aligned}
move(s'_0, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{A\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{B\}) = \{B\} = s'_1 \\
move(s'_0, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{A\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\}
\end{aligned}$$

Vi fortsætter på lignende måde:

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}\} \\
move(s'_1, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{B\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{C\}) = \{C\} = s'_2 \\
move(s'_1, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{B\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\}
\end{aligned}$$

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}\} \\
move(s'_2, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{C\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{D, G\}) = \{D, G\} = s'_3 \\
move(s'_2, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{C\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{F\}) = \{F\} = s'_4
\end{aligned}$$

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}, \underline{s'_3}, s'_4\} \\
move(s'_3, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{D, G\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{F, G\}) = \{F, G\} = s'_5 \\
move(s'_3, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{D, G\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{E, H\}) = \{E, H\} = s'_6
\end{aligned}$$

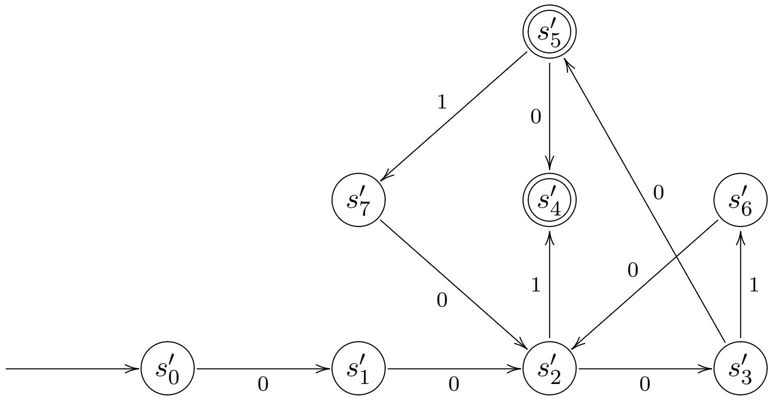
$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}, \underline{s'_3}, \underline{s'_4}, s'_5, s'_6\} \\
move(s'_4, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{F\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\} \\
move(s'_4, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{F\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\}
\end{aligned}$$

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}, \underline{s'_3}, \underline{s'_4}, \underline{s'_5}, s'_6\} \\
move(s'_5, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{F, G\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{F\}) = \{F\} = s'_4 \\
move(s'_5, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{F, G\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{H\}) = \{H\} = s'_7
\end{aligned}$$

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}, \underline{s'_3}, \underline{s'_4}, \underline{s'_5}, \underline{s'_6}, s'_7\} \\
move(s'_6, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{E, H\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{C\}) = \{C\} = s'_2 \\
move(s'_6, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{E, H\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\}
\end{aligned}$$

$$\begin{aligned}
S' &= \{\underline{s'_0}, \underline{s'_1}, \underline{s'_2}, \underline{s'_3}, \underline{s'_4}, \underline{s'_5}, \underline{s'_6}, \underline{s'_7}\} \\
move(s'_7, 0) &= \varepsilon\text{-}closure(\{t \mid s \in \{H\} \wedge s^0 t \in T\}) = \varepsilon\text{-}closure(\{C\}) = \{C\} = s'_2 \\
move(s'_7, 1) &= \varepsilon\text{-}closure(\{t \mid s \in \{H\} \wedge s^1 t \in T\}) = \varepsilon\text{-}closure(\{\}) = \{\}
\end{aligned}$$

Der er nu ikke flere umærkerde elementer tilbage i S', og vi kan konstruere vores DFA. Vores to accepterende tilstande er $F' = \{s'_5, s'_5\}$:



Opgave 3 (Exercise 2.5)

Vi opdeler indledningsvist vores DFA i to grupper:

$$\begin{aligned} \underline{G_1} &= \{0\} \\ G_2 &= \{1,2,3,4\} \end{aligned}$$

G_1 forbliver markeret under resten af denne gennemgang, da den allerede er singleton. Vi markerer derfor G_2 og tjekker, om den er konsistent:

G_2	a	b
1	G_2	–
2	G_2	–
3	G_2	G_1
4	G_2	G_1

G_2 er ikke konsistent, så vi opdeler i maksimale konsistente delgrupper:

$$\begin{aligned} \underline{G_1} &= \{0\} \\ G_3 &= \{1,2\} \\ G_4 &= \{3,4\} \end{aligned}$$

Vi ser nu på G_3:

G_3	a	b
1	G_3	–
2	G_4	–

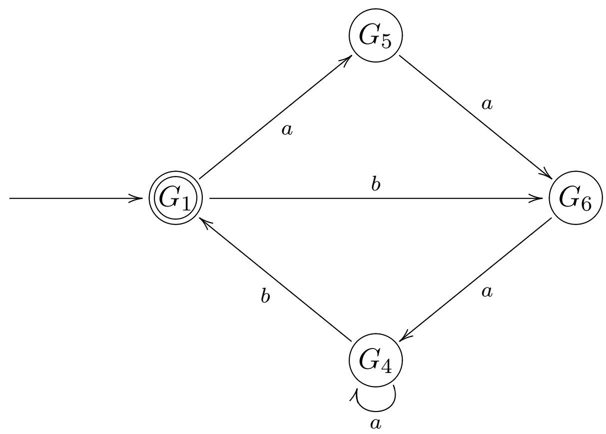
G_3 er ikke konsistent, så vi deler den op i G_5 og G_6, som begge nu er singletons:

$$\begin{aligned} \underline{G_1} &= \{0\} \\ G_4 &= \{3,4\} \\ \underline{G_5} &= \{1\} \\ \underline{G_6} &= \{2\} \end{aligned}$$

Vi ser nu på G_4:

G_4	a	b
3	G_4	G_1
4	G_4	G_1

G_4 er konsistent, så denne markerer vi nu. Da vi ikke har flere umarkerede grupper, svarer disse nu til en minimal DFA, som vi gengiver nedenfor:



Opgave 4

Nedenstående algoritme fjerner (forhåbentlig!) ε-transitioner fra en NFA en af gangen. Hver gang der fjernes en ε-transition, fjernes der også en tilstand, og vi kommer dermed tættere på en optimal NFA. Vi antager, at hvis vi under denne proces kommer frem til to eller flere identiske transitioner mellem de samme to tilstande, anses disse som værende den samme transition.

1. Vælg en vilkårlig ε-transition $s^{\varepsilon}t \in T$. $O(1)$
2. Se på alle udgående transitioner $t^{c_i}u_i$ fra t. Hvis $u_i = s$ og $c_i \neq \varepsilon$, tilføj transitionen $s^{c_i}s$ og gå til trin 1. Ellers erstat $t^{c_i}u_i$ med transitionen $s^{c_i}u_i$. $O(|T|)$
3. Erstat alle indgående transitioner $v^{c_j}t$ til t med transitionerne $v^{c_j}s$. $O(|T|)$
4. Hvis $t \in F$, så sæt $F = (F \setminus \{t\}) \cup \{s\}$. $O(1)$
5. Sæt $S = S \setminus \{t\}$. $O(|1|)$

Algoritmen gentages, til der ikke er flere ε-transitioner i vores NFA, dvs. $O(|T|)$ gange. Den samlede køretid bliver da lineær i antallet af transitioner, og derfor også specielt polynomiel. Dette er dog selvfølgelig afhængigt af brugen af effektive datastrukturer (hvor ε-transitioner kan findes hurtigt, samt hvor transitioner til og fra tilstande kan findes hurtigt).