

Indhold

1	Endelige Automater	2
1.1	Deterministiske Endelige Automater	2
1.2	Nondeterministiske Endelige Automater	6
1.3	Regulære Operationer	7
1.4	Pumpelemmaet og ikke-regulære sprog	10
1.5	Opgaver	12

1

Endelige Automater

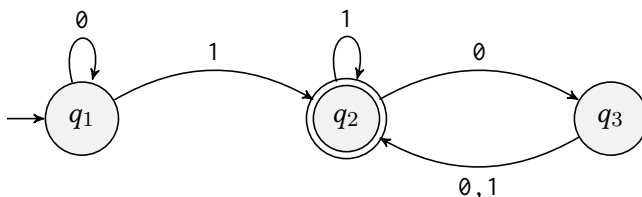
Endelige automater er en type abstrakt maskine med endelig hukommelse, som kan bruges til for eksempel at matche mønstre med.

1.1 Deterministiske Endelige Automater

I en deterministisk endelig automat (DFA) vil man altid vide hvilken state er den næste man havner i, efter et givet symbol. Dette bestemmes ud fra *transitionsfunktionen*, som tager en state, et symbol, og giver en ny state som output. Denne nye state er automatens næste state.

Vi kan beskrive en DFA grafisk ved hjælp af *state diagrams*. Et eksempel på et sådan diagram kan ses i Figur 1.1. Som kan ses på denne figur består den af nogle cirkler, states, nogle pile, transition pile, og en underlig dobbelt-cirkel, accept state. States er en form for hukommelse som en DFA kan være i på et givet tidspunkt. Efter at have læst tre karakterer af en streng, kan den eksempelvis være i q_3 , eller q_{90} , eller q_{fisk} , eller bare *fisk*. Pointen er at navnet er ligegyldig, og der er

ikke en nødvendig rækkefølge for *alle* strenge. Den første pil, som peger ind til q_1 er startpilen, som indikerer *initial state*, altså den state der startes i når strengen skal læses. q_2 , som er den state med dobbeltcirkler, er den *accepterende state*, og betyder at hvis en DFA ender i denne state når den er færdig med at læse en streng, er strengen accepteret. Hvis den ikke ender i denne state, er strengen ikke accepteret. Det skal siges at det er tilladt at have mere end en accept state.



Figur 1.1: Et eksempel på et state diagram for en automat M_1

Spørgsmålet om hvor mange accept states der kan være, samt hvordan transitionspilene fungerer, samt mere, kan forklares ved hjælp af en matematisk definition. Vi introducerer nu den formelle definition af en deterministisk endelig automat:

Definition 1.1

En deterministisk endelig automat er en 5-tuple $(Q, \Sigma, \delta, q_0, F)$, hvor:

1. Q er et endeligt sæt af states (hukommelse)
2. Σ er et endeligt sæt af symboler kaldet alfabetet
3. δ er *transitionsfunktionen*
4. q_0 er den state maskinen starter i (start staten)

5. F er et endeligt sæt af *accept states* ($F \subseteq Q$)

For at give et eksempel beskriver vi M_1 i Figur 1.1 formelt.
 $M_1 = (Q, \Sigma, \delta, q_1, F)$, hvor

1. $Q = \{q_1, q_2, q_3\}$

2. $\Sigma = \{\emptyset, 1\}$

3. δ beskrives som en tabel:

	\emptyset	1
q_1	q_1	q_2
q_2	q_3	q_2
q_3	q_2	q_2

4. q_1 er start staten

5. $F = \{q_2\}$

Vi siger at, hvis A er sættet af alle strenge som M accepterer, så er A sproget af maskinen M , eller $L(M) = A$. Ligeledes siger vi at M genkender A , eller M accepterer A . Dog er “genkender” mest brugt, da accept har andre betydninger her.

Læg mærke til at **hvis en maskine ikke accepterer nogen strenge**, så genkender den stadig et sprog: det tomme sprog, \emptyset .

Vi vil gerne have en bedre forståelse af hvordan en DFA egentlig *komputerer*, altså, hvordan den kommer fra A til B . Vi introducerer her en formel definition af komputering:

Definition 1.2

Lad $M = (Q, \Sigma, \delta, q_0, F)$ være en endelig automat, og lad $w = w_1 w_2 \cdots w_n$

være en streng hvor alle w_i er et symbol som er medlem af alfabetet Σ . M accepterer her w hvis en sekvens af states r_0, r_1, \dots, r_n i Q eksisterer med tre betingelser:

1. $r_0 = q_0$
2. $\delta(r_i, w_{i+1}) = r_{i+1}$ for $i = 0, \dots, n - 1$
3. $r_n \in F$

Det vil altså sige at **betingelse 1** siger at maskinen starter i den initiale state. **Betingelse 2** siger at ud fra reglerne af transitionsfunktionen går maskinen fra state til state. **Betingelse 3** siger at maskinen accepterer dets input hvis den ender i en accept state.

Definition 1.3

Et sprog kaldes et **regulært sprog** hvis en endelig automat genkender det.

1.2 Nondeterministische Endelige Automater

1.3 Regulære Operationer

Frem for at have dette i DFA sektionen som der bliver gjort i Sipser, mener jeg at det giver mere mening at have den til sidst, hvor man kender til alle koncepterne. De regulære operationer er operationer der kan bruges på sprog, som resulterer i et nyt regulært sprog. Vi vil kigge på følgende regulære operationer:

Definition 1.4

Lad A og B være sprog. Vi definerer de regulære operationer **fællesmængde** (union), **sammenkædning** (concatenation) og **stjerne** (star, kleene star), som følger:

1. **Fællesmængde:** $A \cup B = \{x \mid x \in A \text{ eller } x \in B\}$.
2. **Sammenkædning:** $A \circ B = \{xy \mid x \in A \text{ og } y \in B\}$.
3. **Stjerne:** $A^* = \{x_1x_2 \dots x_k \mid k \geq 0 \text{ og hvert } x_i \in A\}$.

Vi kigger på et eksempel for at demonstrere brugen af alle operationer.

Eksempel 1.5

Lad alfabetet Σ være det danske alfabet $\{a, b, \dots, \text{\AA}\}$. Hvis $A = \{\text{god, dårlig}\}$ og $B = \{\text{bog, video}\}$, så:

$$A \cup B = \{\text{god, dårlig, bog, video}\}$$

$$A \circ B = \{\text{godvideo, godbog, dårligvideo, dårligbog}\}$$

$$A^* = \{\epsilon, \text{god, dårlig, godgod, dårligdårlig, goddårlig,} \\ \text{dårliggod, godgodgod, godgoddårlig, goddårliggod, \dots}\}$$

Tidligere blev det nævnt at, givet to regulære sprog (eller et, i tilfældet af stjerne-operationen), så vil de producere et nyt, regulært sprog.

Dette vil vi gerne for hver af disse operationer, først kigger vi på den eneste vi kan bevise med DFA'er, og efter går vi videre til dem som lettere kan bevises med NFA'er. Vi kalder det lukket, når en operation der tager to argumenter af samme klasse, giver et output i samme klasse. Eksempelvis er multiplikation og addition lukket under positive heltal, men hverken subtraktion eller division er¹.

Teorem 1.6

Klassen af regulære sprog er lukket under fællesmængde operationen.

Vi beviser dette ved at konstruere en DFA der holder styr på alle de mulige states den kan være i givet begge sprog.

Bevis: Lad M_1 genkende A_1 , hvor $M_1 = (Q_1, \Sigma, \delta_1, q_1, F_1)$ og M_2 genkende A_2 , hvor $M_2 = (Q_2, \Sigma, \delta_2, q_2, F_2)$.

Vi konstruerer en ny DFA, M som genkender $A_1 \cup A_2$, hvor $M = (Q, \Sigma, \delta, q_0, F)$, hvor:

1. $Q = \{(r_1, r_2) \mid r_1 \in Q_1 \text{ og } r_2 \in Q_2\}$.

Dette sæt er det **Kartesiske Produkt** af sætterne Q_1 og Q_2 og er skrevet $Q_1 \times Q_2$. Det er altså sættet af alle par af states hvor den første af fra Q_1 og den anden fra Q_2 .

2. Σ alfabetet, er det samme som i både M_1 og M_2 . Hvis det ikke er ville $\Sigma = \Sigma_1 \cup \Sigma_2$.
3. δ er transitionsfunktionen som defineres som følgende. For hvert $(r_1, r_2) \in Q$ og hvert $a \in \Sigma$, lad

$$\delta((r_1, r_2), a) = (\delta_1(r_1, a), \delta_2(r_2, a))$$

¹Eksempelvis giver $1 - 100$ et negativt tal, og $1/2$ et rationelt tal.

Altså er den næste state parret af de næste states for hver af funktionerne.

4. q_0 er parret (q_1, q_2)
5. F er sættet af par hvori hvert medlem enten er en accept state af M_1 eller M_2 :

$$F = \{(r_1, r_2) | r_1 \in F_1 \text{ eller } r_2 \in F_2\}$$

Dette er det samme som $F = (F_1 \times Q_2) \cup (Q_1 \times F_2)$, men **ikke** det samme som $F_1 \times F_2$ som ville give snittet (intersection) af to sprog.

□

1.4 Pumpelemmaet og ikke-regulære sprog

Der findes mange sprog der er regulære, men det er også vigtigt at vide om et givet sprog er regulært. Et klassisk eksempel på et sprog der ikke er regulært er sproget der beskriver **Balanced Parentheses Problem**, altså $L = \{w | w \text{ has balanced parentheses}\}$. Dette gælder strenge såsom $(((((()))))((()))$ og $(((((((((((((\dots)))))))))))))$. Et hovedteorem i at finde ud af om et sprog, som for eksempel sproget beskrevet før, er regulært, er *pumpe lemmaet*. Ifølge pumpe lemmaet gælder det at alle regulære sprog har en specifik egenskab, og dermed, hvis et sprog ikke har denne egenskab er det ikke et regulært sprog.

Teorem 1.7 (Pumpelemmaet)

Lad A være et regulært sprog, og p (pumpelængden) være et tal, hvor hvis s er en streng i A af længde mindst p , så kan s brydes ned i mindre stykker, $s = xyz$, som opfylder de følgende betingelser:

1. $(\forall i \geq 0)(xy^iz \in A)$
2. $|y| > 0$
3. $|xy| \leq p$

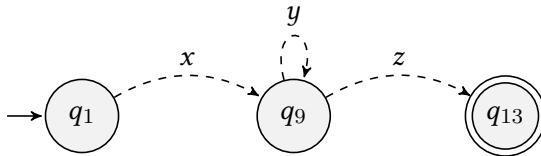
Betingelse 1 siger at en del af strengen, y skal kunne “pumpes” et uendeligt antal gange (hvor $y^0 = \varepsilon$.) Det vil sige at hvis strengen “abc-def” er en del af et regulært sprog, og $y = cde$, så vil “abf” også være en del af sproget, ligesom “abdecdecdecdecdef”. **Betingelse 2** siger at $y \neq \varepsilon$, altså må y ikke være en tom streng. Et enkelt symbol er nok. **Betingelse 3** siger at første del af strengen, uden z -delen, skal være mindre end eller lig med p (lig med hvis $z = \varepsilon$.)

Læg mærke til i betingelserne at det er tilladt for enten x eller z at være lig ε , men ikke y , da denne skal kunne “pumpes”.

Bevis: Lad $M = (Q, \Sigma, \delta, q_1, F)$ være en DFA som genkender sproget A . Lad *pumpelængden* p være antallet af states i M , altså $|Q|$.

Lad $s = s_1 s_2 \cdots s_n$ være en streng i A af længde n , hvor $n \geq p$. Lad r_1, \dots, r_{n+1} være sekvensen af states som M går i når den gennemgår s . Dermed er $r_{i+1} = \delta(r_i, s_i)$ for $1 \leq i \leq n$. Siden sekvensen har længde $n + 1$ må mindst to states være ens.² Vi kalder den første r_j og den anden r_l . Fordi r_l . Da $n + 1 \geq p + 1$, gennemgås r_l indenfor de første $p + 1$ states, i en sekvens der starter ved r_1 , har vi at $l \leq p + 1$. Lad nu $x = s_1 \cdots s_{j-1}$, $y = s_j \cdots s_{l-1}$ og $z = s_l \cdots s_n$. Altså har vi splittet strengen op i sekvenser af states.

Da s_j og s_l er ens, betyder det at y kan gentages et uendeligt antal gange, og opfylder dermed betingelse 1. Siden $j \neq l$ er betingelse 2 også opfyldt, og $l \leq p + 1$ betyder at betingelse 3 også er opfyldt. \square



Figur 1.2: Hvordan en endelig automat kan deles op ud fra beviset.

Figur 1.2 viser en simpel version af hvordan x , y og z defineres grafisk ud fra M . Altså har du første del af strengen, som ikke er den der kan gentages. Denne del, x bliver genkendt indtil vi kommer til staten der genkender starten af y , som kan gentages flere gange. Til sidst kommer vi til de states der genkender z . Læg mærke til at navnene på statesne er arbitrære, og er valgt for at vise et eksempel på den mulige længde mellem states.

²Af dueslagsprincippet

1.5 Opgaver