

Fordítóprogramok

2. előadás

Aszalós László

2015. szeptember 29.

Reguláris kifejezések

Legyen Σ egy ábécé. A Σ feletti reguláris kifejezések halmaza a $(\Sigma \cup \{\lambda, (,), +, *\}^*)$ halmaz legszűkebb olyan U részhalmaza, melyre teljesülnek az alábbi feltételek:

- $\lambda \in U$
- $\Sigma \subset U$
- Ha $R_1, R_2 \in U$, akkor $(R_1) + (R_2)$, $(R_1)(R_2)$ és $(R_1)^*$ is elemei U -nak.

Definíció

Legyen R egy Σ feletti reguláris kifejezés. Az R által meghatározott $|R|$ nyelvet a következőképp definiáljuk:

- $|\lambda| = \emptyset$
- Ha $a \in \Sigma$, akkor $|a| = \{a\}$
- Ha $R = (R_1) + (R_2)$, akkor $|R| = |R_1| \cup |R_2|$
- Ha $R = (R_1)(R_2)$, akkor $|R| = |R_1||R_2|$
- Ha $R = (R_1)^*$, akkor $|R| = |R_1|^*$

Továbbá egy $L \subset \Sigma^*$ nyelvről azt mondjuk, hogy *reguláris nyelv*, ha van olyan Σ feletti R reguláris kifejezés, melyre $L = |R|$.

Tétel

A reguláris nyelvek osztálya zárt a reguláris és a Boole műveletekre nézve.

Tétel

Tetszőleges Σ ábécé esetén a Σ feletti 3 típusú nyelvek és a Σ feletti reguláris nyelvek osztálya megegyezik.

A zárójelek számának csökkentéséért:

Precedencia: Leggyengébb az összeadás, erősebb a szorzás (egymás mellé írás), legerősebb a hatványozás.

Megfelelés $+ \rightarrow |$, $a_1 + \dots + a_n \rightarrow [a_1 \dots a_n]$, $RR^* \rightarrow R+$, $\lambda + R \rightarrow R?$, $R \dots R \rightarrow R\{n\}$, további variánsok: $R\{n, \}$ és $R\{n, m\}$.

Jelölje L a betűk, D a számjegyek reguláris kifejezését.

- egyszerű azonosító: $L(L + D)^*$
- pozitív szám: $(DD^*(\lambda + .)) + (D^*.DD^*)$
- szám tudományos jelölésben: ???

Speciális problémák

- kulcsszavak felismerése `if if then else=then;`
 - ▶ minden kulcsszóhoz reguláris kifejezés
 - ▶ kulcsszavak táblázata
- előreolvasás `++`, `+=`, `--`, `--=`, `!=`
- szimbólumtábla
- hibakezelés

Thompson algoritmus

reguláris kifejezésből determinisztikus automata

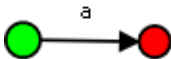
- első lépésben nem-determinisztikus automata készítése
 - ▶ automata csúcsainak száma maximum kétszerese a reguláris kifejezés hosszának
- második lépésben az automata determinizálása
- harmadik lépésben a determinisztikus automata minimalizálása

Felismerő automaták

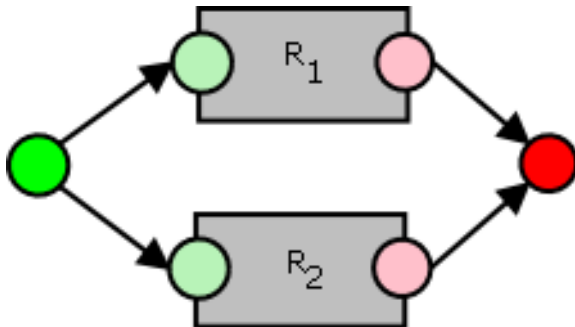
λ



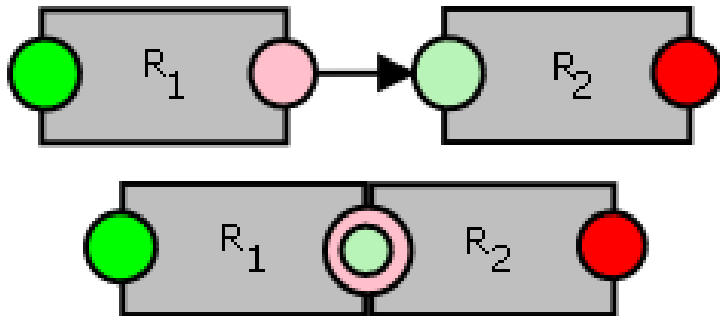
$a \in \Sigma$



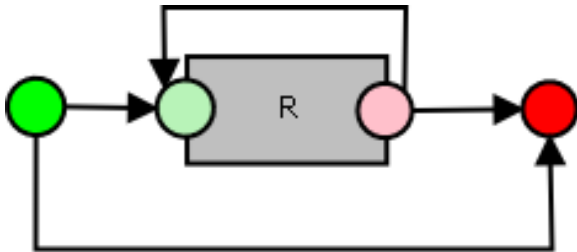
Felismerő automaták $R_1 + R_2$



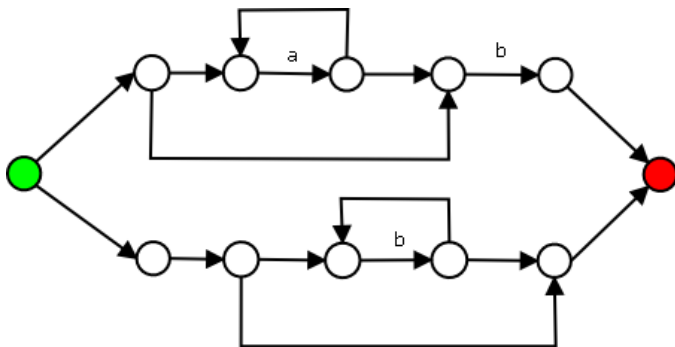
Felismerő automaták $R_1 \cdot R_2$



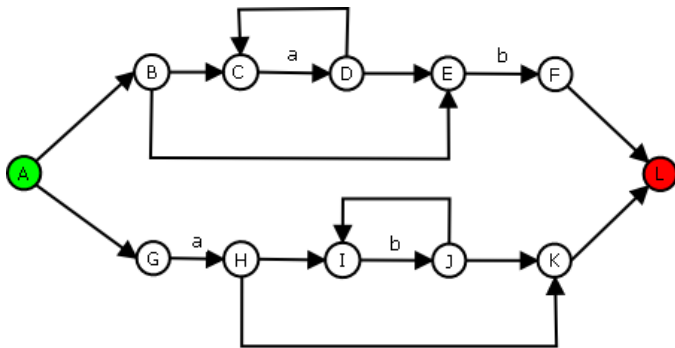
Felismerő automaták R_*



Példa: $a * b + ab *$



Determinisztikus automata készítése



kiinduló állapot: A , végállapot: L

Az automata azon csúcsainak meghatározása, melyekbe az adott pontból csak λ átmenetekkel eljuthatunk.

A lezártja esetünkben $\{A, B, C, E, G\}$. Hagyománytiszteletből és a rövidség okán erre a halmazra $[ABCEG]$ jelöléssel hivatkozunk.

A determinisztikus automatának csak a és b -átmenetei lehetségesek. Az $[ABCEG]$ halmazból a -átmenet C és G csúcsokban történhet, b -átmenet pedig csak E -ben. Az átmenet után a D , a H , illetve az F csúcsokba jutunk. Ezeknek is elkészítjük a lezártját, s ez folytatódik mindaddig, amíg új halmazokat kapunk.

Determinisztikus automata

	<i>a</i>	<i>b</i>
$[\underline{A}BCEG]$	$[C\underline{D}EHIKL]$	$[F\underline{L}]$
$[CDEHIKL]$	$[C\underline{D}E]$	$[FI\underline{J}KL]$
$[F\underline{L}]$	$[]$	$[]$
$[CDE]$	$[C\underline{D}E]$	$[F\underline{L}]$
$[FIJKL]$	$[]$	$[I\underline{J}KL]$
$[]$	$[]$	$[]$
$[IJKL]$	$[]$	$[I\underline{J}KL]$

Amely halmaz tartalmazza a nem-determinisztikus automata kezdőállapotát, az lesz a kezdőállapot, s mindazok, melyek a eredeti automata végállapotát tartalmazzák, lesznek a végállapotok.

Hagyományos jelöléssel

	<i>a</i>	<i>b</i>
\rightarrow 1	2	3
<u>2</u>	4	5
<u>3</u>	6	6
4	4	3
<u>5</u>	6	7
6	6	6
<u>7</u>	6	7

Automata minimalizálása

- 1 Két külön halmazba rakjuk a determinisztikus automata végállapotait és nem végállapotait.
- 2 Ha egy halmazon belül van két elem, mely átmenetei szerint megkülönböztethető, akkor ezt a halmazt daraboljuk, egyébként STOP.
- 3 goto 2

Példa minimalizálásra

$(a + b) * a$ automatája:

	a	b
$\rightarrow 1$	2	3
<u>2</u>	2	3
3	2	3

	a	b
$\rightarrow 1, 3$	2	1, 3
<u>2</u>	2	1, 3

Végállapotok: $\{2\}$, nem végállapotok $\{1, 3\}$. Az 1 és 3 egyetlen átmenettel sem különböztethető meg, így ez a halmaz nem vágható szét.