۳ زبان‌های منظم و گرامرهای منظم

REGULAR LANGUAGES AND REGULAR GRAMMARS

- یک زبان منظم زبانی است که برای آن یک پذیرنده‌ی متناهی (FA) موجود باشد.
- هر زبان منظم می‌تواند توسط یک DFA یا NFA توصیف شود.

۳-۱ عبارت‌های منظم

۳-۱-۱ تعریف رسمی عبارت منظم

تعریف

عبارت منظم فرض می‌کنیم که Σ یک الفبای داده شده باشد:

۱) $\varnothing$، λ و a $(a \in \Sigma)$ عبارت‌های منظم هستند (عبارات منظم ابتدایی).

۲) اگر r_1 و r_2 دو عبارت منظم باشد، $r_1 + r_2$، $r_1.r_2$، r_1^* و (r_1) هم عبارات منظم هستند.

۳) یک رشته عبارت منظم است، اگر و فقط اگر بتوان آن را از ترکیب عبارات منظم ابتدایی با بکارگیری تعداد متناهی قانون ۲ به دست آورد.

مثال

هر یک از موارد زیر، یک عبارت منظم روی الفبای $\Sigma = \{a, b, c\}$ است:

$$(a + bc)^*, \quad (a + bc)^*(c + \varnothing)$$

۳-۱-۲ زبان متناظر با یک عبارت منظم

اگر r یک عبارت منظم باشد، $L(r)$ زبان مرتبط با r است و مطابق جدول زیر تعریف می‌شود:

$L(\varnothing) = \varnothing$
$L(\lambda) = \{\lambda\}$
$L(a) = \{a\} \quad , a \in \Sigma$
$L(r_1 + r_2) = L(r_1) \cup L(r_2)$
$L(r_1.r_2) = L(r_1).L(r_2)$
$L((r)) = L(r)$
$L(r^*) = (L(r))^*$

◄ **تذکر** عبارت منظم r^+ به صورت $r.r^*$ و عبارت منظم $r^?$ به صورت $r + \lambda$ تعریف می‌شود.

◄ **تذکر** برای اجتناب از پرانتزگذاری‌های زیاد، از تقدم عملگرها استفاده می‌کنیم: بالاترین تقدم مربوط به بستار ستاره‌ای و پس از آن به الحاق و سپس اجتماع می‌باشد.

مثال

زبان متناظر با عبارت منظم $a^*.(a+b)$ عبارت است از:

$$L(a^*.(a+b)) = L(a^*).L(a+b) = (L(a))^*.(L(a) \cup L(b)) = \{a\}^*.\{a,b\}$$

مثال

مجموعه‌ی تمام رشته‌های روی الفبای $\{a,b\}$ که به a یا bb ختم می‌شوند، با عبارت منظم زیر توصیف می‌شود:

$$(a+b)^*(a+bb)$$

مثال

مجموعه‌ی تمام رشته‌های روی الفبای $\{a,b\}$ که با تعداد زوجی a آغاز می‌شوند و پس از آن به تعداد فردی b ختم می‌شوند، با عبارت منظم زیر توصیف می‌شود:

$$(aa)^*(bb)^*b$$

مثال

مجموعه‌ی تمام رشته‌های روی الفبای $\{۰, ۱\}$ که حداقل یک زوج صفر متوالی در آن‌ها وجود دارد، با عبارت منظم زیر توصیف می‌شود:

$$(0+1)^*00(0+1)^*$$

مثال

مجموعه‌ی تمام رشته‌های روی الفبای $\{0,1\}$ که هیچ زوج صفر متوالی در آن‌ها وجود ندارد، با عبارت منظم زیر توصیف می‌شود:

$$(1+01)^*(0+\lambda)$$

۳-۱-۳ تساوی عبارت‌های منظم

تعریف **تساوی دو عبارت منظم** دو عبارت منظم را مساوی گویند اگر و فقط اگر زبان‌های متناظر با آنها برابر باشند.

$$r_1 = r_2 \quad \textbf{iff} \quad L(r_1) = L(r_2)$$

چند رابطه برای عبارت منظم فرض می‌کنیم r و s عبارت‌های منظم روی الفبای Σ باشند.

$(r+s)^* = (r^*+s^*)^*$

$(r+s)^* = (r^*+s)^*$

$(r+s)^* = (r+s^*)^*$

$(r+s)^* = (r^*s^*)^*$

$(r+s)^* = r^*(sr^*)^*$

$(r+s)^* = s^*(rs^*)^*$

$(r+s)^* = (r^*s+rs^*)^*$

$(rs)^*r = r(sr)^*$

$r^*r^* = r^*$

$(r^*)^* = r^*$

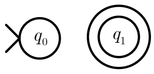
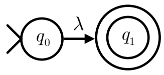
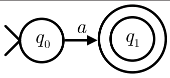
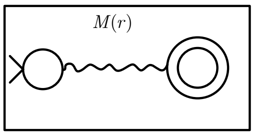
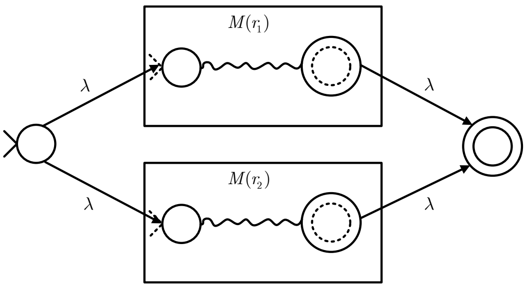
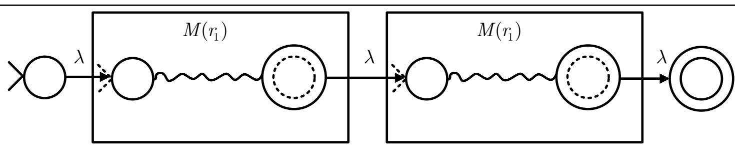
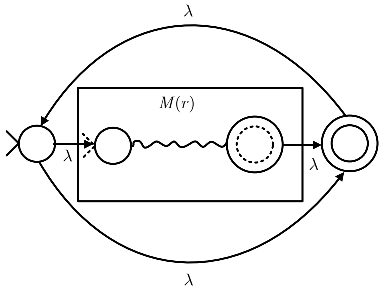
$$L(s) \subseteq L(r) \quad \Rightarrow \quad (r+s)^* = r^*$$

۳-۲ ارتباط میان عبارت‌های منظم و زبان‌های منظم

۳-۲-۱ تعیین آتوماتون متناهی متناظر با یک عبارت منظم

قضیه اگر r یک عبارت منظم باشد، برای آن یک λ-NFA وجود دارد که $L(r)$ را می‌پذیرد.

ساخت NFA با گذر تهی از روی عبارت منظم مطابق جدول زیر عمل می‌کنیم:

Regular expression	Finite automaton
$\varnothing$	
λ	
$a \qquad (a \in \Sigma)$	
r	
$r_1 + r_2$	
$r_1.r_2$	
r^*	

نتیجه اگر r یک عبارت منظم باشد، $L(r)$ منظم است.

۳-۲-۲ تعیین عبارت منظم متناظر با یک آتوماتون متناهی

قضیه

اگر L یک زبان منظم باشد، آن‌گاه عبارت منظم r وجود دارد به طوری که $L = L(r)$.

◄ برای هر زبان منظم یک NFA وجود دارد که آن را می‌پذیرد.
برای هر NFA یک عبارت منظم متناظر با آن وجود دارد.

حل دستگاه معادلات منظم برای یافتن عبارت منظم متناظر با یک آتوماتون متناهی :
برای تعیین عبارت منظم متناظر با یک آتوماتون متناهی

- برای هر حالت یک معادله‌ی منظم می‌نویسیم.
  - اگر $q \notin F$ و گذرهای $q_1 \in \delta(q, a)$ و $q_2 \in \delta(q, b)$ را داشتیم، معادله‌ی منظم

$$q = aq_1 + bq_2$$

را به دستگاه اضافه می‌کنیم.
  - اگر $q \in F$ و گذرهای $q_1 \in \delta(q, a)$ و $q_2 \in \delta(q, b)$ را داشتیم، معادله‌ی منظم

$$q = aq_1 + bq_2 + \lambda$$

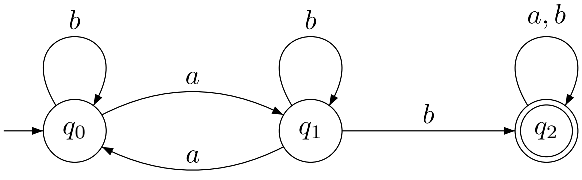
را به دستگاه اضافه می‌کنیم.
- دستگاه حاصل را با شروع از یک معادله و با جایگذاری حل می‌کنیم و q_0 را می‌یابیم.
اگر معادله به صورت بازگشتی بود، آن را به کمک قاعده‌ی Arden حل می‌کنیم.
- عبارت منظم متناظر با آتوماتون متناهی در q_0 قرار دارد.

قاعده‌ی Arden

$$x = ax + b \quad \Rightarrow \quad x = a^*b \quad , \lambda \notin L(a)$$

مثال

عبارت منظم متناظر با آتوماتون متناهی غیرقطعی زیر را محاسبه می‌کنیم:



دستگاه معادلات منظم برای آتوماتون فوق به صورت زیر می‌باشد:

$$\begin{aligned} q_0 &= bq_0 + aq_1 \\ q_1 &= bq_1 + aq_0 + bq_2 \\ q_2 &= (a+b)q_2 + \lambda \end{aligned}$$

که با اعمال قاعده‌ی آردن روی معادله‌ی سوم به دست می‌آوریم $q_2 = (a+b)^*$. با جایگذاری عبارت q_2 در معادله‌ی حالت q_1 داریم:

$$q_1 = bq_1 + aq_0 + b(a+b)^*$$

که با اعمال قاعده‌ی آردن روی آن خواهیم داشت:

$$\begin{aligned} q_1 &= b^*(aq_0 + b(a+b)^*) \\ &= b^*aq_0 + b^*b(a+b)^* \end{aligned}$$

با جایگذاری عبارت q_1 در معادله‌ی حالت q_0 داریم:

$$\begin{aligned} q_0 &= bq_0 + ab^*aq_0 + ab^*b(a+b)^* \\ &= (b + ab^*a)q_0 + ab^*b(a+b)^* \end{aligned}$$

که با اعمال قاعده‌ی آردن، به عبارت منظم زیر برای q_0 و آتوماتون مفروض می‌رسیم:

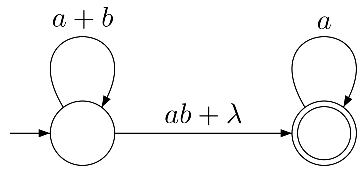
$$q_0 = (b + ab^*a)^*ab^*b(a+b)^*$$

۳-۲-۳ گراف گذر حالت تعمیم‌یافته

گراف گذر حالت تعمیم‌یافته، همانند گراف گذر حالت عادی است، با این تفاوت که یال‌های آن با یک عبارت منظم برچسب‌گذاری می‌شود. از این گراف می‌توان برای یافتن عبارت منظم متناظر با یک آتوماتون متناهی استفاده کرد.

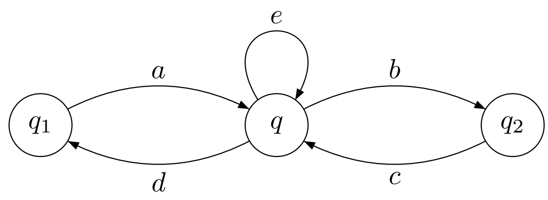
مثال

شکل زیر یک گراف گذر حالت تعمیم‌یافته را نشان می‌دهد:

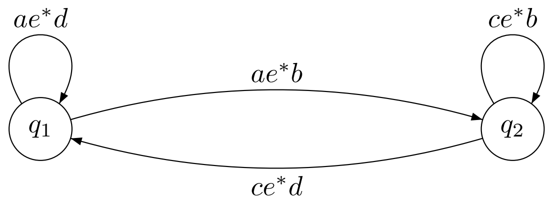


مثال

گراف گذر حالت زیر با سه حالت



را می‌توان به یک گراف گذر حالت تعمیم‌یافته‌ی زیر با دو حالت تبدیل کرد:



۳-۲-۴ کاربرد عبارت‌های منظم در توصیف الگوهای ساده

- تحلیل‌گر لغوی در یک کامپایلر
- تطابق الگو (برنامه‌های sed، grep در UNIX)
- ...

۳-۳ گرامرهای منظم

تعریف **گرامرهای خطی از راست و خطی از چپ** گرامر $G = (V, T, S, P)$ با $A, B \in V$ و $x \in T^*$ را در نظر می‌گیریم.
گرامر خطی از راست *(right linear)* گرامری است که همه‌ی قواعد آن به صورت زیر باشد.

$$A \rightarrow xB \quad \text{یا} \quad A \rightarrow x$$

گرامر خطی از چپ *(left linear)* گرامری است که همه‌ی قواعد آن به صورت زیر باشد.

$$A \rightarrow Bx \quad \text{یا} \quad A \rightarrow x$$

تعریف **گرامر منظم** گرامر منظم، گرامری است که خطی از راست یا خطی از چپ باشد.

◄ **تذکر** در یک گرامر منظم، حداکثر یک متغیر در سمت راست هر قاعده ظاهر می‌شود و بعلاوه این متغیر باید همواره آخرین نماد یا اولین نماد سمت راست باشد.

مثال

گرامر زیر یک گرامر خطی از راست است:

$$S \rightarrow abS \mid a$$

همچنین گرامر زیر یک گرامر خطی از چپ است:

$$\begin{array}{lcl} S & \rightarrow & Aab \\ A & \rightarrow & Aab \mid B \\ B & \rightarrow & a \end{array}$$

و مطابق تعریف هر دوی این گرامرها منظم هستند.

مثال

گرامر زیر یک گرامر منظم نیست:

$$\begin{array}{lcl} S & \to & A \\ A & \to & aB \mid \lambda \\ B & \to & Ab \end{array}$$

زیرا نه خطی از راست و نه خطی از چپ است.

▲

قضیه

اگر $G=(V,T,S,P)$ یک گرامر خطی از راست باشد، آن‌گاه $L(G)$ یک زبان منظم است.

قضیه

اگر L یک زبان منظم بر روی الفبای Σ باشد، آن‌گاه یک گرامر خطی از راست مانند G وجود دارد که $L=L(G)$.

۳-۳-۱ ساخت گرامر خطی راست از روی آتوماتون متناهی

- برای هر حالت $q_i \in Q$ $(0 \le i \le |Q|-1)$ یک ناپایانه A_i را در نظر می‌گیریم.
- نماد شروع S را متناظر با حالت اولیه q_0 به صورت A_0 در نظر می‌گیریم.
- اگر $q_j \in \delta(q_i, a)$ در تابع گذر وجود داشت $(a \in \Sigma)$ (مصرف a)، آن‌گاه قاعده‌ی $A_i \to aA_j$ را به گرامر می‌افزاییم (تولید a).
- اگر $q_j \in \delta(q_i, \lambda)$ در تابع گذر وجود داشت، آن‌گاه قاعده‌ی $A_i \to A_j$ را به گرامر می‌افزاییم.
- اگر $q_i \in F$، آن‌گاه قاعده‌ی $A_i \to \lambda$ را به گرامر می‌افزاییم.

۳-۳-۲ ساخت آتوماتون متناهی از روی گرامر خطی از راست

- هر قاعده به صورت $A \to a_1a_2\ldots a_k$ با $k \ge 2$ $(A \in V, a_1, a_2, \ldots, a_k \in T)$ را به $k+1$ قاعده به شکل

$$A \to a_1B_1,\ B_1 \to a_2B_2,\ B_2 \to a_3B_3,\ \ldots,\ B_{k-1} \to a_kB_k,\ B_k \to \lambda$$

تبدیل می‌کنیم که در آن B_1، B_2، $\ldots$، B_k ناپایانه‌های جدید هستند.

- هر قاعده به صورت $A \to a_1a_2\ldots a_kB$ با $k \ge 2$ $(A, B \in V, a_1, a_2, \ldots, a_k \in T)$ را به k قاعده به شکل

$$A \to a_1B_1,\ B_1 \to a_2B_2,\ B_2 \to a_3B_3,\ \ldots,\ B_{k-1} \to a_kB,$$

تبدیل می‌کنیم که در آن B_1، B_2، $\ldots$، B_{k-1} ناپایانه‌های جدید هستند.

- هر قاعده به صورت $A \to a$ $(A \in V, a \in T)$ را به دو قاعده به شکل

$$A \to aB, \qquad B \to \lambda$$

تبدیل می‌کنیم که در آن B یک ناپایانه‌ی جدید است.

- به ازای هر ناپایانه‌ی A_i موجود درگرامر، یک حالت q_i را در نظر می‌گیریم.
- حالت اولیه q_0 را متناظر با نماد شروع S (A_0) در نظر می‌گیریم.
- اگر قاعده‌ی $A_i \to \lambda$ درگرامر موجود بود، حالت q_i را به مجموعه‌ی حالات نهایی F می‌افزاییم.
- اگر قاعده‌ی $A_i \to aA_j$ درگرامر موجود بود، گذر $q_j \in \delta(q_i, a)$ را به تابع δ می‌افزاییم.
- اگر قاعده‌ی $A_i \to A_j$ درگرامر موجود بود، گذر $q_j \in \delta(q_i, \lambda)$ را به تابع δ می‌افزاییم.

۳-۳-۳ گرامر خطی چپ و زبان‌های منظم

قضیه اگر $G = (V, T, S, P)$ یک گرامر خطی از چپ باشد، آن‌گاه $L(G)$ یک زبان منظم است.

قضیه اگر L یک زبان منظم بر روی الفبای Σ باشد، آن‌گاه یک گرامر خطی از چپ مانند G وجود دارد که $L = L(G)$.

۳-۳-۴ گرامر منظم و زبان منظم

قضیه زبان L منظم است اگر و فقط اگر گرامر منظمی مانند G وجود داشته باشد که $L = L(G)$.

۳-۴ روش‌های مختلف برای توصیف زبان‌های منظم

مشاهده کردیم که روش‌های گوناگونی برای توصیف زبان‌های منظم وجود دارد:

- پذیرنده‌ی متناهی قطعی (DFA)
- پذیرنده‌ی متناهی غیرقطعی (NFA)
- عبارت منظم
- گرامر منظم (گرامر خطی راست و خطی چپ)

همگی این روش‌ها قدرت یکسانی دارند و تعریف‌های کامل و خالی از ابهامی از زبان‌های منظم ارائه می‌دهند.