

Teoria dos Autômatos e Linguagens Formais

aula 07: Expressões regulares

1 Introdução

→ *Tudo o que os autômatos fazem é reconhecer padrões repetitivos*

Será?

Bom para saber se isso é verdade ou não, nós precisamos de uma definição precisa do que é um *padrão repetitivo*.

E para obter essa definição precisa, nós chamamos atenção para um padrão repetitivo que aparece nas nossas descrições de padrões repetitivos.

Quer dizer, quando a gente vai descrever um padrão repetitivo, a gente sempre acaba escrevendo algo como

→ Um bloco da forma tal ou tal
seguido por uma concatenação de blocos da forma tal ou tal
seguido (opcionalmente) por isso ou aquilo

É sempre assim.

Daí que, a gente pode dizer que um padrão repetitivo é qualquer coisa que a gente pode descrever dessa maneira.

E para que essa definição ganhe precisão matemática, basta dizer de maneira matematicamente precisa que maneira é essa.

É isso o que nós vamos ver hoje.

2 Expressões regulares

De fato, a coisa é realmente simples

- A operação de concatenação será denotada pelo símbolo $*$

Quer dizer, quando a gente escreve algo como

$$(abb)^*$$

a gente quer dizer: “uma concatenação de blocos abb ”

- O termo ou será denotada pelo símbolo $\cup$

Quer dizer, quando a gente escreve algo como

$$b \cup ab$$

a gente quer dizer: “um b ou um bloco ab ”.

- E a ideia de seguido por (ou precedido por) será representada escrevendo simplesmente uma coisa depois da outra.

Quer dizer, quando a gente escreve algo como

$$(abb)^* (b \cup ab)$$

a gente quer dizer: “uma concatenação de blocos abb seguida por um a ou ab ”.

Pronto.

É só isso.

As expressões produzidas com essa notação são conhecidas como *expressões regulares*.

E agora nós vamos ver que é possível descrever uma grande coleção de propriedades utilizando essas expressões.

Exemplos

- a.* Uma palavra qualquer formada por a 's e b 's

$$(a \cup b)^*$$

Faz sentido, não é?

Quer dizer, qualquer palavra pode ser vista como uma concatenação dos “blocos” a ou b .

◇

- b.* Blocos de a 's de tamanho ímpar

$$(aa)^* a \quad \text{ou} \quad a(aa)^*$$

Porque um bloco de a 's de tamanho ímpar é basicamente um bloco de a 's de tamanho par seguido ou precedido por um a .

◇

- c.* Palavras que não possuem o símbolo b

$$(a)^* = (e \cup a \cup aa \cup \dots)$$

Isto é, qualquer concatenação do “bloco” a (inclusive a concatenação vazia).

◇

- d.* Palavras que possuem exatamente um b

$$(a)^* b (a)^*$$

Nesse caso os parênteses não são realmente necessários, e nós podemos escrever a coisa assim

$$a^* b a^*$$

◇

e. Palavras que possuem no máximo uma ocorrência do símbolo b

$$a^* \cup a^*ba^*$$

Quer dizer, ou a palavra não tem b 's, ou então ela tem exatamente um b .

◇

f. Palavras que possuem ao menos uma ocorrência do símbolo b

$$(a \cup b)^*b(a \cup b)^*$$

Quer dizer, um b precedido e sucedido por qualquer coisa.

◇

g. Palavras que possuem exatamente dois b 's

Concatenando duas palavras com exatamente um b , nós obtemos

$$a^*ba^*a^*ba^*$$

Mas, essa expressão pode ser simplificada para

$$a^*ba^*ba^*$$

porque a^*a^* é a mesma coisa que a^* — (*Você consegue ver?*)

◇

h. Palavras que possuem no máximo dois b 's

Aqui, de novo, nós podemos concatenar duas palavras que possuem no máximo um b

$$(a^* \cup a^*ba^*)^*(a^* \cup a^*ba^*)^*$$

Mas, isso aqui também funciona

$$a^*(b \cup e)a^*(b \cup e)a^*$$

(*Você consegue ver?*)

◇

i. Palavras que não possuem dois símbolos consecutivos iguais.

Como essas palavras são basicamente sequências de a 's e b 's alternados, nós poderíamos escrever

$$(ab)^* \cup (ba)^*$$

Mas, note que essa expressão descreve palavras que começam e terminam com símbolos diferentes.

Esse problema pode ser corrigido da seguinte maneira:

$$b(ab)^* \cup (ab)^* \cup a(ba)^* \cup (ba)^*$$

Finalmente, é possível simplificar a expressão para

$$(b \cup e)(ab)^* \cup (a \cup e)(ba)^*$$

◇

j. Palavras que não possuem dois a's consecutivos

$$b^*(abb^*)^*(a \cup e)$$

Quer dizer, sempre que aparece um a logo em seguida vem ao menos um b, a menos que ele seja o último símbolo da palavra.

◇

l. Palavras onde todo bloco de a's de tamanho par é seguido por bb, e todo bloco de a's de tamanho ímpar é seguido por b.

Primeira ideia:

$$\left((aa)^*bb \cup a(aa)^*b \right)^*$$

Defeito: essa expressão inclui a palavra abbb, pois o primeiro bloco $(aa)^*$ pode ser vazio.

Segunda ideia:

$$\left(aa(aa)^*bb \cup a(aa)^*b \right)^*$$

Defeito: agora a expressão só inclui palavras que começam com o símbolo a.

Terceira ideia:

$$b^* \left(aa(aa)^*bb \cup a(aa)^*b \right)^*$$

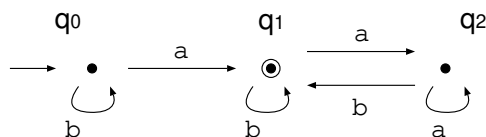
◇

3 Autômatos e expressões regulares

Certo.

Agora que nós temos uma notação adequada, nós podemos utilizá-la para descrever a propriedade reconhecida por um autômato.

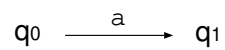
Por exemplo, considere o diagrama abaixo



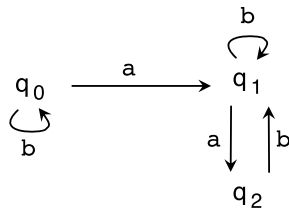
- Que expressão regular descreve as palavras reconhecidas por esse autômato ?

Como usual, o primeiro passo é analisar os caminhos que levam o autômato do estado inicial q_0 até o estado final q_1 .

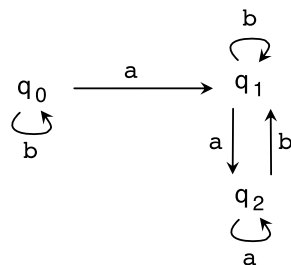
O caminho mais simples desse tipo é:



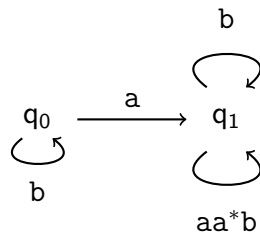
Esse caminho pode ficar mais complicado quando ocorrem laços em torno de q_0 e q_1



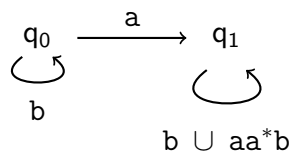
E ele pode ficar mais complicado ainda quando ocorrem laços em torno de q_2



A seguir, nós podemos representar os laços em torno de q_1 assim



e depois assim



Agora é bem fácil chegar na expressão regular

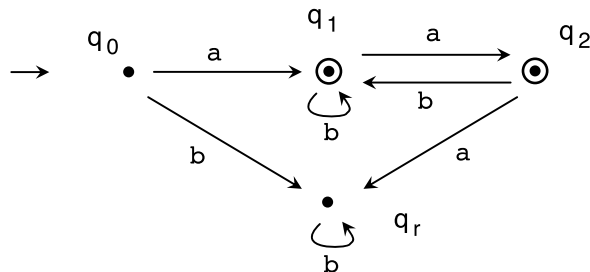
$$b^* a (b \cup aa^*b)^*$$

A seguir nós vamos ver outros exemplos.

Exemplos

a. Dois estados finais

Considere o autômato abaixo



- *Que expressão regular descreve as palavras reconhecidas por esse autômato?*

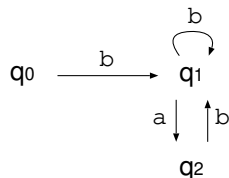
Dessa vez nós precisamos analisar os caminhos que começam no estado inicial q_0 e terminam nos estados finais q_1 e q_2 .

Nós vamos começar com o estado final q_1 .

O caminho mais simples que começa em q_0 e termina em q_1 é

$$q_0 \xrightarrow{b} q_1$$

Esse caminho pode ficar mais complicado quando ocorrem voltas em torno de q_1



E aqui já é fácil escrever a expressão regular que corresponde a esse esquema

$$b(b \cup ab)^*$$

A seguir, nós consideramos o estado final q_2 .

E a observação chave é que qualquer caminho de q_0 até q_2 pode ser decomposto como

- um caminho que vai de q_0 até q_1
- seguido pela transição $q_1 \rightarrow q_2$



Daí que a expressão regular que corresponde a esses caminhos é

$$b(b \cup ab)^* a$$

Portanto, o autômato aceita as palavras descritas pela expressão regular

$$b(b \cup ab)^* a \cup b(b \cup ab)^* a$$

que pode ser simplificada assim

$$b(b \cup ab)^* (e \cup a)$$

◇

b. Raciocinando ao contrário

Agora considere a expressão regular

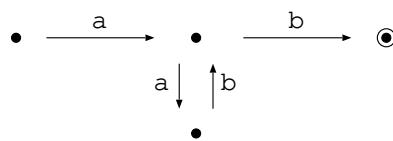
$$a(ab)^*b \cup bb(ab)^*b$$

e a seguinte tarefa

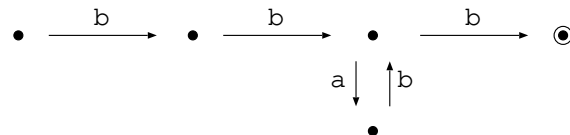
- Construa um autômato que aceita as palavras descritas por essa expressão

Bom, uma ideia é construir caminhos separados para cada parte da expressão regular.

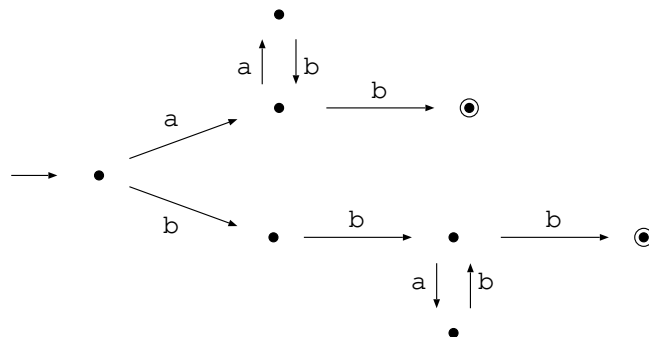
O caminho abaixo corresponde à primeira parte



E o caminho abaixo corresponde à segunda parte

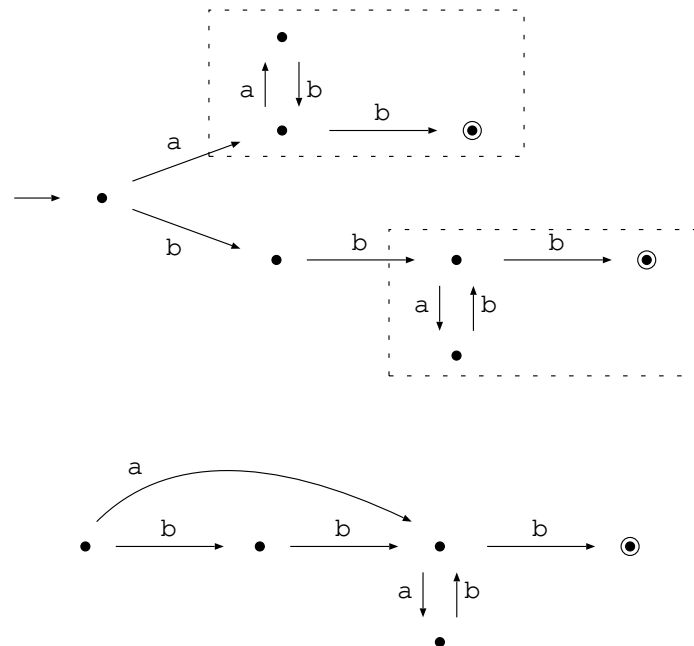


Juntando os dois caminhos em um único esquema, nós obtemos

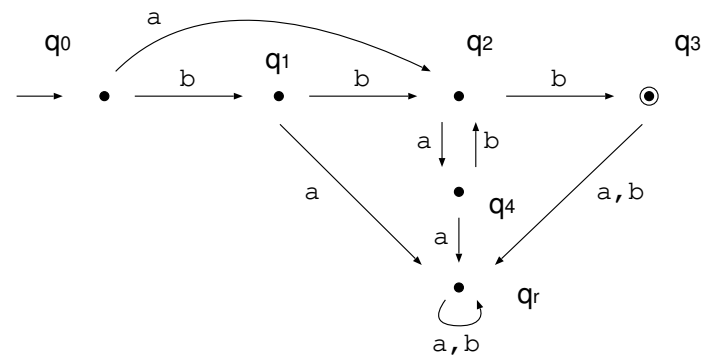


Mas, antes de finalizar a tarefa com a inclusão do estado de rejeição, note que a parte direita dos dois caminhos é idêntica:

Isso significa que esse bloco pode ser reaproveitado, reorganizando o diagrama da seguinte maneira:



E abaixo nós temos o diagrama completo



◇

c. Trabalhando em etapas

Agora considere a tarefa de construir um autômato que reconhece

→ palavras que não possuem a's isolados

A ideia aqui é trabalhar em etapas

1. Primeiro nós capturamos o padrão repetitivo associado a essa propriedade por meio de uma expressão regular
2. E depois nós construímos o autômato com base na expressão regular

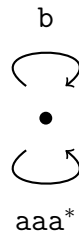
Na primeira etapa, a observação chave é que, se não existem a's isolados então todo bloco de a's tem tamanho no mínimo 2.

E, como não existem restrições sobre os b's, isso corresponde à expressão regular

$$(b \cup aaa^*)^*$$

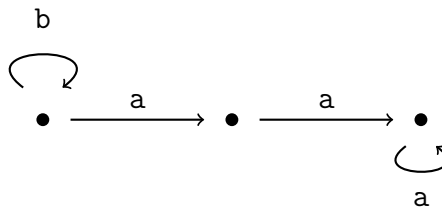
Certo.

Essa expressão regular corresponde a dois laços em torno de um estado

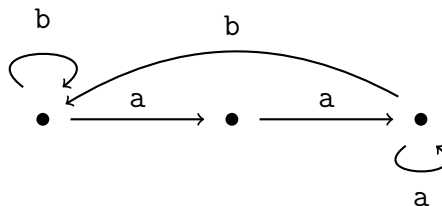


Só que, para expandir o segundo laço em um caminho, nós vamos precisar fazer uma pequena esperteza.

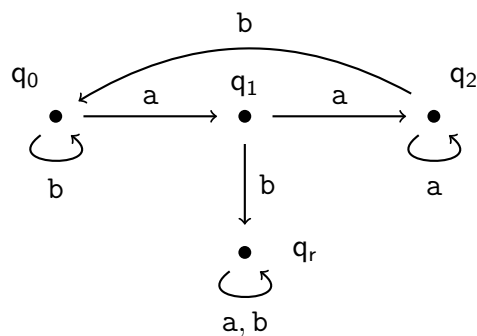
Quer dizer, a ideia é construir o caminho assim



E daí, quando o bloco de a's chega ao fim, com o aparecimento de um b, nós voltamos ao estado inicial



O diagrama completo fica assim



◇

d. Trabalhando em etapas ao contrário

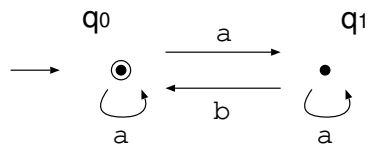
Agora considere a tarefa de construir uma expressão regular que descreve as palavras

→ palavras que possuem um número par de b's

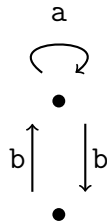
A ideia outra vez é trabalhar em etapas

1. Primeiro nós construímos um autômato que reconhece essa propriedade
2. E depois nós construímos a expressão regular com base nesse autômato

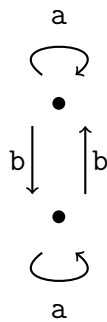
Para a primeira etapa, nós lembramos do autômato que foi construído na aula 01



Daí, para obter a expressão regular, nós analisamos os caminhos que começam e terminam em q_0 . Os caminhos mais simples desse tipo são dois laços



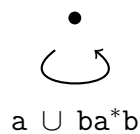
Sendo que o segundo laço pode ficar mais complicado quando ocorrem voltas em torno de q_1



Daí, a gente reorganiza o esquema assim



e depois assim



E aqui é bem fácil ver que a expressão regular é

$$(a \cup ba^*b)^*$$

◇