

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

$$a^*ba^*$$

◇

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:

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

◇