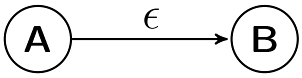
Material de Apoio #0
Algoritmo de Subconjuntos

O algoritmo de subconjuntos é responsável pela conversão de um autômato finito *não-determinístico* para o autômato *determinístico* correspondente. Esse algoritmo é fundamental na geração automática de código de reconhecimento de analisadores léxicos, nas situações onde expressões regulares são definidas pelo projetista para descrever os *tokens* da linguagem a ser implementada no compilador. As expressões regulares são transformadas através da **Construção de Thompson** em autômatos finitos não-determinísticos que são por sua vez convertidos em autômatos finitos determinísticos através do algoritmo de subconjuntos.

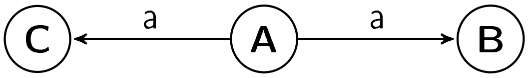
1 Preliminares

Um autômato finito não-determinístico se defina por ter pelo menos uma destas duas características:

- Tem pelo menos uma transição vazia (ϵ) entre dois estados



- Ter mais de uma transição com o mesmo símbolo a partir de um mesmo estado



No primeiro ponto, a transição vazia (ϵ) implica que uma vez no estado de origem, não sabe-se de forma determínistica se devemos não consumir a entrada realizando portanto a transição vazia, ou adotar outra transição no autômato com consumo da entrada. No segundo ponto, não há uma forma direta de decidir se devemos realizar a transição do estado central (A) para (B) consumindo a ou ir para o outro estado (C) consumindo esse mesmo símbolo. Estamos portanto com situações não-determinísticas. Na prática, isso implica que existem múltiplas, e possivelmente infinitas, formas de se reconhecer uma determinada entrada válida de um autômato não-determinista. Este fato complica a implementação desse tipo de autômato em um programa de computador.

2 Visão geral

Considerando um autômato finito *não-determinístico* N e um autômato finito *determinístico* D, a visão geral que inspira o algoritmo de subconjuntos é que cada estado no autômato determinista equivale a um conjunto de estados do autômato não-determinista. A ideia do algoritmo é portanto detectar quais são estes (sub)conjuntos de estados. Essa detecção é feita através de duas operações fundamentais: o fechamento vazio (Fechamento-ϵ) e o movimento.

3 Operações

Existem duas operações fundamentais no algoritmo de subconjuntos: Fechamento-ϵ e Movimento. Os exemplos abaixo são realizados sobre o autômato finito não-determinístico ilustrando no diagrama de transição da Figura 1.

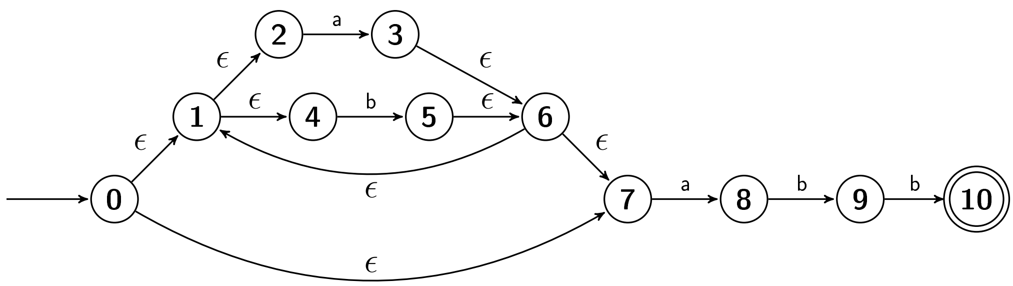


Figura 1: Autômato finito não-determinístico para a expressão $(a|b)^{*}abb$

3.1 Fechamento-ϵ

O Fechamento-ϵ (**e**) é definido como o conjunto de estados alcançados a partir do estado **e** utilizando única e exclusivamente transições vazias e incluem por definição e próprio estado **e**. Considerando o autômato não-determinístico $(a|b)^*abb$ ilustrado na Figura 1, o Fechamento-ϵ (1) é o subconjunto de estados { 1, 2, 4 }, onde o estado 1 é adicionado por definição e os outros dois estados são adicionados ao subconjunto pois é possível atingí-los somente com transições vazias diretas a partir do estado 1. O Fechamento-ϵ (5) é o subconjunto de estados { 5, 6, 7, 1, 2 e 4 }. O estado 5 é adicionado pela definição do fechamento, enquanto que os estados 6, 7, 1, 2 e 4 são atingidos unicamente por transições vazias (observe o autômato da Figura 1).

O Fechamento-ϵ pode também ser calculado a partir de um conjunto de estados. Por exemplo, podemos querer calcular o Fechamento-ϵ ({ 3, 5 }). Neste caso, simplesmente calculamos o Fechamento-ϵ (3) e realizamos a união deste resultado com o obtido do Fechamento-ϵ (5). Sendo assim, temos:

$$\text{Fechamento-}\epsilon\ (\{\ 3,\ 5\ \}) = \text{Fechamento-}\epsilon\ (3) \cup \text{Fechamento-}\epsilon\ (5)$$

O fechamento de um subconjunto de estados é necessário para a operação de Movimento, descrita a seguir.

3.2 Movimento

A operação de Movimento é utilizada para calcular uma transição a partir de um determinado estado e utilizando um determinado símbolo da gramática. Portanto, denota-se Movimento(**e**, **s**). O movimento neste caso, caracteriza-se pelo conjunto de estados alcançados a partir do estado **e** do autômato consumindo unicamente o símbolo **s** através de uma única transição. Sempre, imediatamente após o cálculo do Movimento, realiza-se o Fechamento-ϵ do conjunto resultante. Considerando o autômato da Figura 1, o Movimento(2, a) é { 3 }, mas calculando-se o fechamento do resultado, portanto de Fechamento-ϵ ({ 3 }), obtemos o conjunto resultante do movimento { 3, 6, 1, 2, 4, 7 }, pois a partir do estado 2 atinge-se o estado 3 consumindo apenas o símbolo a; os demais estados são calculados realizando-se o fechamento do estado 3. A operação de movimento é portanto descrita como Fechamento-ϵ (Movimento (**e**, **s**)).

Da mesma forma que no fechamento, a operação de movimento pode ser calculada a partir de um conjunto de estados. Sobre o exemplo da Figura 1, o Fechamento-ϵ (Movimento ({ 2, 7 }, a)) é definido como { 3, 8 } pois há uma transição direta entre os estados 2 e 3 com o símbolo a; e entre 7 e 8 com o mesmo símbolo. Realizando o Fechamento-ϵ deste resultado, obtemos o conjunto final { 3, 8, 6, 1, 2, 4, 7 }.

4 Funcionamento

O algoritmo de subconjuntos inicia-se pelo cálculo do Fechamento-ϵ do estado inicial do autômato finito. não-determinístico. O próximo passo é calcular o fechamento do movimento do conjunto de estados obtidos considerando cada um dos símbolos pertecentes a linguagem. O processo se repete iterativamente até todos os subconjuntos terem sido tratados e quando nenhum novo subconjunto é criado. Dois subconjuntos são diferentes se a quantidade de estados difere ou se há pelo menos um estado presente em um e não no outro. Os diferentes subconjuntos de estados representam os novos estados do autômato finito *determinístico* correspondente, e os movimentos com os diferentes símbolos da linguagem configuram as transições entre estes estados. O estado inicial do novo autômato determinístico é aquele que representa o fechamento-ϵ do estado inicial do autômato não-determinístico. Os estados finais do determinista são todos aqueles que tem em seus respectivos subconjuntos pelo menos um estado final do autômato não-determinista de origem.

5 Exemplo baseado na Figura 1

O autômato da Figura 1 é *não-determinista* pois existem transições com ϵ. Ele tem 11 estados: de 0 a 10; e dois símbolos pertencentes a gramática: **a** e **b**. Vamos convertê-lo para o autômato determinista correspondente. Iniciamos através do cálculo do Fechamento-ϵ do estado 0 (estado inicial):

- Fechamento-ϵ ({ 0 }) = { 0, 1, 2, 4, 7 }

Como o subconjunto é original { 0, 1, 2, 4, 7 }, vamos nomeá-lo subconjunto **A**. O próximo passo é calcular o fechamento do movimento do subconjunto A acima com os dois símbolos da linguagem:

- Fechamento-ϵ (Movimento(A, **a**)) = { 3, 8, 6, 1, 2, 4, 7 }
- Fechamento-ϵ (Movimento(A, **b**)) = { 5, 6, 1, 2, 4, 7 }

Acabamos de criar dois novos subconjuntos originais, pois são diferentes do subconjunto A já existente. Nomeamos o subconjunto { 3, 8, 6, 1, 2, 4, 7 } como subconjunto **B** e o subconjunto { 5, 6, 1, 2, 4, 7 } como **C**. O próximo passo é realizar o fechamento do movimento com os símbolos a partir destes dois novos subconjuntos. Começamos pelo subconjunto B:

- Fechamento-ϵ (Movimento(B, **a**)) = { 3, 8, 6, 1, 2, 4, 7 } (subconjunto já existe, é o próprio **B**)
- Fechamento-ϵ (Movimento(B, **b**)) = { 9, 5, 6, 1, 2, 4, 7 }

O subconjunto { 3, 8, 6, 1, 2, 4, 7 } já existe e é o **B**, mas o subconjunto { 9, 5, 6, 1, 2, 4, 7 } criado da transição de **B** com **b** é um original que nomeamos de subconjunto **D**. Já tratamos neste ponto os movimentos a partir dos subconjuntos A e B, devemos agora tratar os subconjuntos C e D. Começando por C, temos:

- Fechamento-ϵ (Movimento(C, **a**)) = { 8, 3, 6, 1, 2, 4, 7 } (que é o subconjunto B)
- Fechamento-ϵ (Movimento(C, **b**)) = { 5, 6, 1, 2, 4, 7 } (que é o próprio subconjunto C)

Em seguida, tratamos os movimentos a partir do subconjunto D:

- Fechamento-ϵ (Movimento(D, **a**)) = { 3, 8, 6, 1, 2, 4, 7 } (que é o subconjunto B)
- Fechamento-ϵ (Movimento(D, **b**)) = { 10, 5, 6, 1, 2, 4, 7 }

O subconjunto { 10, 5, 6, 1, 2, 4, 7 } configura um novo estado **E** que é final pois ele tem o estado 10 do autômato de origem. Embora já tenhamos detectado que o subconjunto E é final, devemos ainda assim calcular o fechamento do movimento a partir deste novo estado pois sempre há a possibilidade de se criar novos estados. Então temos:

- Fechamento-ϵ (Movimento(E, **a**)) = { 3, 8, 6, 1, 2, 4, 7 } (B)
- Fechamento-ϵ (Movimento(E, **b**)) = { 5, 6, 1, 2, 4, 7 } (C)

A Tabela 1 resume as informações. Podemos representar através de uma diagrama de transições as informações da Tabela 1, criando o autômato determinista correspondete. A Figura 2 ilustra esse novo autômato.

Tabela 1: Resumo das transições da conversão do autômato da Figura 1

		Transições		
Subconjunto	**Identificador**	**a**	**b**	**Comentário**
{ 0, 1, 2, 4, 7 }	A	B	C	Calculado a partir do Fechamento-ϵ (0)
{ 3, 8, 6, 1, 2, 4, 7 }	B	B	D	
{ 5, 6, 1, 2, 4, 7 }	C	B	C	
{ 9, 5, 6, 1, 2, 4, 7 }	D	B	E	
{ 10, 5, 6, 1, 2, 4, 7 }	E	B	C	Estado E é final pois tem 10

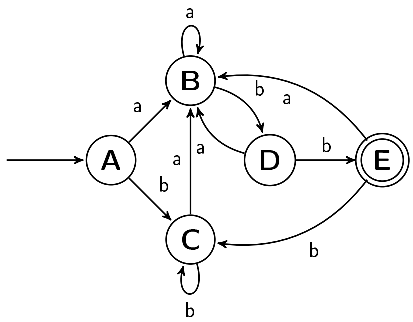


Figura 2: Autômato finito determinístico para a expressão $(a|b)^*abb$