

# Reconhecimento de padrões

---

ALGORITMOS E ESTRUTURAS DE DADOS III

Prof. Marcos André S. Kutova

# Reconhecimento de padrões

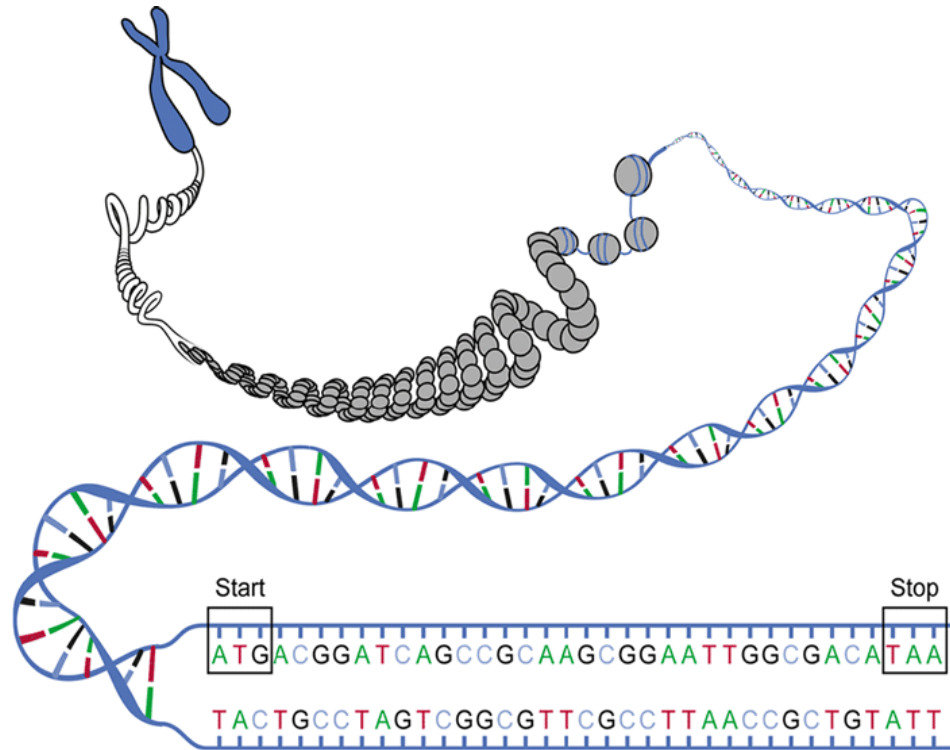
- O reconhecimento de padrões é uma busca por uma cadeia de símbolos em uma sequência maior
  - Ex.: busca de palavras em textos

# Reconhecimento de padrões

- Os algoritmos são identificados como:
  - *Pattern searching*
  - *String searching*
  - *String matching*

# Reconhecimento de padrões

- O reconhecimento de padrões não se limita a busca de texto.
- Exemplos:
  - busca de DNA,
  - detecção de intrusão,
  - correção ortográfica,
  - detecção de plágio,
  - reconhecimento de voz, etc.



# Reconhecimento padrões

- Reconhecimento exato de padrões
  - Força Bruta
  - KMP
  - Aho Corasick
  - Boyer Moore
- Reconhecimento aproximado de padrões
  - Distância de Levenshtein

**Força bruta**

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
T	R	I	S	T	E													

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 1

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
T	R	I	S	T	E													

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 2



# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
T	R	I	S	T	E													

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 3

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
	T	R	I	S	T	E												

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 4

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
		T	R	I	S	T	E											

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 5

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
			T	R	I	S	T	E										

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 6

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
				T	R	I	S	T	E									

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 7

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
					T	R	I	S	T	E								

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 8

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
					T	R	I	S	T	E								

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 9

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
						T	R	I	S	T	E							

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 10



# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
							T	R	I	S	T	E						

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 11

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
								T	R	I	S	T	E					

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 12

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
									T	R	I	S	T	E				

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 13

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
										T	R	I	S	T	E			

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 14

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S

T	R	I	S	T	E
---	---	---	---	---	---

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 15

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
												T	R	I	S	T	E	

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 16

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
												T	R	I	S	T	E	

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 17

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
												T	R	I	S	T	E	

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 18



# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S

T	R	I	S	T	E
---	---	---	---	---	---

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 19

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
												T	R	I	S	T	E	

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 20

# Força bruta

- Busca do padrão "TRISTE" em um texto

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
T	R	E	S		T	I	G	R	E	S		T	R	I	S	T	E	S
												T	R	I	S	T	E	

Tamanho do texto: 19 caracteres

Tamanho do padrão: 6 caracteres

Quantidade de comparações: 21

# Problema do força bruta

	0	1	2	3	4	5
Sequência	R	Á	P	I	D	O

Padrão	R	A	T	O		
		R	A	T	O	
			R	A	T	O

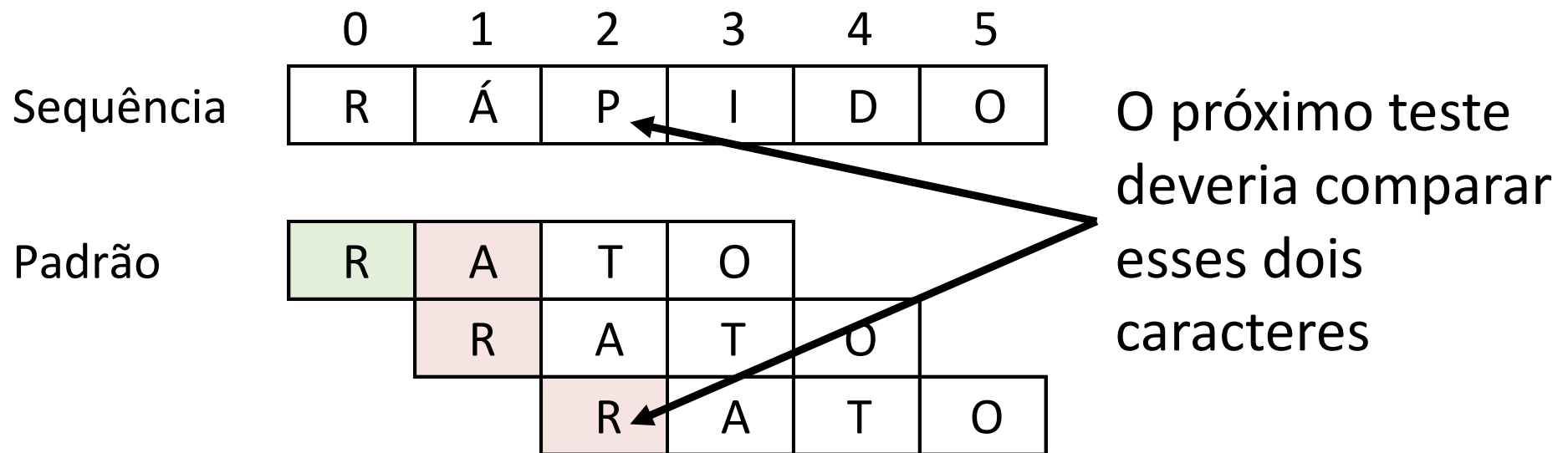
Com o algoritmo da força bruta, o segundo caráter (Á) será testado duas vezes.

# Problema do força bruta

	0	1	2	3	4	5
Sequência	R	Á	P	I	D	O
Padrão	R	A	T	O		
		R	A	T	O	
			R	A	T	O

Mas já sabemos que  
caráter é esse  
e que ele não  
aparece no padrão.

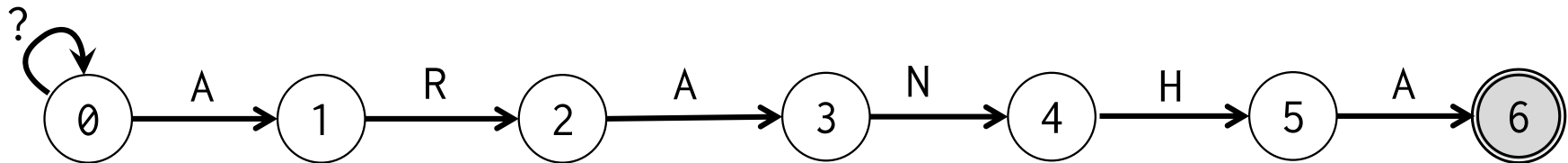
# Problema do força bruta



**KMP**

# KMP

- Criado por Donald Knuth, James Morris e Vaughan Pratt em 1977
- Baseado em uma variação de um diagrama de estados



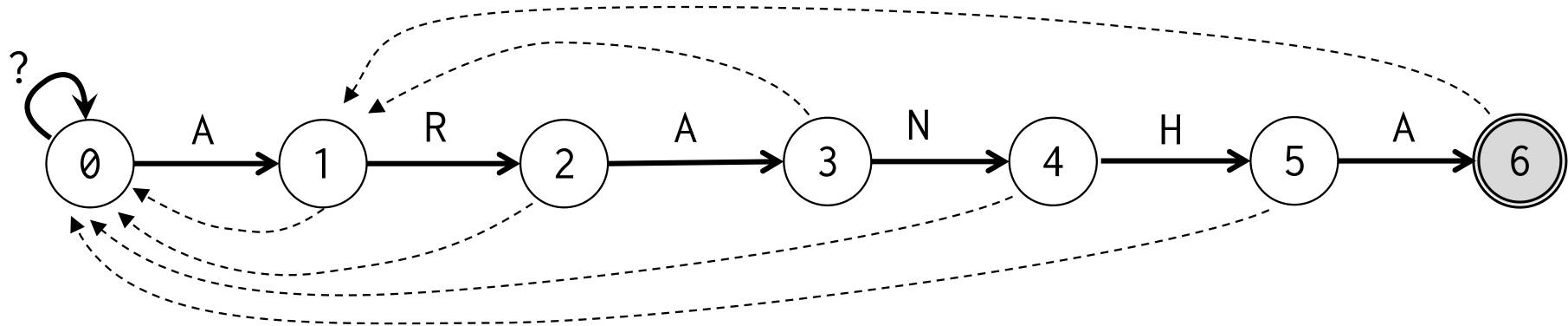
Vetor de transição == padrão  
(de eventos)

0	1	2	3	4	5
A	R	A	N	H	A



# Transições de falhas

- Transição de falha (baseada nas repetições do prefixo)

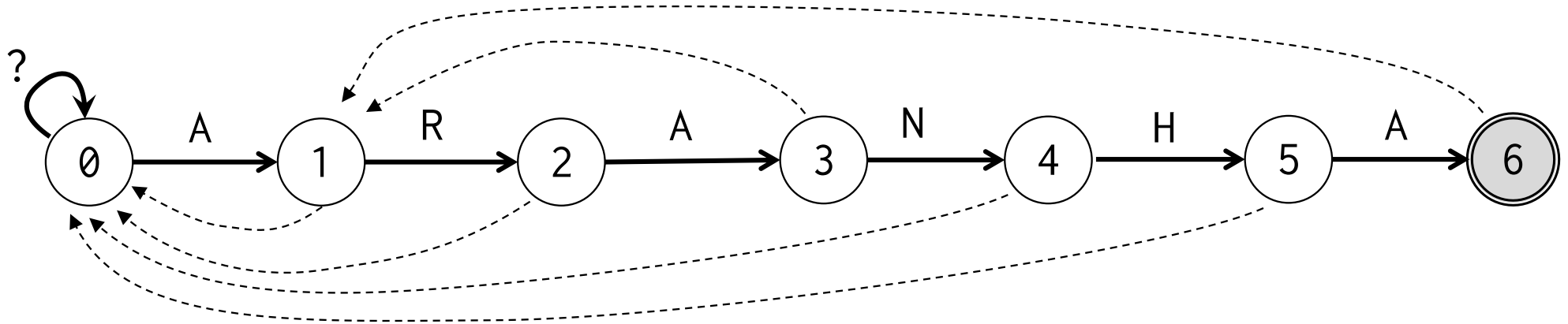


	0	1	2	3	4	5
	A	R	A	N	H	A
Vetor de transição de falha	0	0	1	0	0	1

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



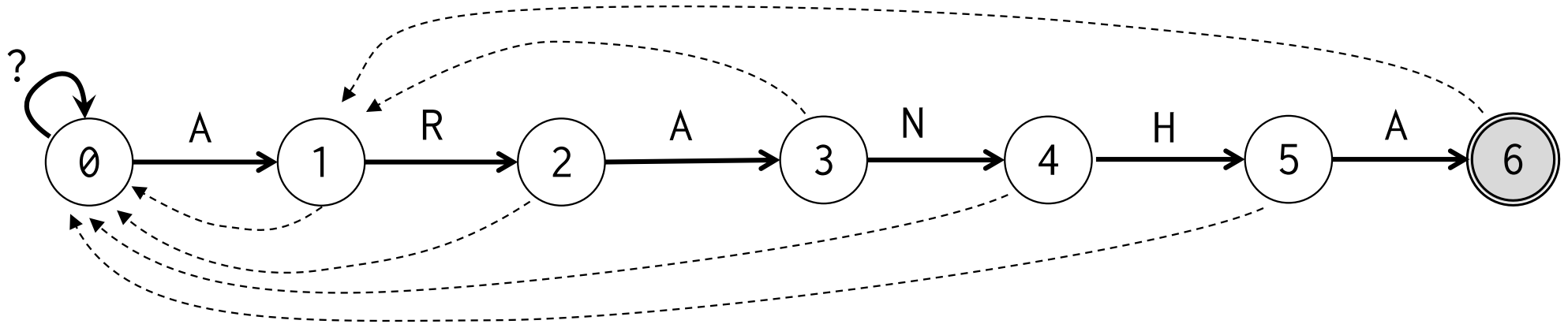
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



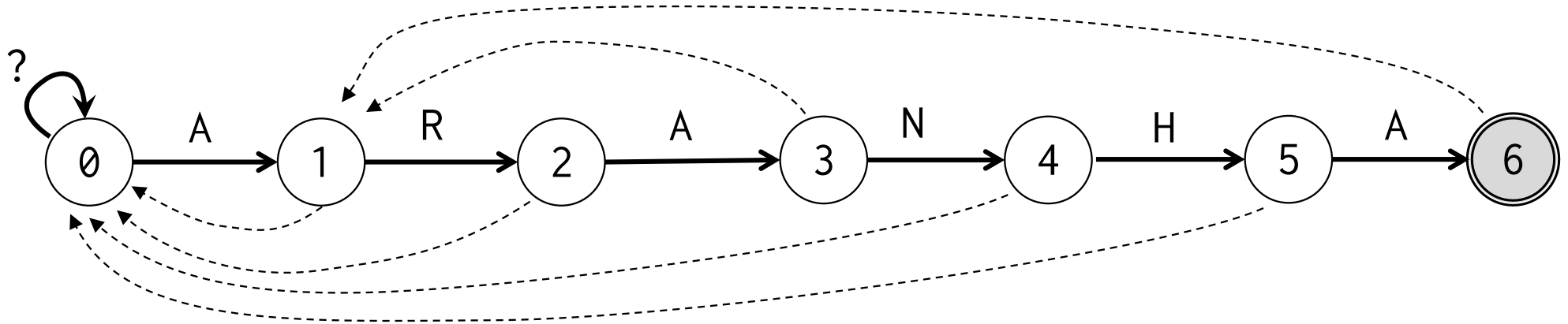
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



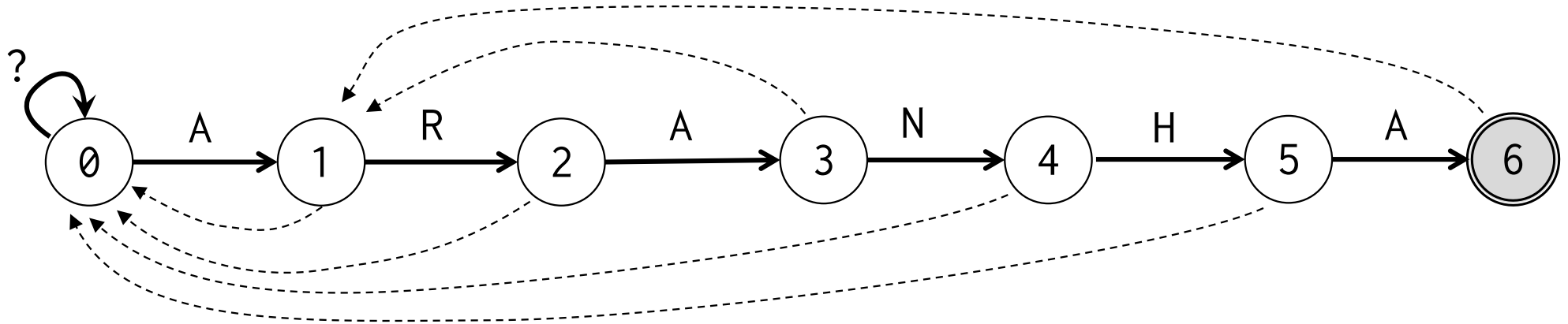
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



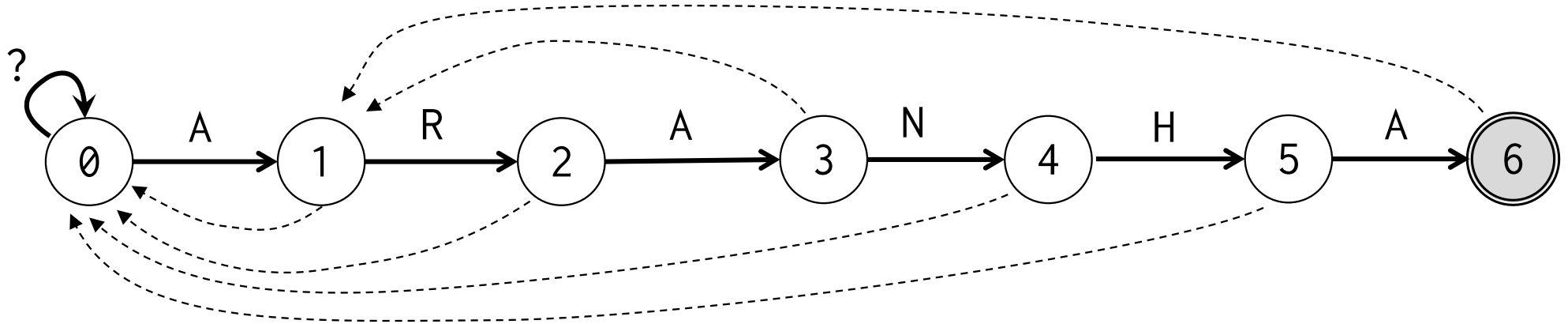
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



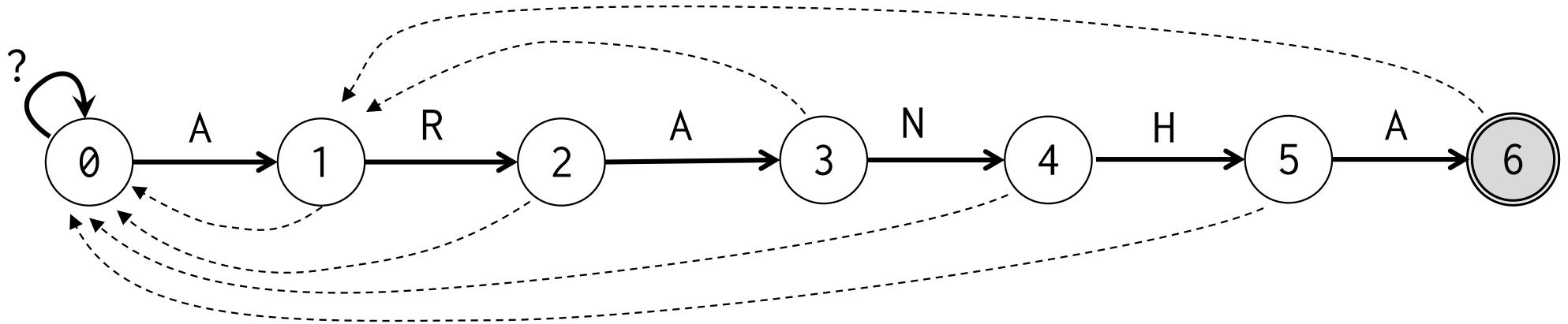
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---

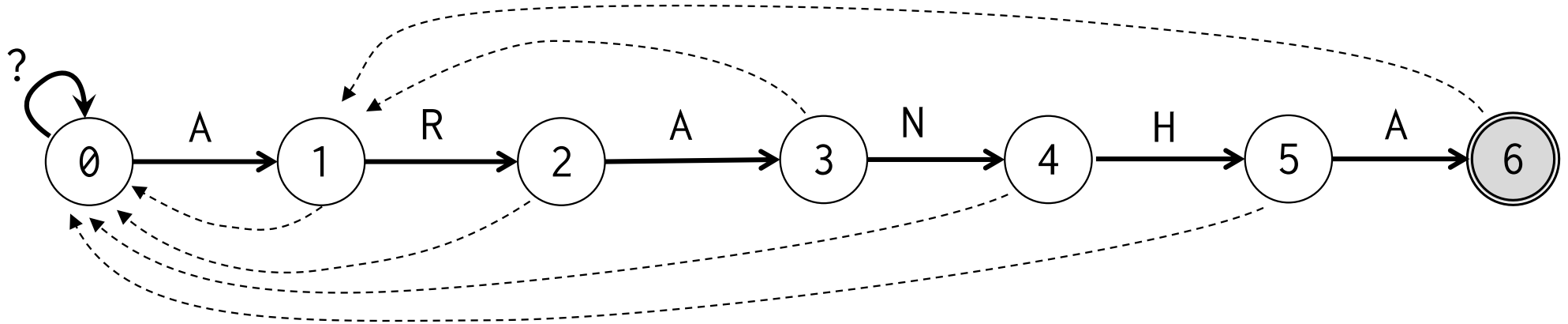


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1



0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

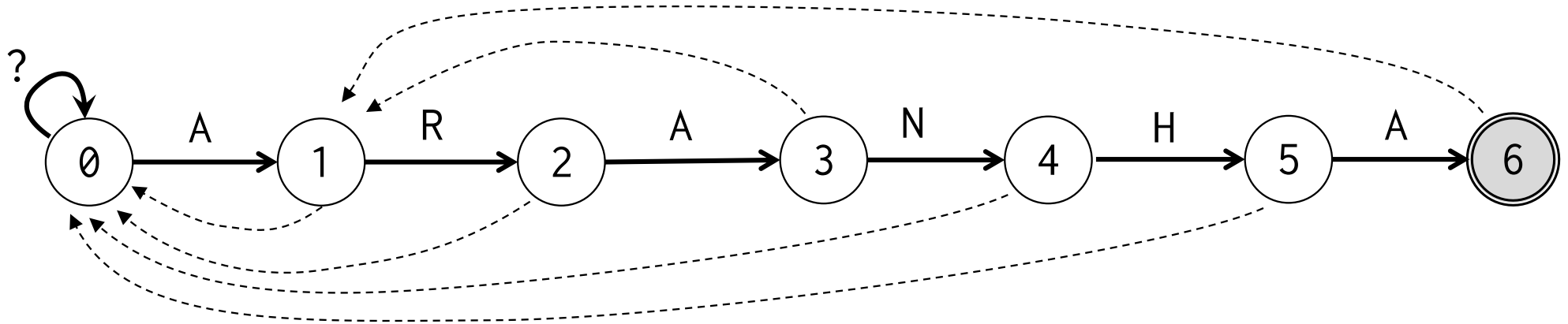
[illegible]



# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---

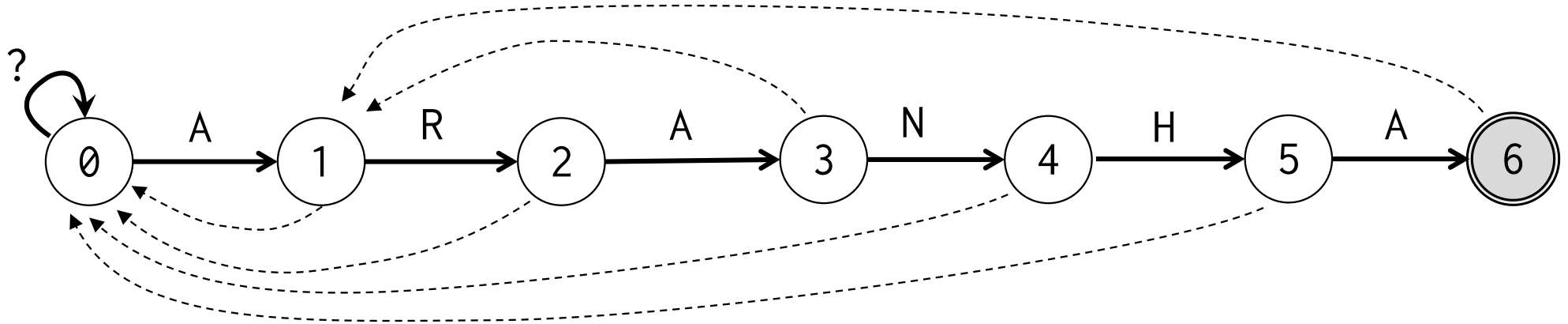


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1

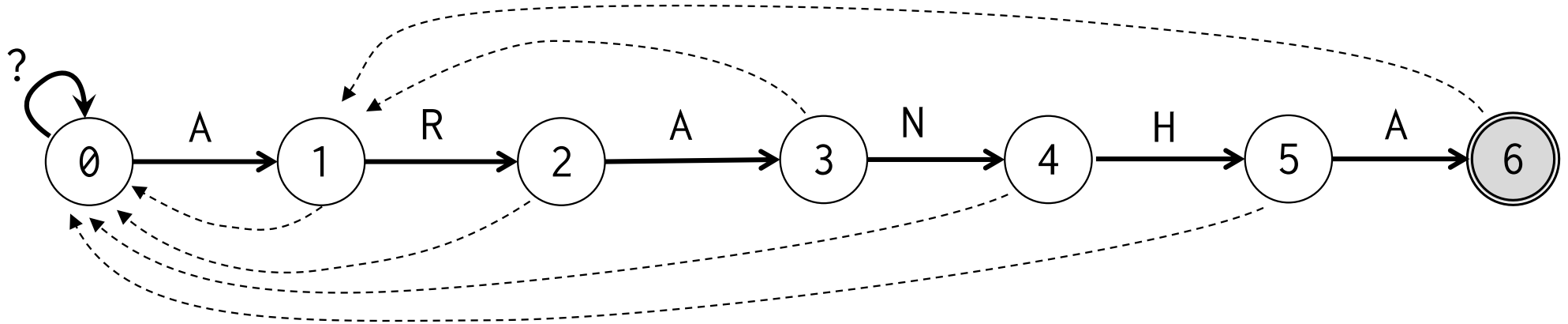


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1



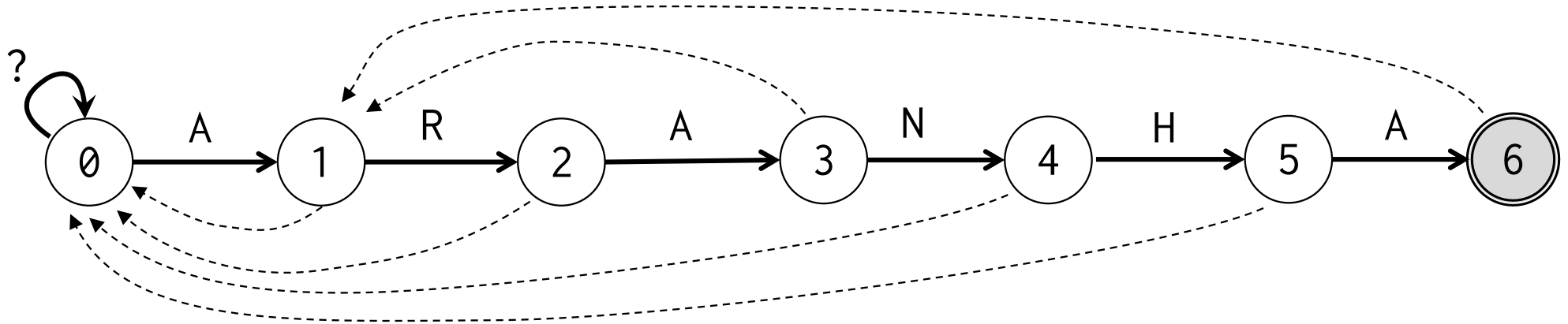
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



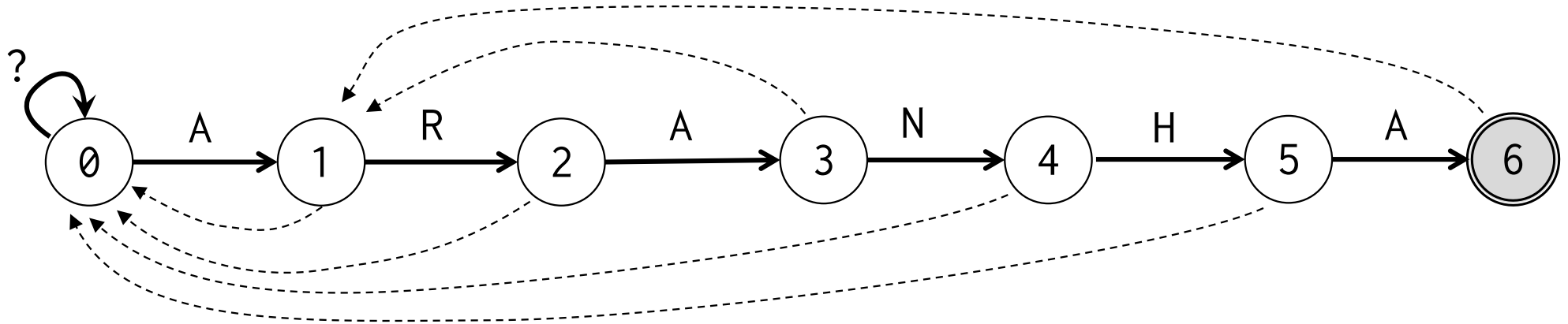
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



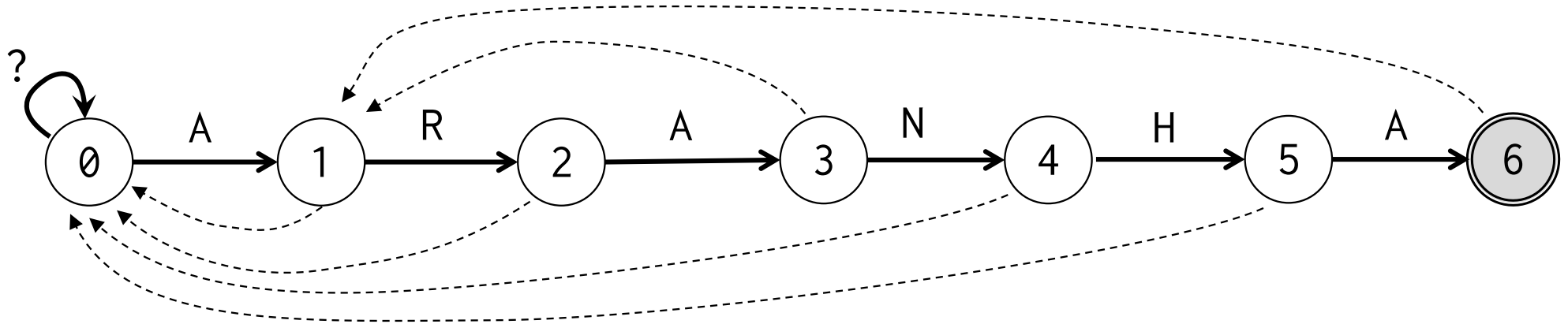
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---



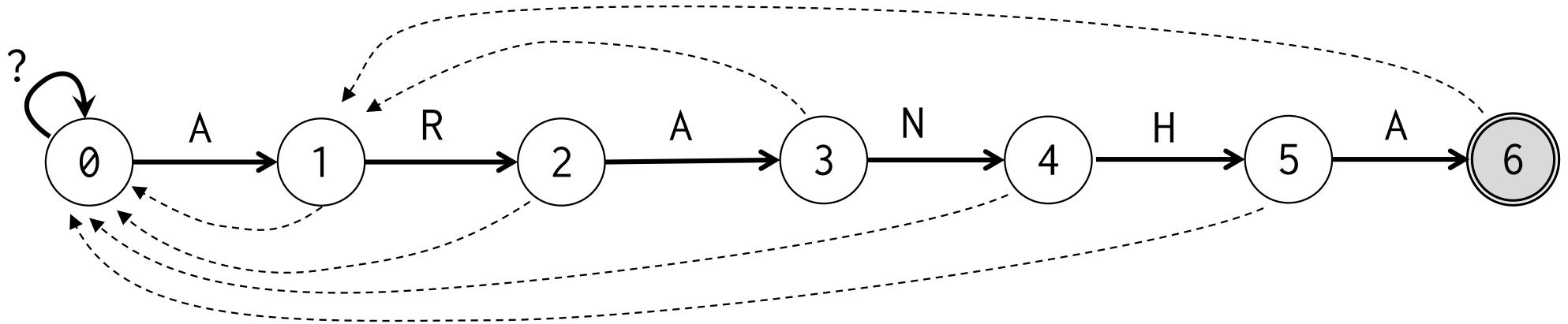
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0 1 2 3 4 5  
A R A N H A

0	0	1	0	0	1
---	---	---	---	---	---

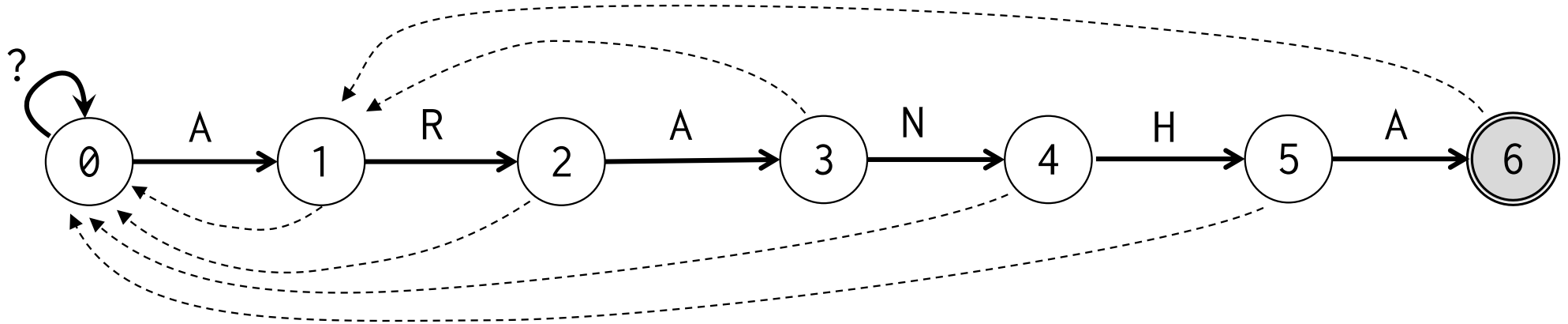


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1



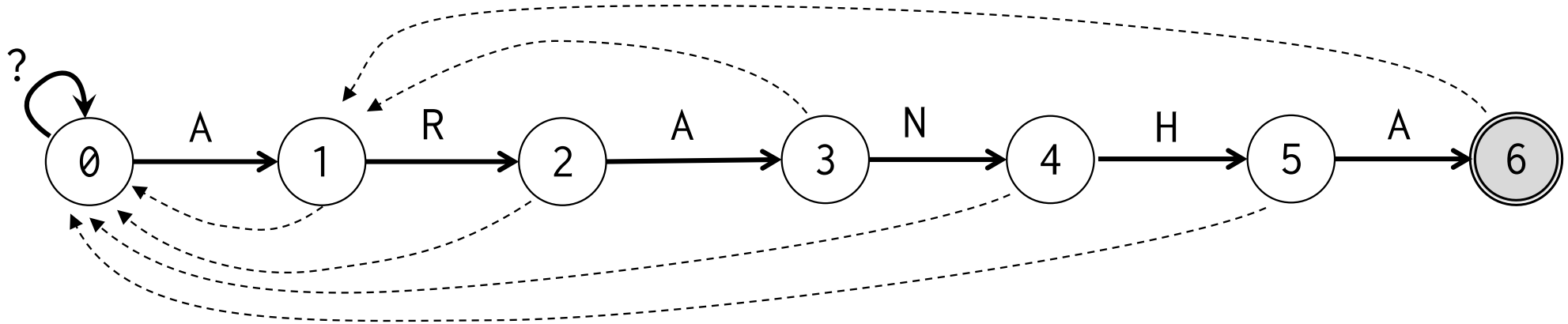
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

[illegible]



# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1

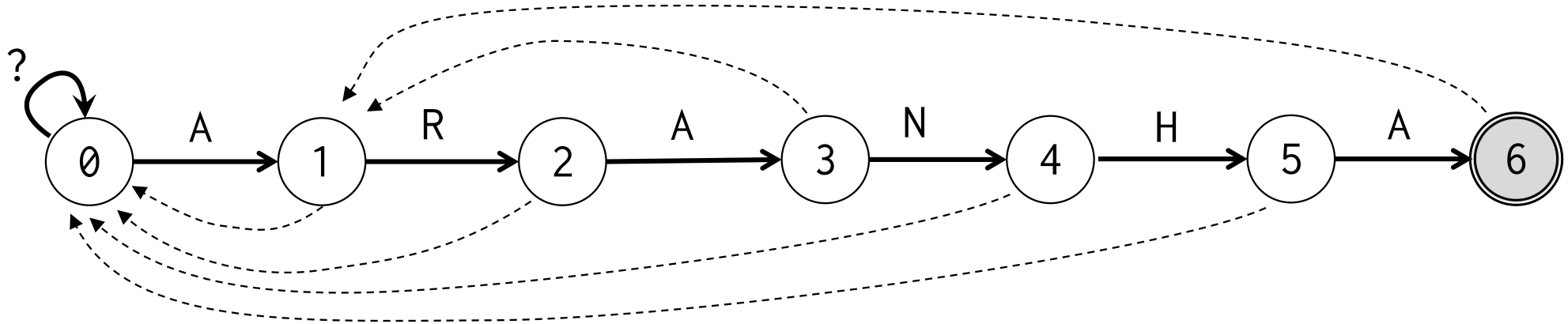


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

Estados	0	1	0	1	2	3	4	5	6	0	1	2	0	1	0	0								
---------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--	--	--

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1

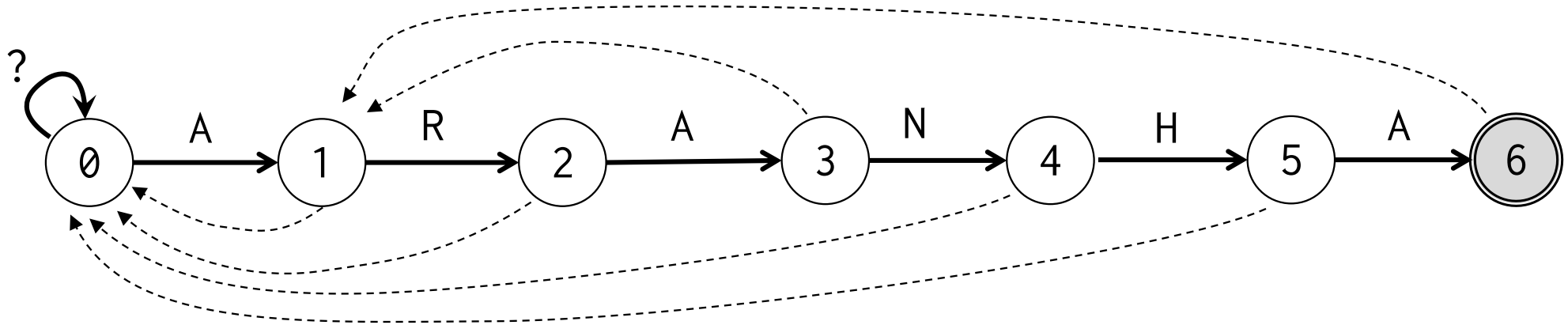


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

Estados	0	1	0	1	2	3	4	5	6	0	1	2	0	1	0	0	1							
---------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--	--

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1

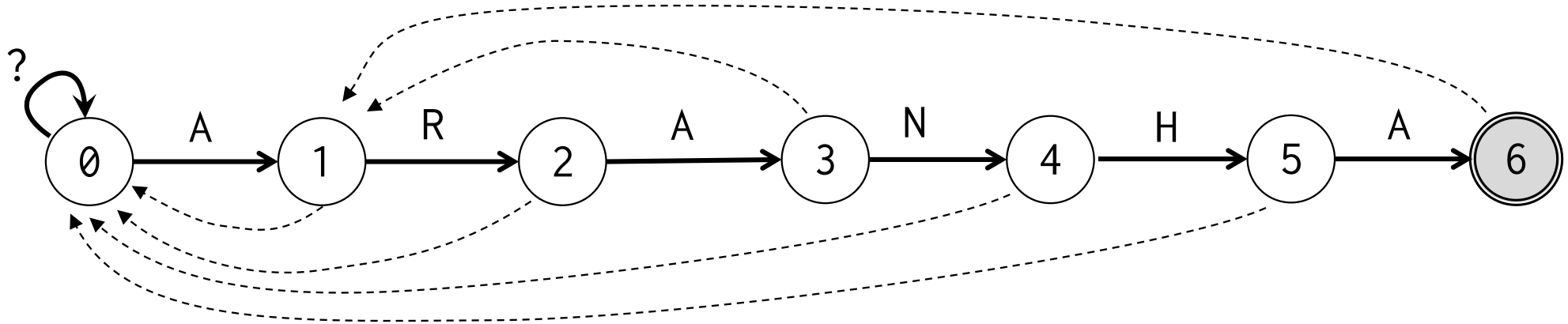


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

Estados	0	1	0	1	2	3	4	5	6	0	1	2	0	1	0	0	1	0						
---------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--	--

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1

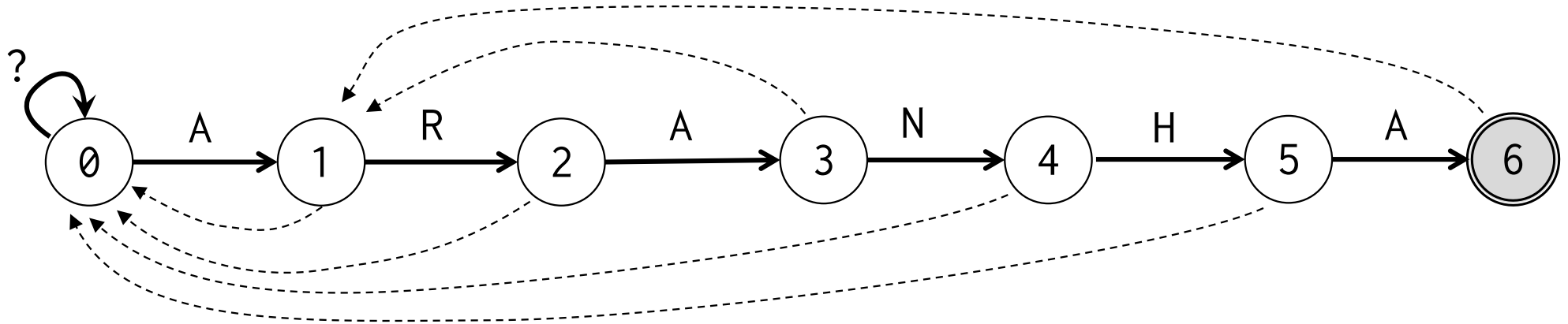


0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

Estados	0	1	0	1	2	3	4	5	6	0	1	2	0	1	0	0	1	0	1					
---------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--

# Exemplo

0	1	2	3	4	5
A	R	A	N	H	A
0	0	1	0	0	1



0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
A		A	R	A	N	H	A		A	R	R	A	N	H	A		A		J	A	R	R	A

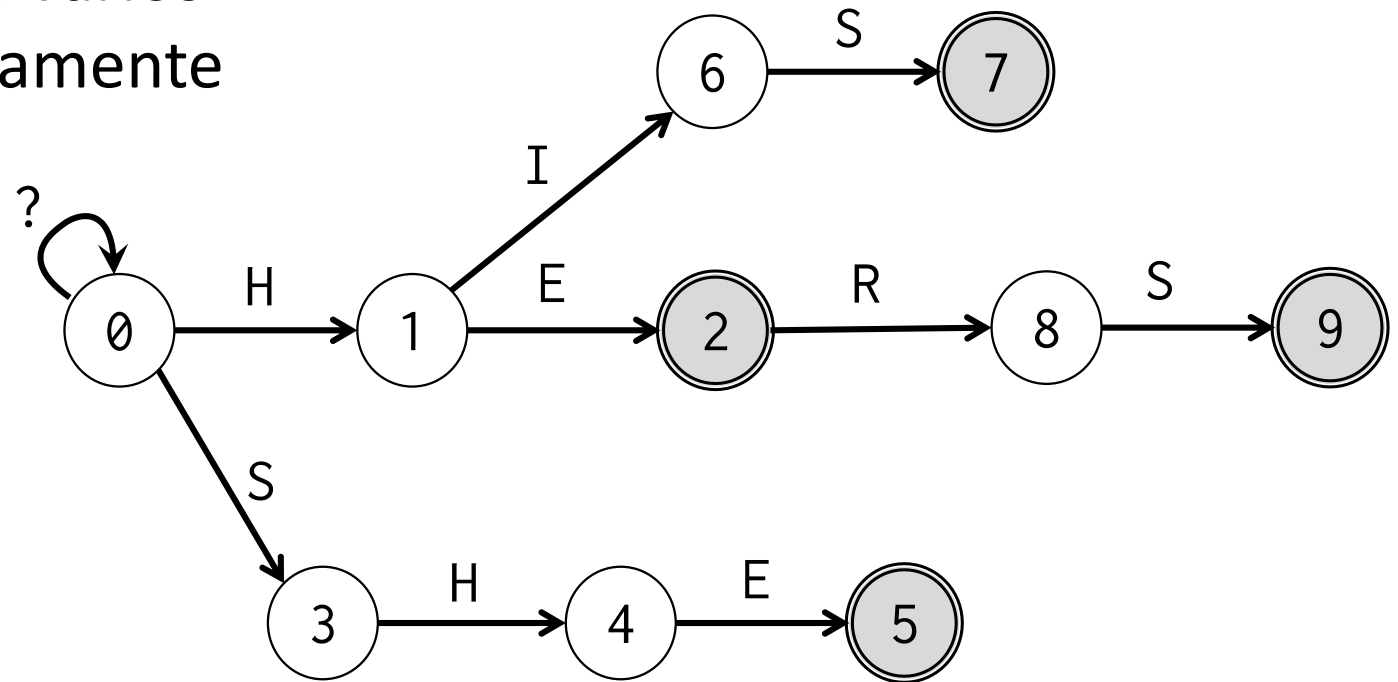
Estados	0	1	0	1	2	3	4	5	6	0	1	2	0	1	0	0	1	0	1	0					
---------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--	--

# Aho-Corasick

# Aho Corasick

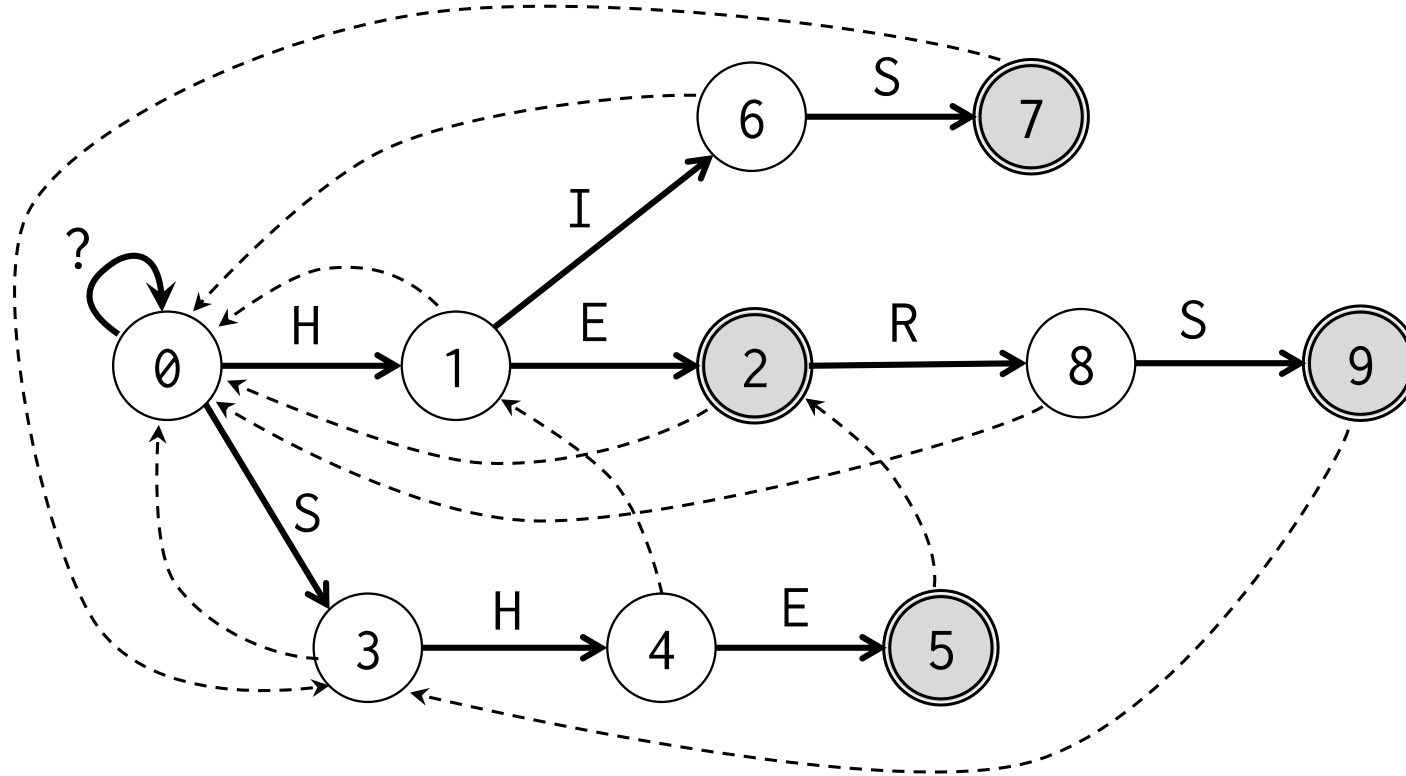
- Criado por Alfred V. Aho e Margaret J. Corasick em 1975
- Permite a busca de vários padrões simultaneamente

Ex.: HE,  
SHE,  
HIS,  
HERS



# Aho Corasick

- Transições de falha continuam baseadas em prefixos

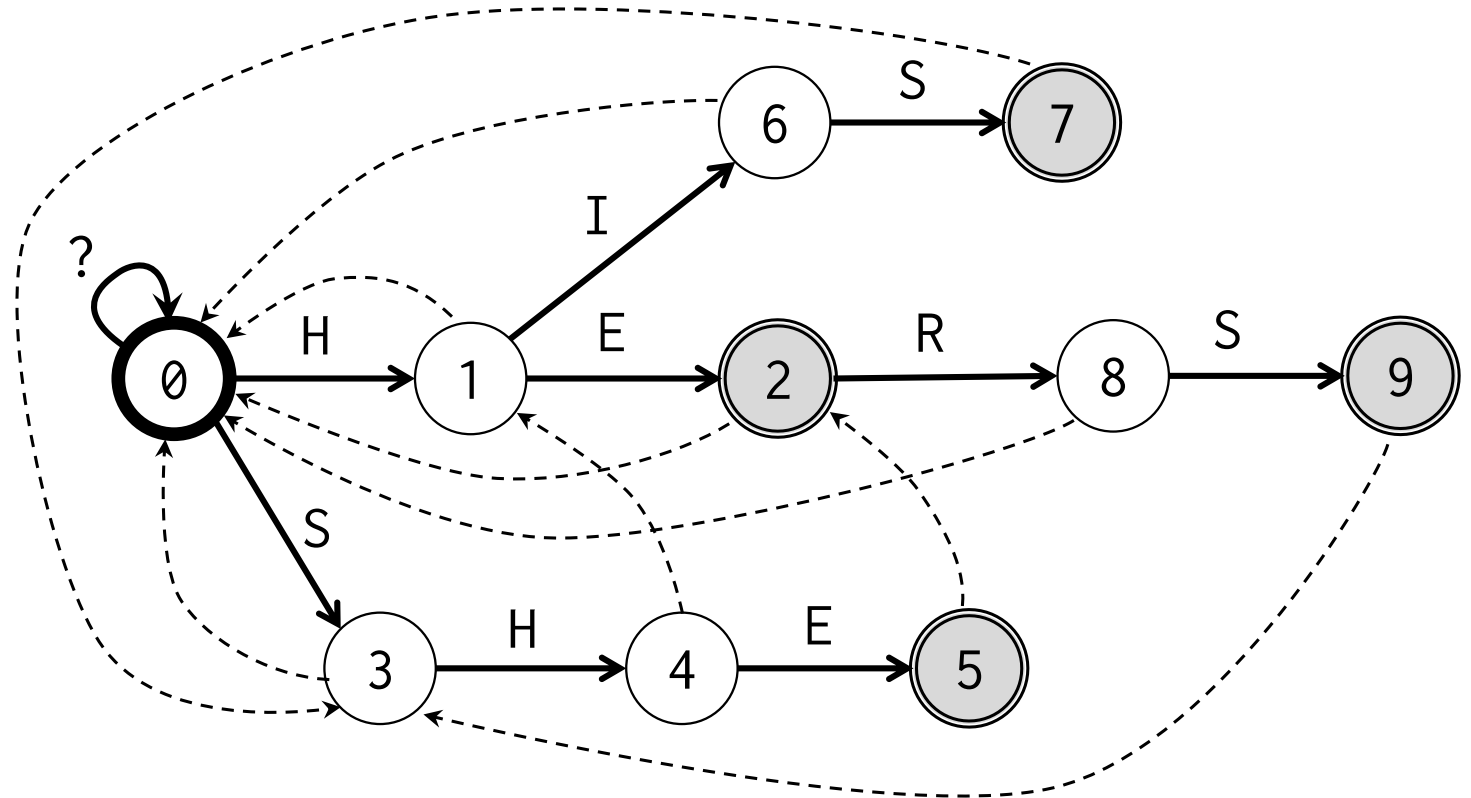




# Aho Corasick

- Reconhecimento de  
HERSHEYS:

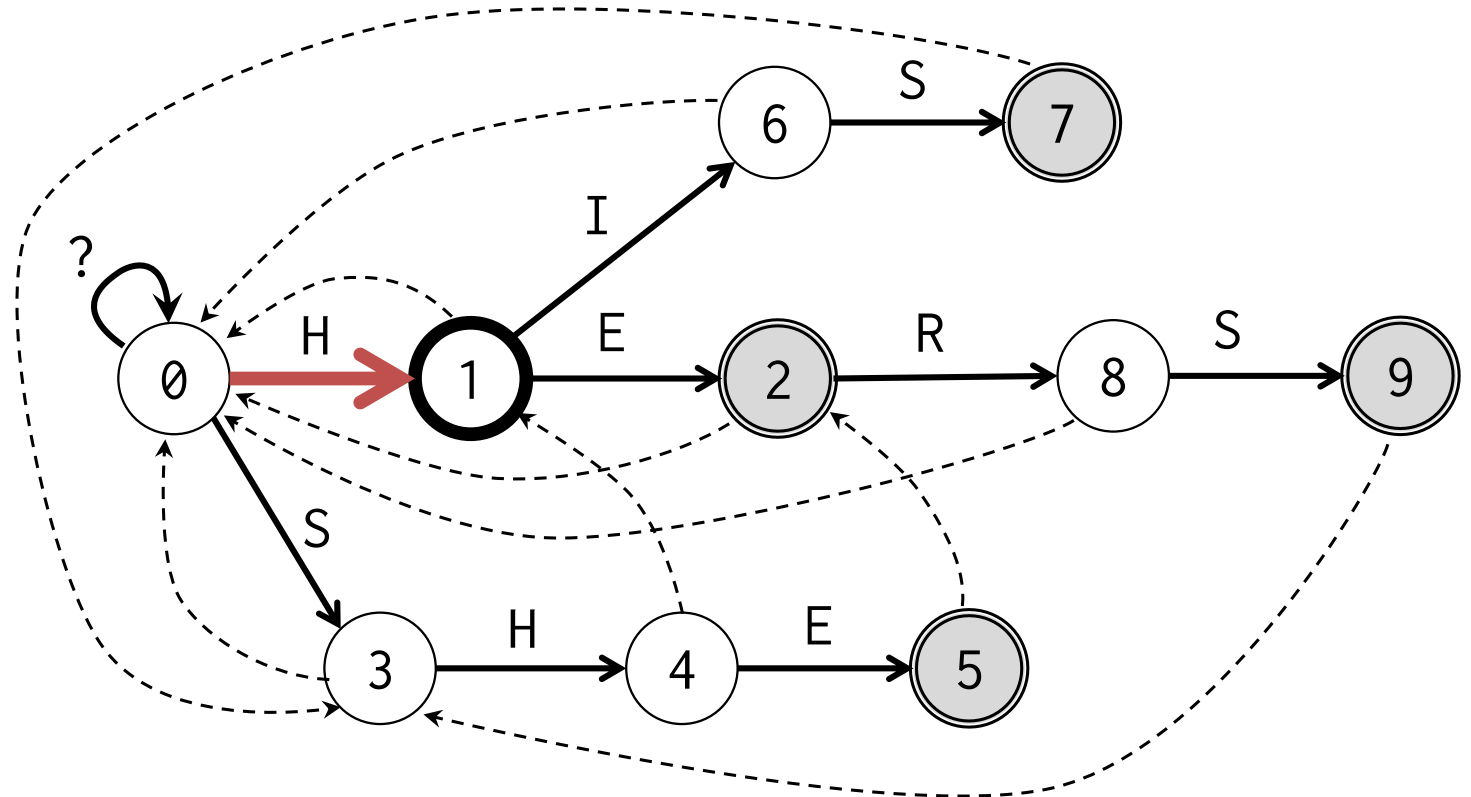
0



# Aho Corasick

- Reconhecimento de  
HERSHEYS:

0  
H: 1



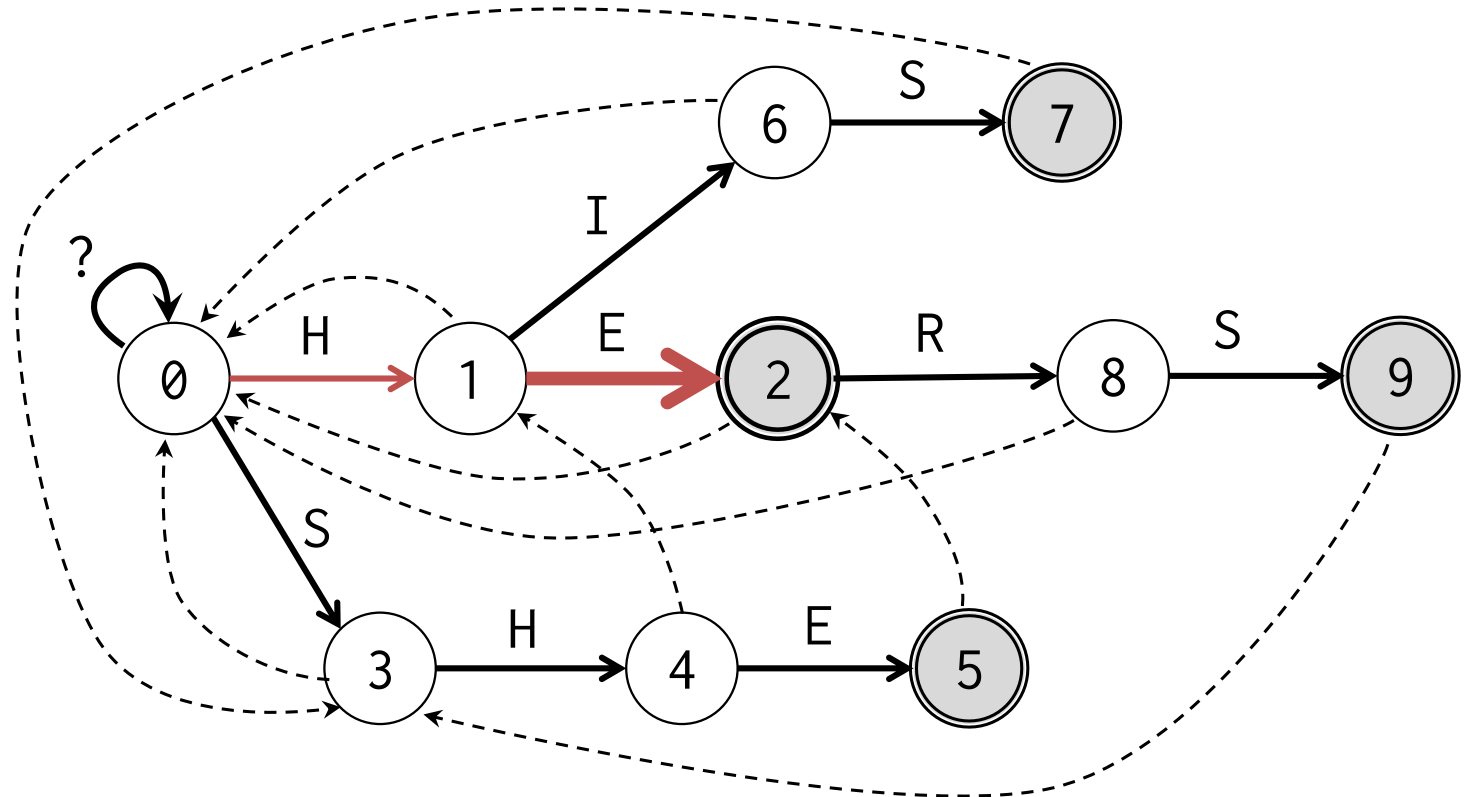
# Aho Corasick

- Reconhecimento de  
HERSHEYS:

0

H: 1

E: 2 {HE}



# Aho Corasick

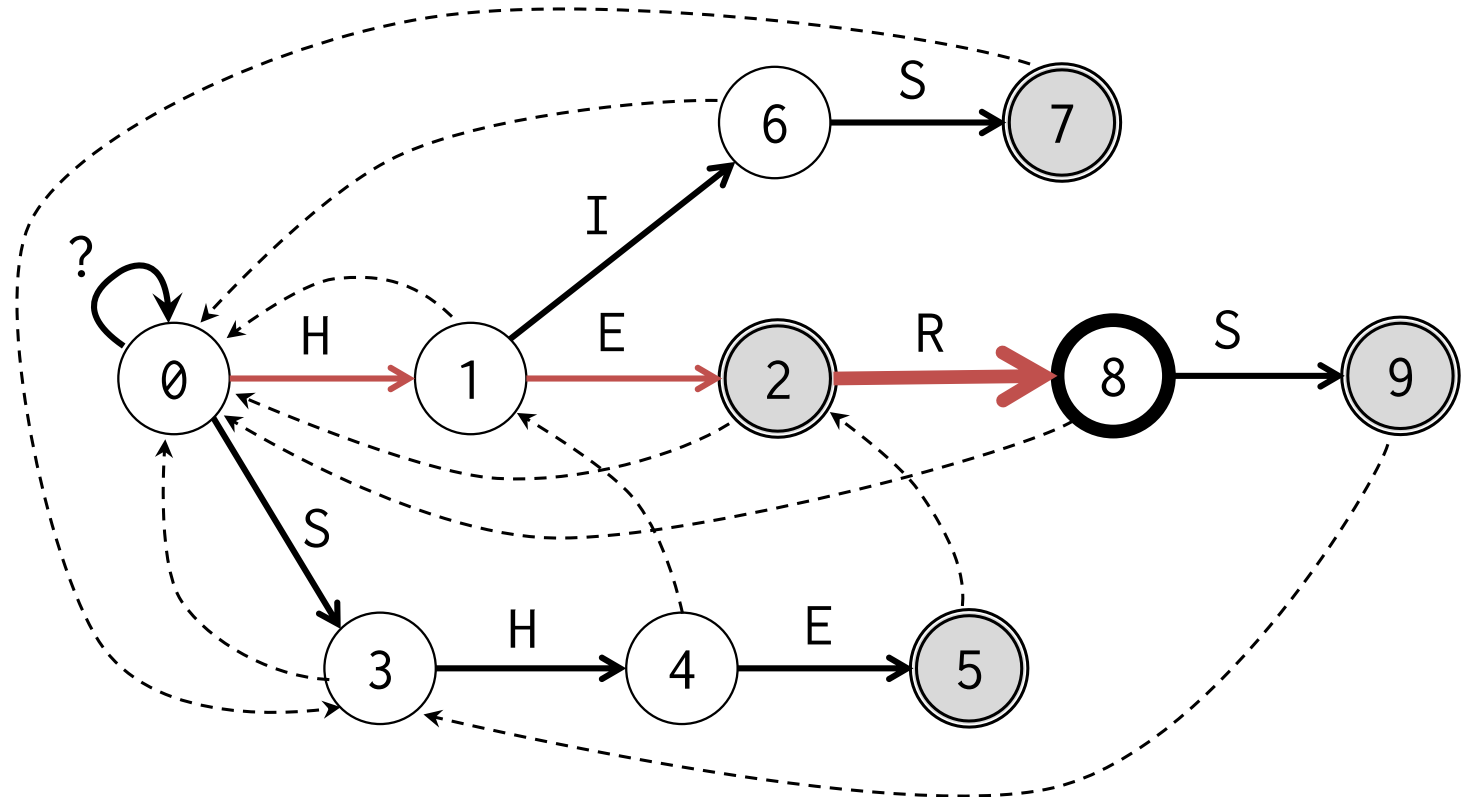
- Reconhecimento de  
HERSHEYS:

0

H: 1

E: 2 {HE}

R: 8



# Aho Corasick

- Reconhecimento de HERSHEY'S:

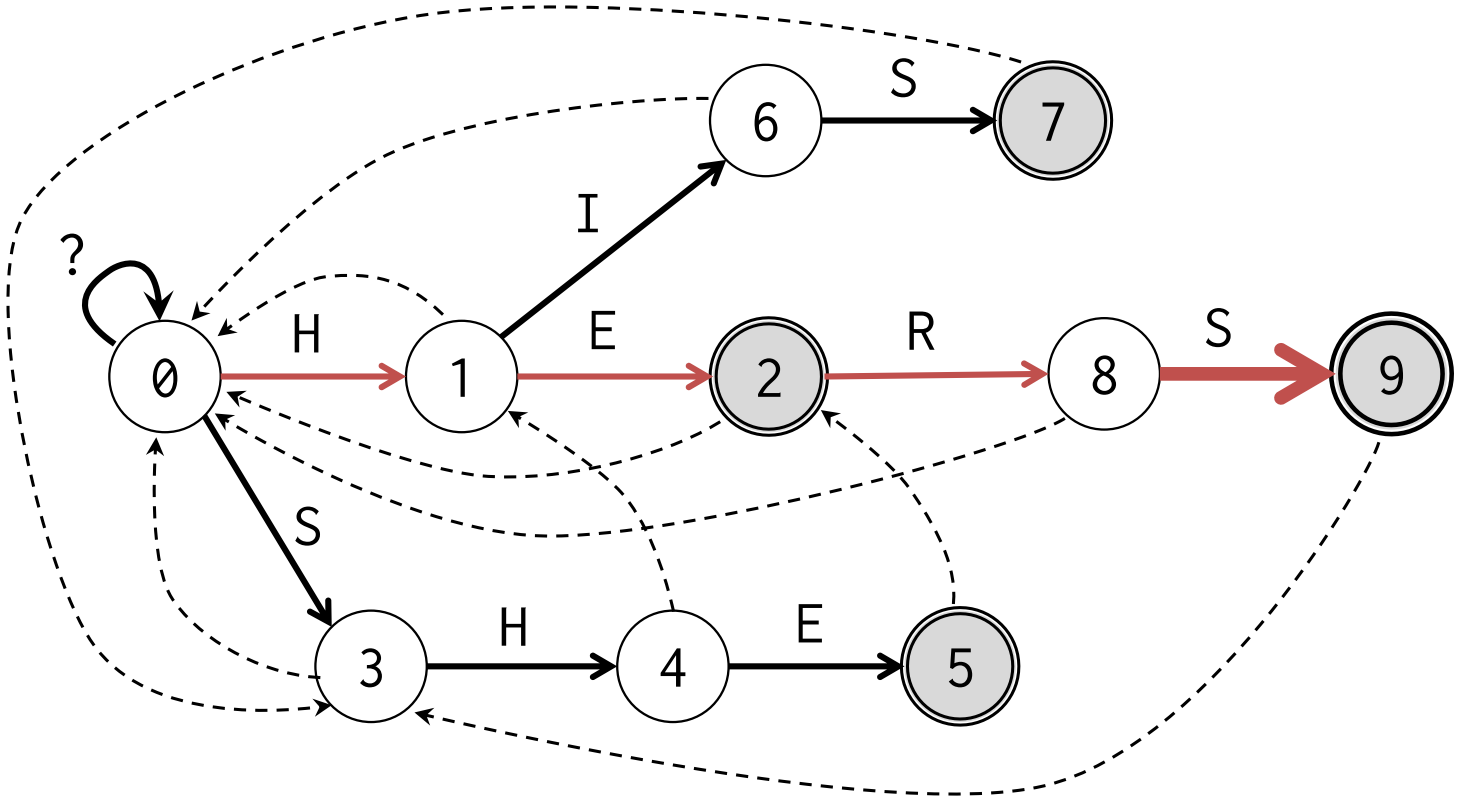
0

H: 1

E: 2 {HE}

R: 8

S: 9 {HERS}



# Aho Corasick

- Reconhecimento de HERSHEYs:

0

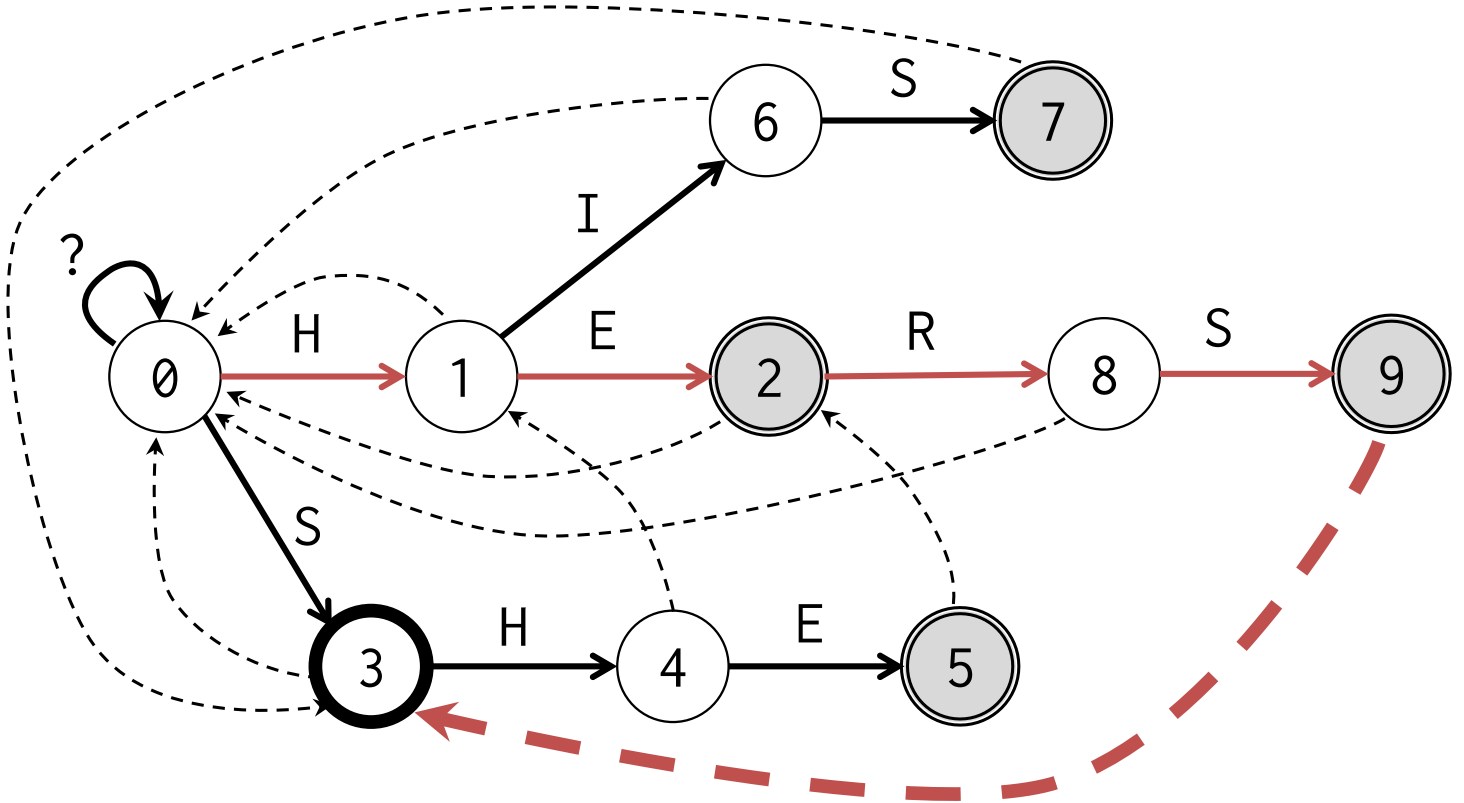
H: 1

E: 2 {HE}

R: 8

S: 9 {HERS}

H: 3



# Aho Corasick

- Reconhecimento de HERSHEYs:

0

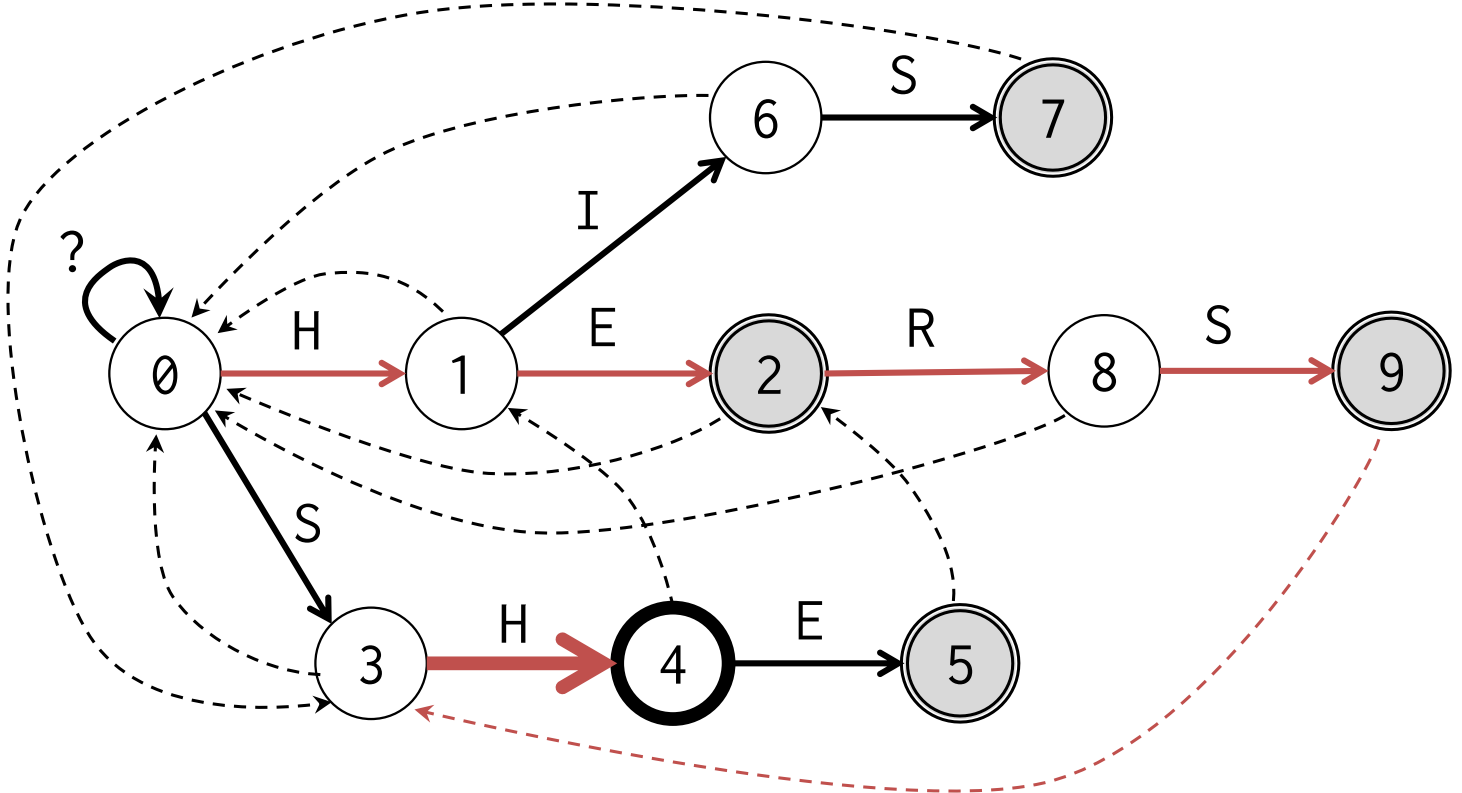
H: 1

E: 2 {HE}

R: 8

S: 9 {HERS}

H: 3, 4



# Aho Corasick

- Reconhecimento de  
HERSHEYS:

0

H: 1

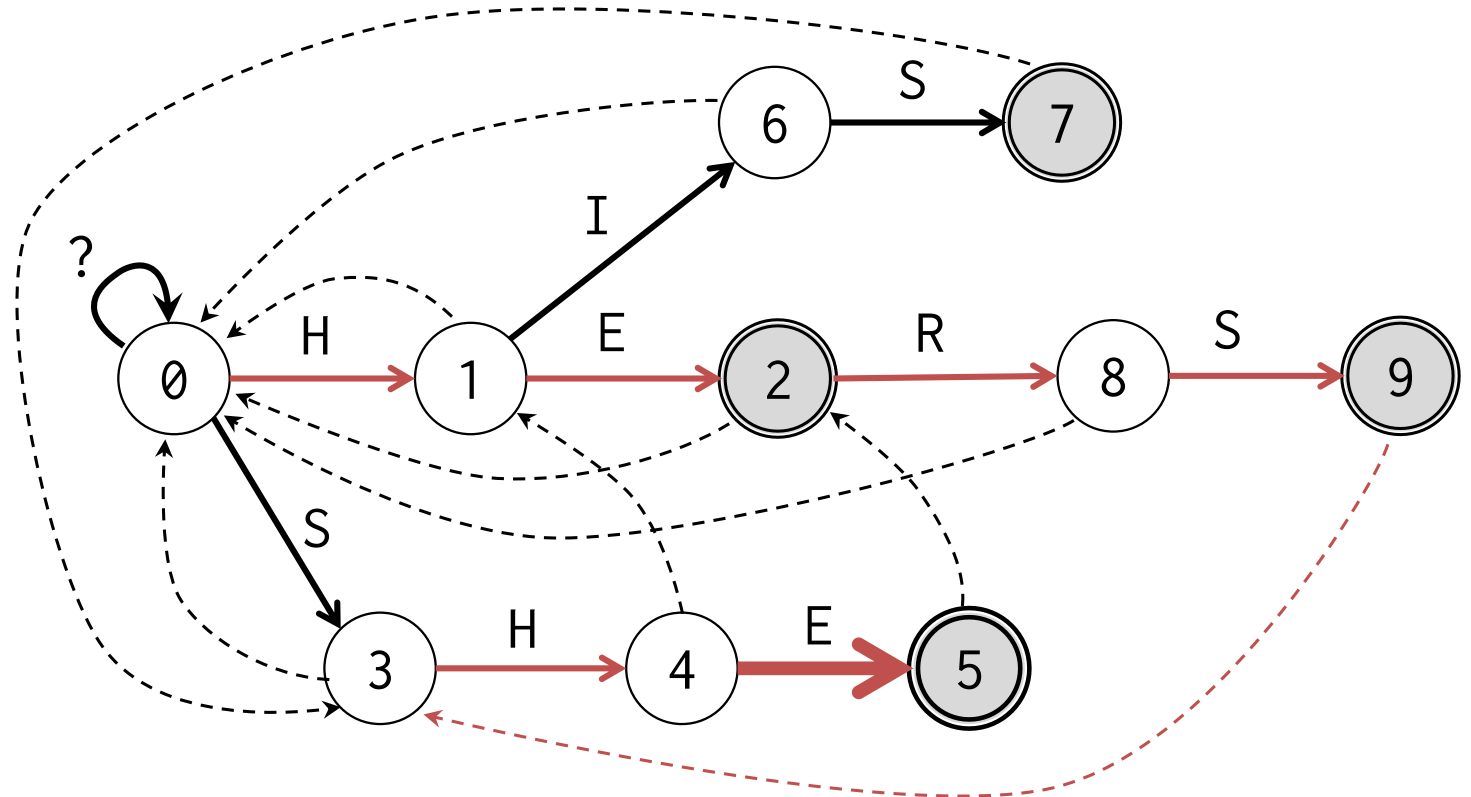
E: 2 {HE}

R: 8

S: 9 {HERS}

H: 3, 4

E: 5 {SHE, HE}

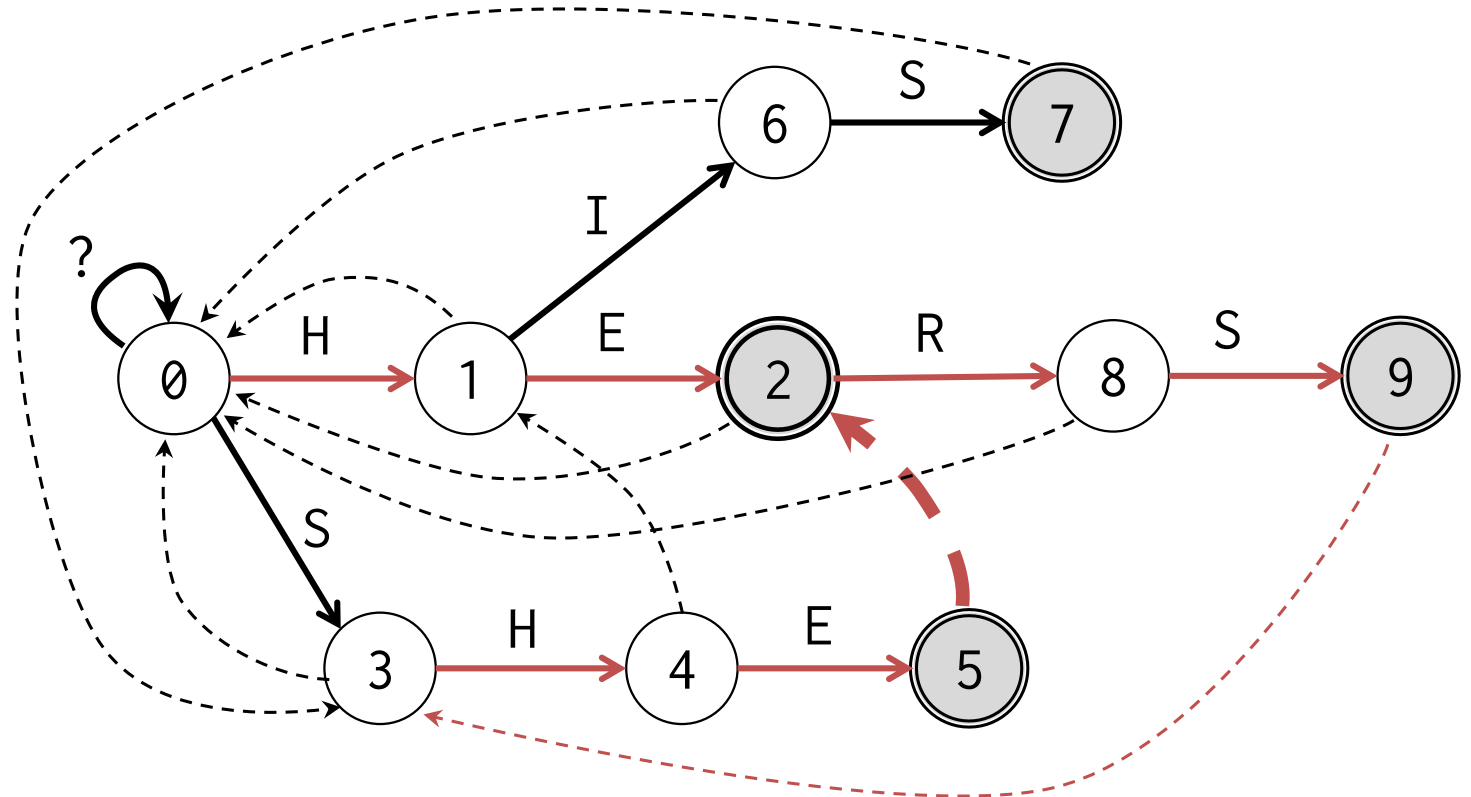




# Aho Corasick

- Reconhecimento de  
HERSHEYS:

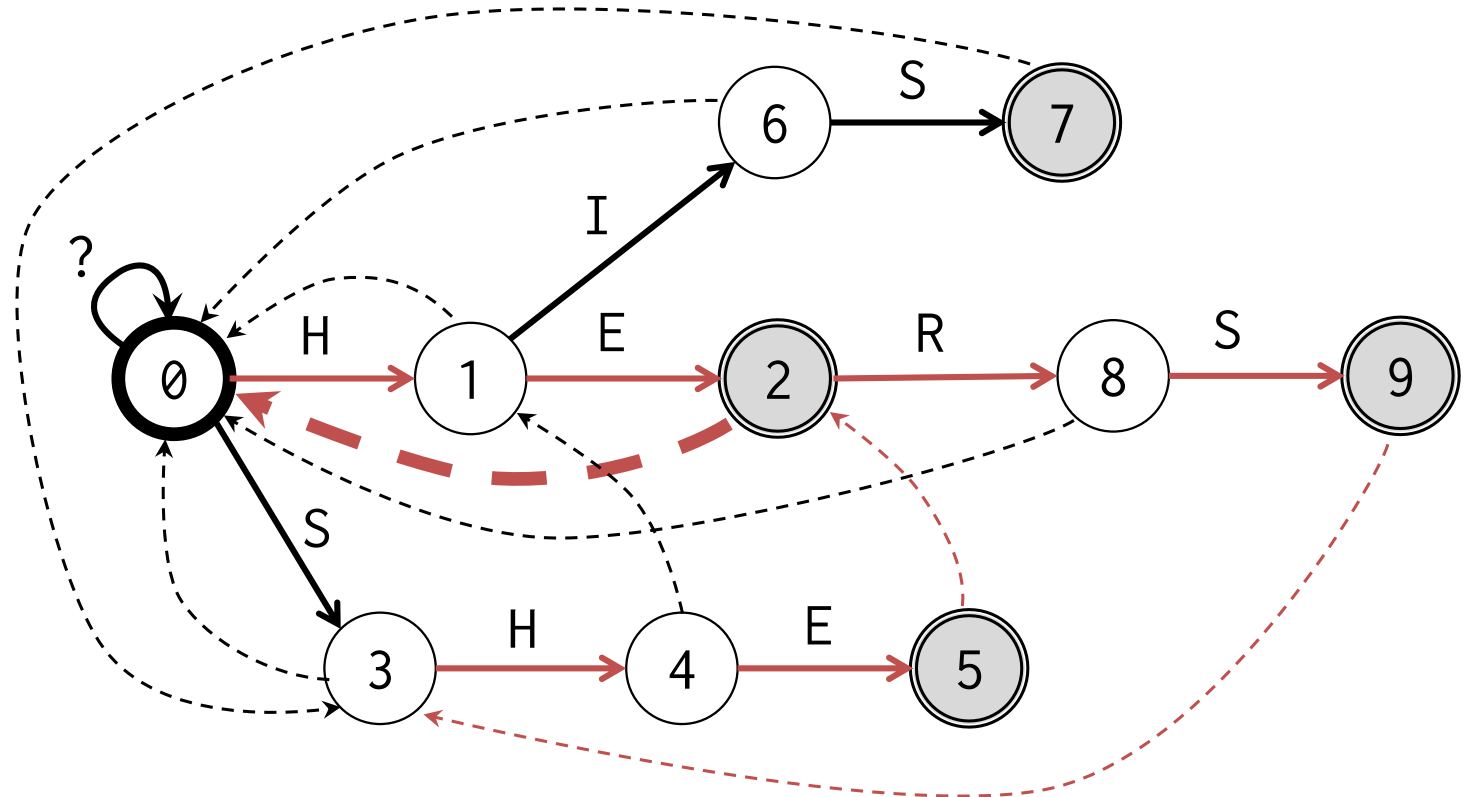
0  
H: 1  
E: 2 {HE}  
R: 8  
S: 9 {HERS}  
H: 3, 4  
E: 5 {SHE, HE}  
Y: 2



# Aho Corasick

- Reconhecimento de  
HERSHEYS:

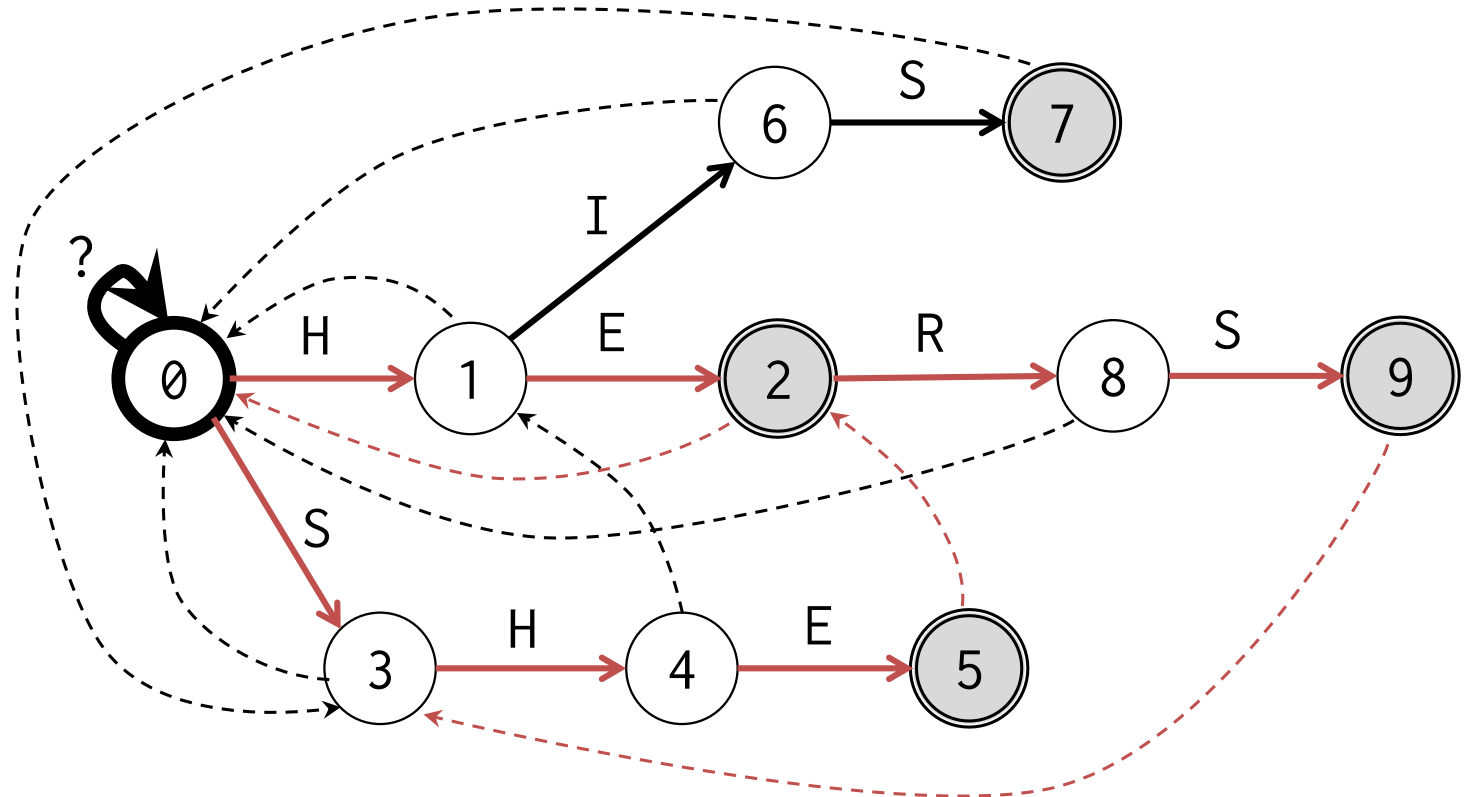
0  
H: 1  
E: 2 {HE}  
R: 8  
S: 9 {HERS}  
H: 3, 4  
E: 5 {SHE, HE}  
Y: 2, 0



# Aho Corasick

- Reconhecimento de  
HERSHEYS:

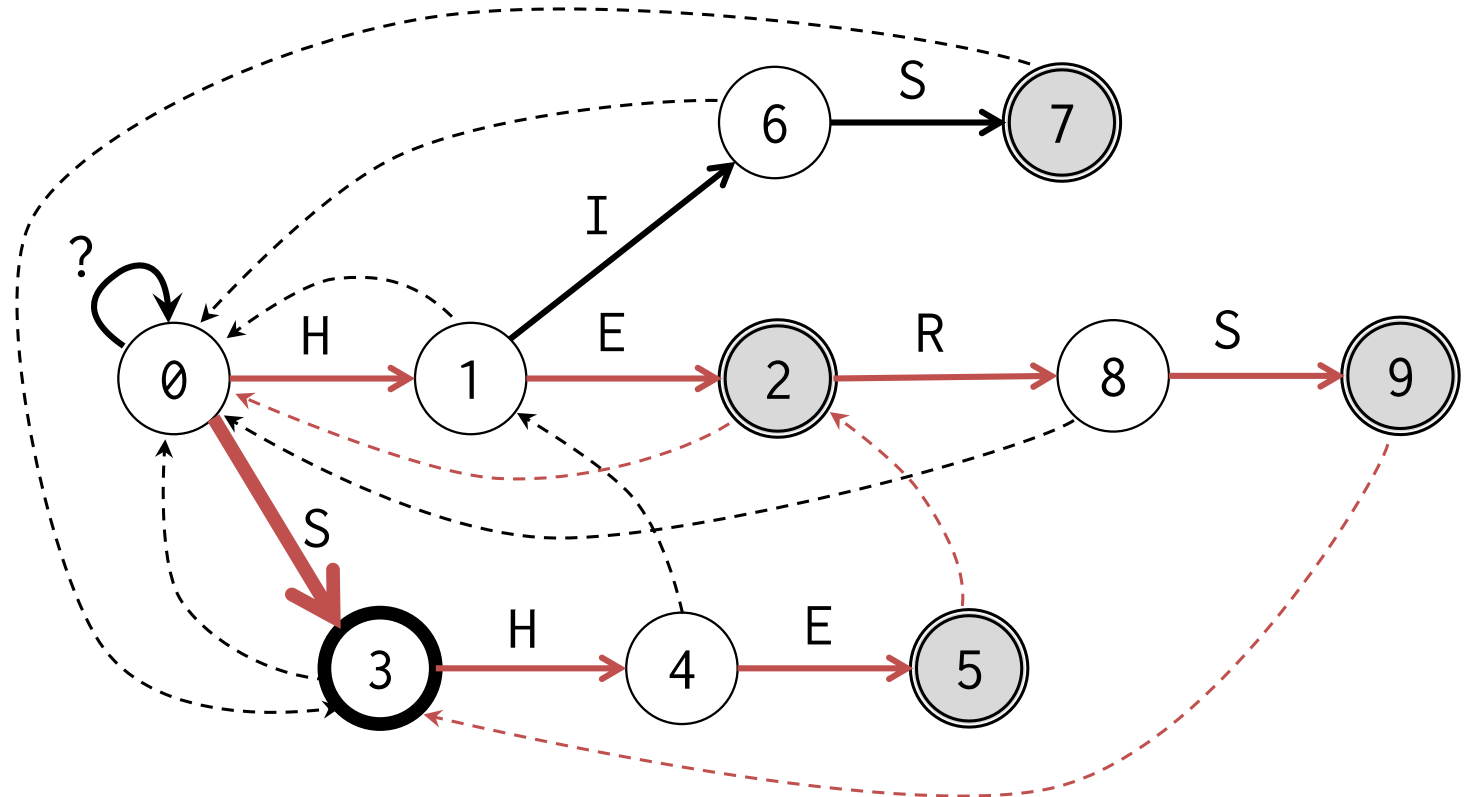
0  
H: 1  
E: 2 {HE}  
R: 8  
S: 9 {HERS}  
H: 3, 4  
E: 5 {SHE, HE}  
Y: 2, 0, 0



# Aho Corasick

- Reconhecimento de  
HERSHEYS:

0  
H: 1  
E: 2 {HE}  
R: 8  
S: 9 {HERS}  
H: 3, 4  
E: 5 {SHE, HE}  
Y: 2, 0, 0  
S: 3



# Boyer Moore (1)

# Boyer Moore

- Criado por Robert S. Boyer e J. Strother Moore em 1977
- **Comparações de caracteres do padrão são feitas da direita para a esquerda**
- São feitos dois testes a cada passo:
  - Deslocamento por caráter ruim
  - Deslocamento por sufixo bom
- O deslocamento final será o maior dos dois

# Boyer Moore

- Deslocamento por carácter ruim
  - O carácter T do texto que não foi encontrado no padrão deve aparecer em outra posição do padrão (ou não adianta fazer comparações que o envolvam)

# Deslocamento por caráter ruim

- Caso 1: o caráter ruim (N) aparece em outra posição do padrão

	0	1	2	3	4	5	6	7
Sequência	A		A	R	A	N	H	A
Padrão	A	R	A	N	H	A		



# Deslocamento por caráter ruim

- Caso 1: o caráter ruim (N) aparece em outra posição do padrão

	0	1	2	3	4	5	6	7
Sequência	A		A	R	A	N	H	A
Padrão	A	R	A	N	H	A		
			A	R	A	N	H	A

# Deslocamento por caráter ruim

- Caso 2: o caráter ruim (U) não aparece em nenhuma outra posição do padrão

	0	1	2	3	4	5	6	7	8	9	10	11
Sequência	A		Á	G	U	A		P	I	N	G	A
Padrão	A	R	A	N	H	A						

# Deslocamento por caráter ruim

- Caso 2: o caráter ruim (U) não aparece em nenhuma outra posição do padrão

	0	1	2	3	4	5	6	7	8	9	10	11
Sequência	A		Á	G	U	A		P	I	N	G	A
Padrão	A	R	A	N	H	A						

# Deslocamento por caráter ruim

- Caso 2: o caráter ruim (U) não aparece em nenhuma outra posição do padrão

	0	1	2	3	4	5	6	7	8	9	10	11
Sequência	A		Á	G	U	A		P	I	N	G	A
Padrão	A	R	A	N	H	A						
						A	R	A	N	H	A	

# Deslocamento por caráter ruim

- Cálculo

0	1	2	3	4	5
A	R	A	N	H	A

Observar a ocorrência mais à direita de cada caráter, exceto o último e preencher o vetor (abaixo) para todos os símbolos possíveis (do texto):

[illegible]

## **Boyer Moore (2)**

# Boyer Moore

- Deslocamento por sufixo bom
  - Analisa repetição de sufixos no padrão (ao invés de prefixos como no KMP)
  - Se um sufixo (ou parte dele) se repetir no padrão, então o deslocamento é feito para testar essa repetição.

# Deslocamento por sufixo bom

- Caso 1: sufixo (BA) se repete com caráter anterior diferente.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	C	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

C	B	A	D	B	A	D	B	A
---	---	---	---	---	---	---	---	---



# Deslocamento por sufixo bom

- Caso 1: sufixo (BA) se repete com caráter anterior diferente.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	C	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

C	B	A	D	B	A	D	B	A			
			C	B	A	D	B	A	D	B	A



Mesmo caráter precede o sufixo

# Deslocamento por sufixo bom

- Caso 1: sufixo (BA) se repete com caráter anterior diferente.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	C	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

C	B	A	D	B	A	D	B	A									
						C	B	A	D	B	A	D	B	A			

# Deslocamento por sufixo bom

- Caso 2: sufixo não se repete com caráter anterior diferente.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	C	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

C	A	B	D	B	A	D	B	A
---	---	---	---	---	---	---	---	---

# Deslocamento por sufixo bom

- Caso 2: sufixo não se repete com caráter anterior diferente.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	C	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

C	A	B	D	B	A	D	B	A															
									C	A	B	D	B	A	D	B	A						

# Deslocamento por sufixo bom

- Caso 3: parte do sufixo se repete no início do padrão

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	D	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

B	A	B	D	B	A	D	B	A
---	---	---	---	---	---	---	---	---

# Deslocamento por sufixo bom

- Caso 3: parte do sufixo se repete no início do padrão

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
B	A	C	B	D	A	D	B	A	D	B	A	D	B	A	C	C	B	A	B	A	D	A	B

B	A	B	D	B	A	D	B	A															
							B	A	B	D	B	A	D	B	A								

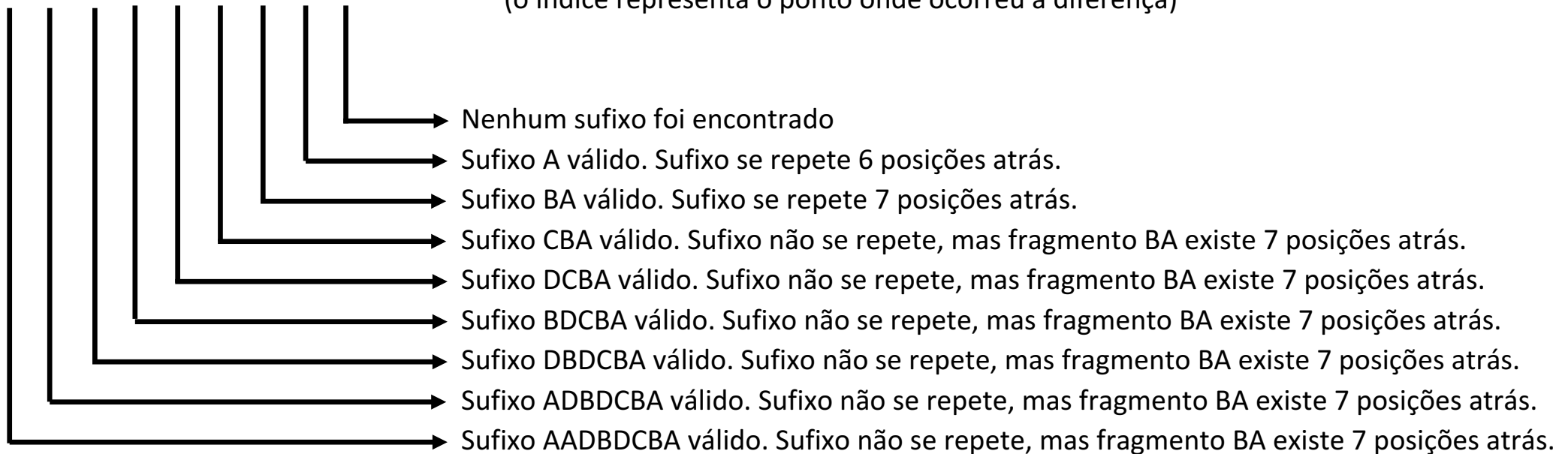
# Deslocamento por sufixo bom

- Cálculo

0	1	2	3	4	5	6	7	8
B	A	A	D	B	D	C	B	A

7	7	7	7	7	7	7	6	1
---	---	---	---	---	---	---	---	---

← Vetor de deslocamentos  
(o índice representa o ponto onde ocorreu a diferença)



# Exemplo

Padrão

0	1	2	3	4	5	6	7
G	C	A	G	A	G	A	G

Vetor para cálculo dos deslocamentos por caráter ruim (DCR)

...	A	B	C	D	E	F	G	H	...
...	6	-1	1	-1	-1	-1	5	-1	...

Vetor para cálculo dos deslocamentos por sufixo bom (DSB)

0	1	2	3	4	5	6	7
G	C	A	G	A	G	A	G
7	7	7	2	7	4	7	1



# Exemplo

- Teste 1

DCR	...	A	B	C	D	E	F	G	H	...
	...	6	-1	1	-1	-1	-1	5	-1	...

DSB	0	1	2	3	4	5	6	7
	7	7	7	2	7	4	7	1

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
G	C	A	T	C	G	C	A	G	A	G	A	G	T	A	T	A	C	A	G	T	A	C	G
							7																
G	C	A	G	A	G	A	G																

Caráter ruim: A. Deslocamento:  $7 - \text{DCR}[ 'A' ] = 7 - 6 = 1$

Sufixo bom: - Deslocamento:  $\text{DSB}[7] = 1$

Melhor deslocamento: 1

# Exemplo

- Teste 2

DCR

...	A	B	C	D	E	F	G	H	...
...	6	-1	1	-1	-1	-1	5	-1	...

DSB	0	1	2	3	4	5	6	7
	7	7	7	2	7	4	7	1

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
G	C	A	T	C	G	C	A	G	A	G	A	G	T	A	T	A	C	A	G	T	A	C	G
						5																	
						G	C	A	G	A	G	A	G										

Caráter ruim: C. Deslocamento:  $5 - \text{DCR}['C'] = 5 - 1 = 4$

Sufixo bom: - Deslocamento:  $\text{DSB}[5] = 4$

Melhor deslocamento: 4

# Exemplo

- Teste 3

...	A	B	C	D	E	F	G	H	...
...	6	-1	1	-1	-1	-1	5	-1	...

0	1	2	3	4	5	6	7
7	7	7	2	7	4	7	1

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
G	C	A	T	C	G	C	A	G	A	G	A	G	T	A	T	A	C	A	G	T	A	C	G

G	C	A	G	A	G	A	G
---	---	---	---	---	---	---	---

Padrão encontrado na posição 5 do texto

**Reconhecimento aproximado de padrões**

# Reconhecimento aproximado

- Algoritmos buscam reconhecer padrões parecidos, considerando:
  - Erros ortográficos (ex.: atensão)
  - Mesmos fonemas (ex.: paço e passo)
  - Variações ortográficas (ex.: Luiz e Luís)
  - Etc.

# Reconhecimento aproximado

- Aplicações:
  - Corretores ortográficos
  - Comparação de DNA
  - Filtragem de SPAM
  - OCR (optical character recognition)

# Reconhecimento aproximado

- Distância de edição
  - Mede a diferença entre duas sequências
  - A diferença é dada pelo número de edições (inserções, exclusões e substituições) necessárias para transformar uma sequência em outra.
- Exemplos:
  - Gato → Pato (1 edição)
  - Paço → Passo (2 edições)

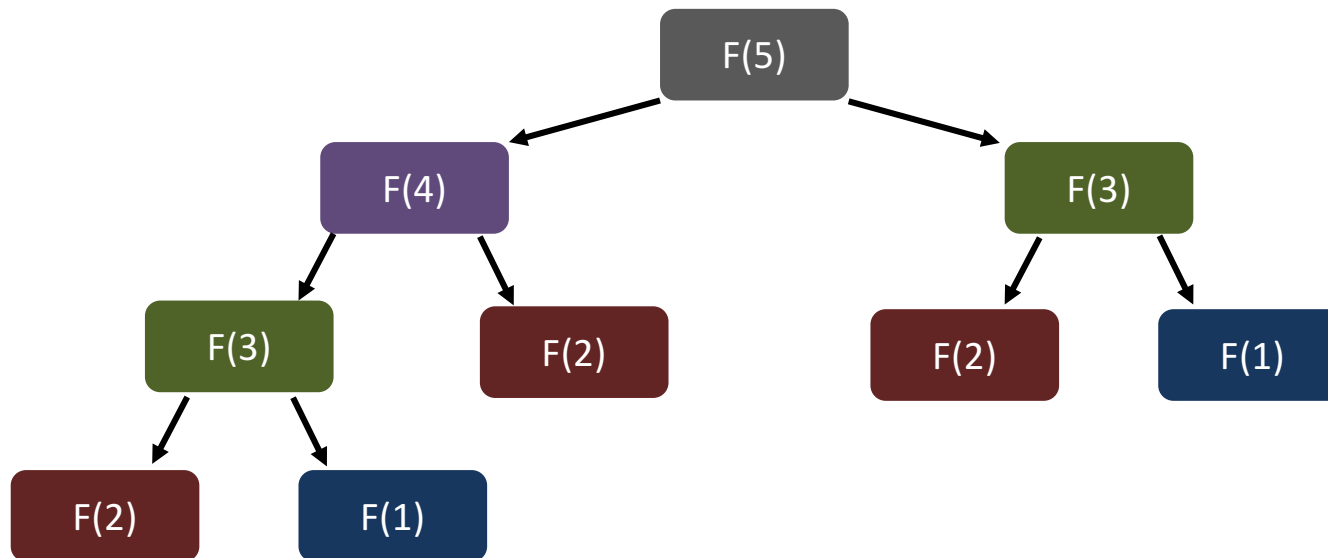
# Reconhecimento aproximado

- Programação dinâmica
  - Método de solução de problemas por meio da sua decomposição em problemas menores
  - Requisitos:
    - Subestrutura ótima – a solução pode ser obtida pela combinação da solução dos subproblemas
    - Sobreposição de subproblemas – a solução do problema pode envolver resolver os mesmos subproblemas várias vezes



# Reconhecimento aproximado

- Programação dinâmica
  - Exemplo de problema
    - Fibonacci –  $F(n) = F(n-1) + F(n-2)$



# Reconhecimento aproximado

- Programação dinâmica
  - Solução – calcular (e guardar em uma tabela) os valores dos “subproblemas”

F(1)	F(2)	F(3)	F(4)	F(5)
1	1	2	3	5



# Reconhecimento aproximado

- Distância de Levenshtein (1965)

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{se } \min(i,j) = 0 \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + C \end{cases} & \text{nos outros casos} \end{cases}$$

em que  $C = 0$ , se  $a_i = b_j$ , ou  $C = 1$ , se  $a_i \neq b_j$ .

# Reconhecimento aproximado

- Exemplo 1:  
CASACO x CASCAO

		C	A	S	A	C	O
	0	1	2	3	4	5	6
C	1						
A	2						
S	3						
C	4						
A	5						
O	6						

# Reconhecimento aproximado

- Exemplo 1:  
CASACO x CASCAO

		C	A	S	A	C	O
	0	1	2	3	4	5	6
C	1	0	1	2	3	4	5
A	2	1	0	1	2	3	4
S	3	2	1	0	1	2	3
C	4	3	2	1	1	1	2
A	5	4	3	2	2	2	2
O	6	5	4	3	3	3	2

# Reconhecimento aproximado

- Exemplo 1:  
PIGARRO x CIGANO

		P	I	G	A	R	R	O
	0	1	2	3	4	5	6	7
C	1	1	2	3	4	5	6	7
I	2	2	1	2	3	4	5	6
G	3	3	2	1	2	3	4	5
A	4	4	3	2	1	2	3	4
N	5	5	4	3	2	2	3	4
O	6	6	5	4	3	3	3	3