
INF2010 – ASD

Recherche de chaînes de caractères

Plan

- I – Problématique
- II – Algorithme naïf
- III – Rabin-Karp
- IV – Automate FSM

I – Problématique

Problématique:

Chercher la chaîne de caractères $P[1..m]$ dans un texte $T[1..n]$.
où $m \leq n$.

Ayant $m \leq n$, on traduit le problème par chercher tous les
 $s \leq n-m+1$ pour lesquels $T[s+1...s+m] = P[1..m]$

Exemples:

- Chercher un mot dans un fichier
- Chercher un fichier dans un volume (HDD, Clé Flash, CD-ROM)
- Chercher un mot dans des fichiers
- Chercher un mot dans le web (google, yahoo)

II – Algorithme naïf

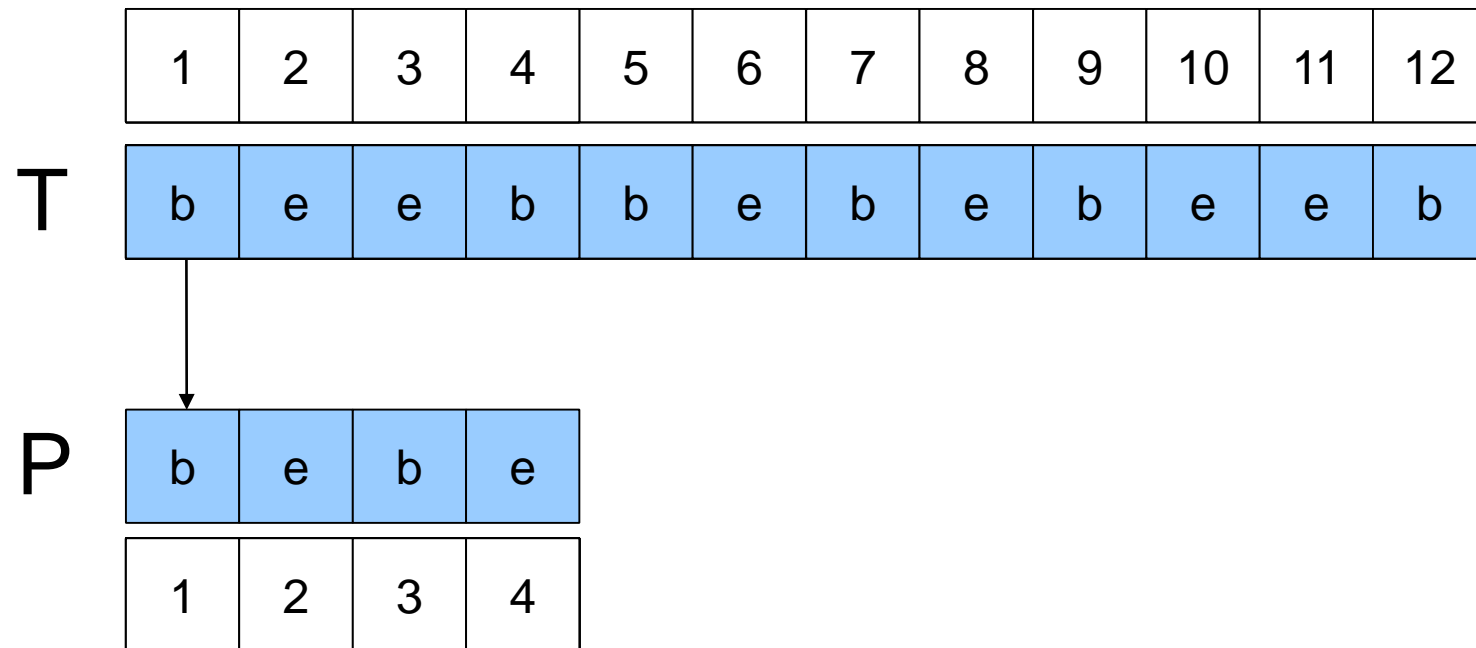
```
Pour s=0 à n-m  
  Pour j= 1 à m  
    Si T[s+j] != P[j]  
      Reprendre au s suivant  
    Sinon Si j=m  
      Inclure s dans S  
Retourner S
```

II – Algorithme naïf

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | b | e | e | b | b | e | b | e | b | e | e | b |
| P | b | e | b | e | | | | | | | | |
| | 1 | 2 | 3 | 4 | | | | | | | | |

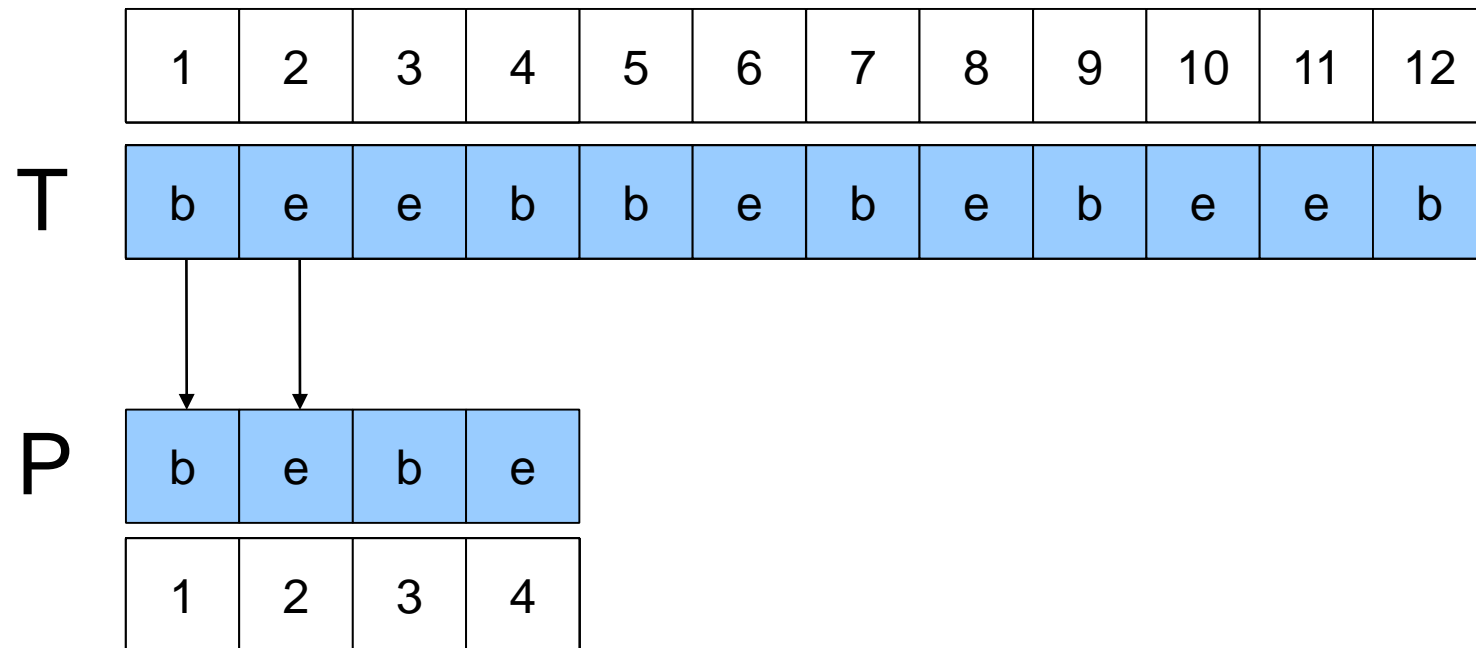
II – Algorithme naïf

$s=0, j=1$



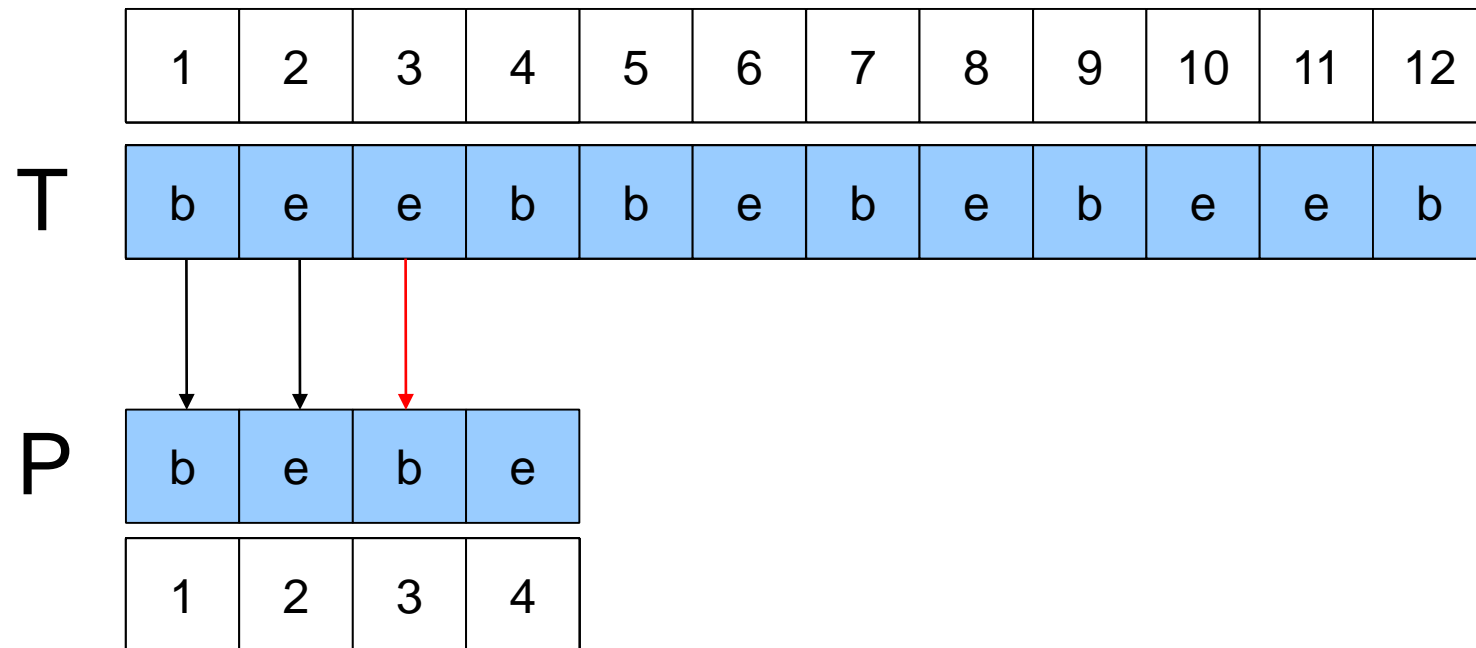
II – Algorithme naïf

$s=0, j=2$



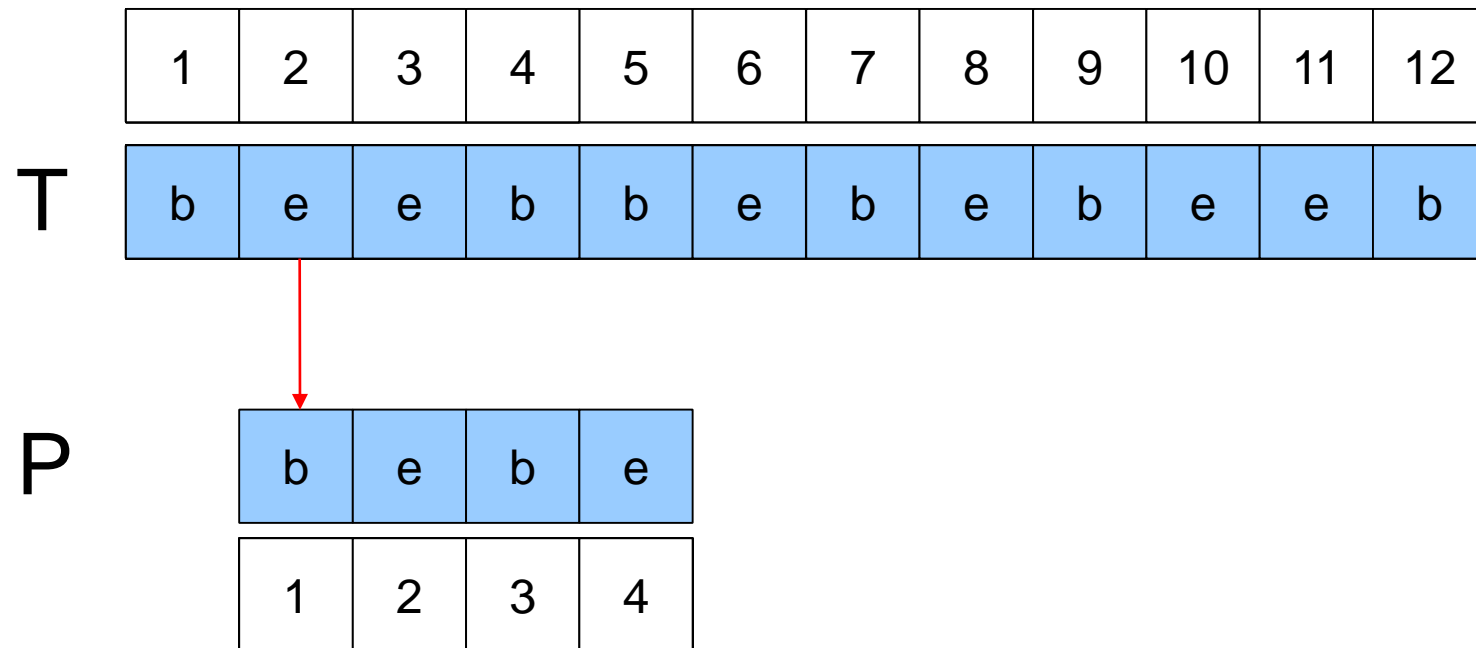
II – Algorithme naïf

$s=0, j=3$



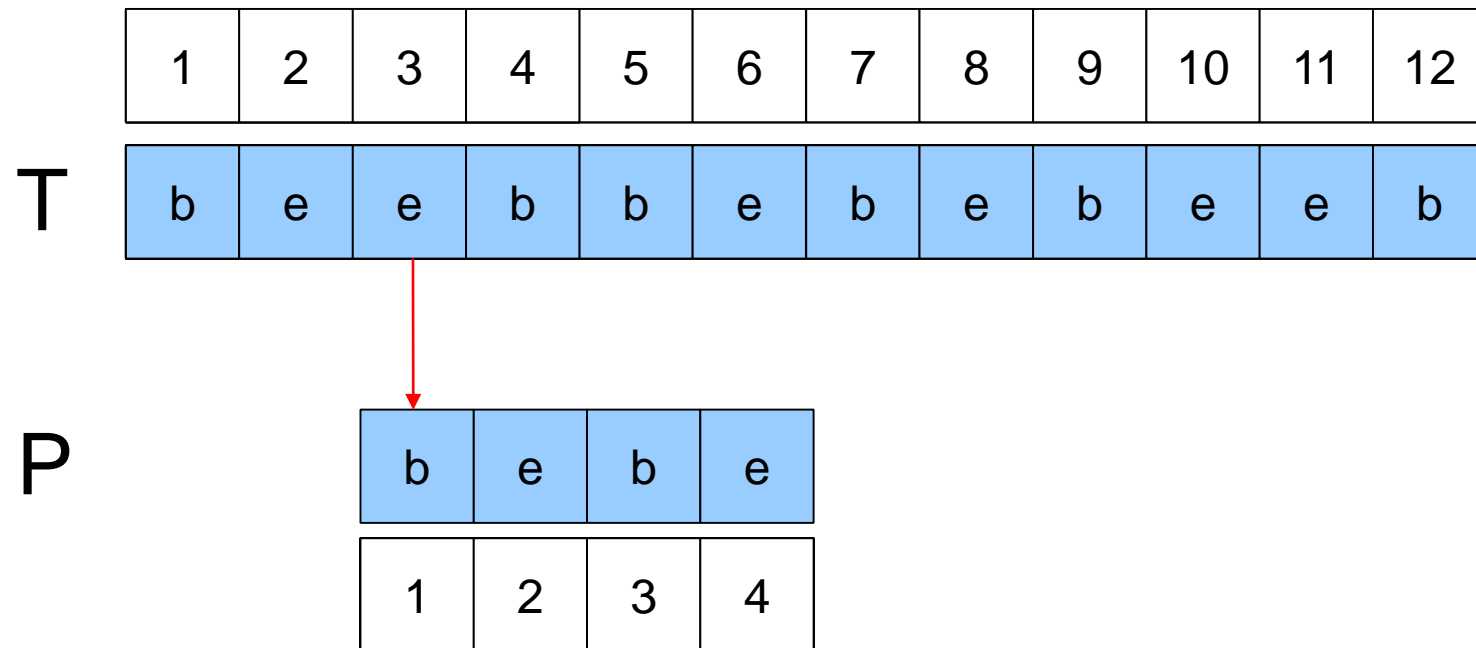
II – Algorithme naïf

$s=1, j=1$



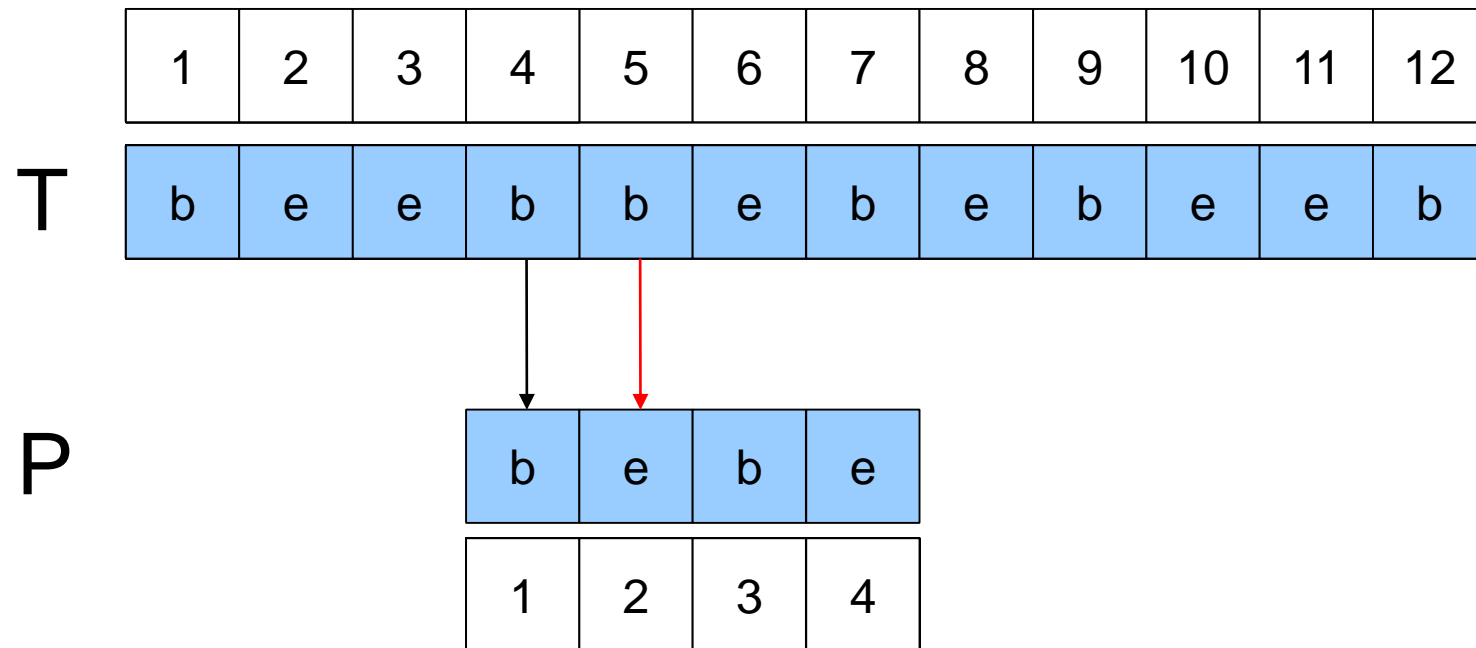
II – Algorithme naïf

$s=2, j=1$



II – Algorithme naïf

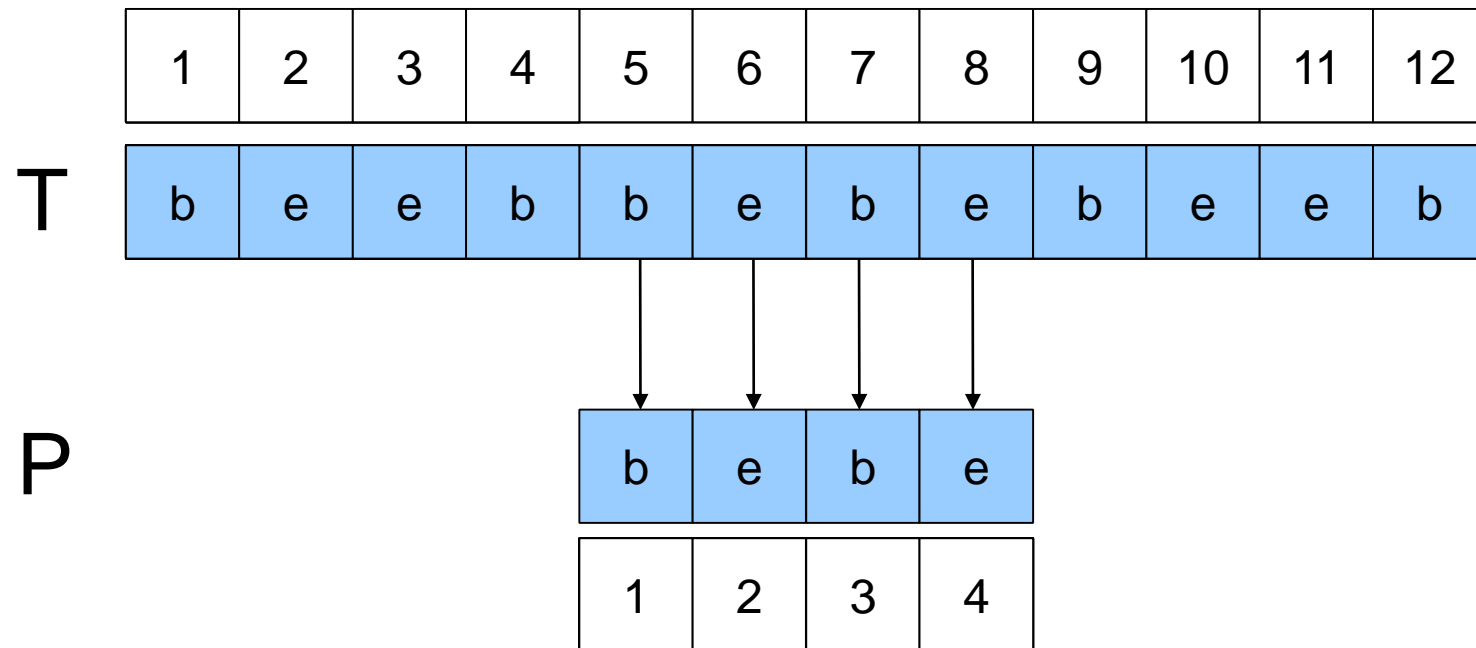
$s=3, j=2$



II – Algorithme naïf

$s=4, j=4=m$

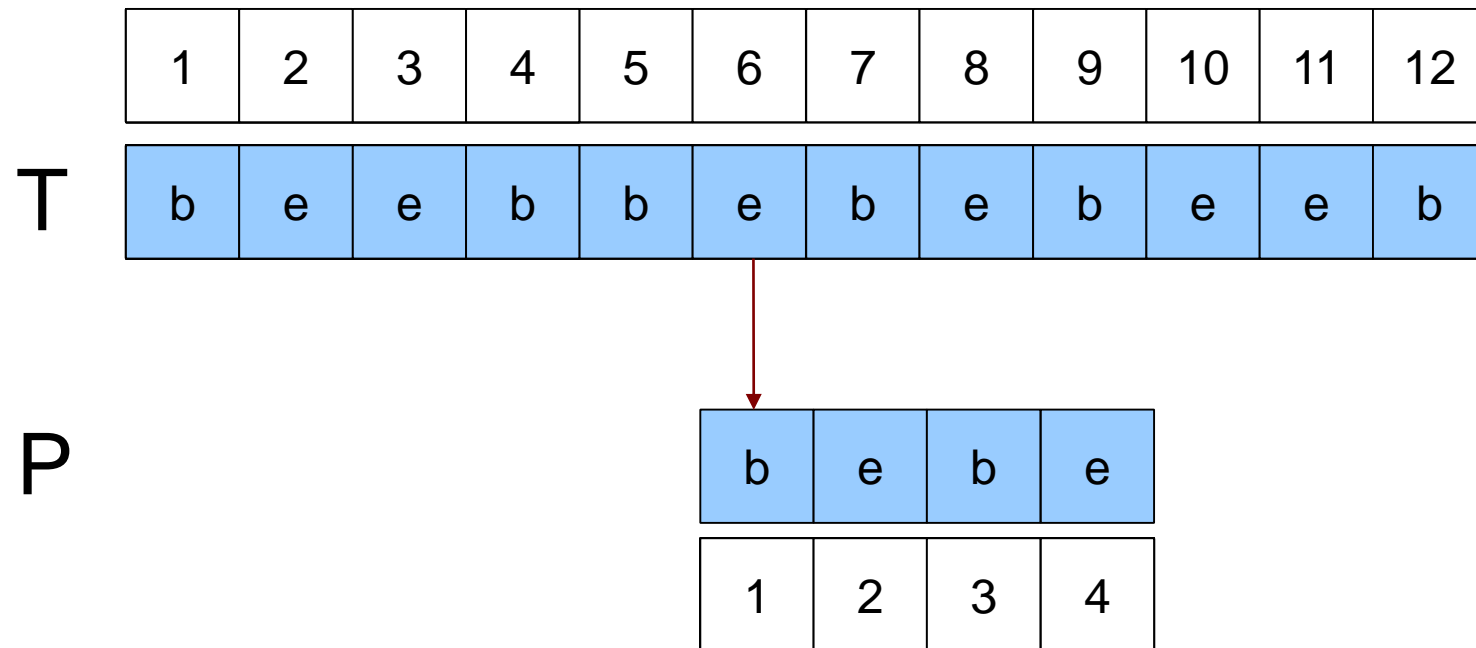
$S=\{4\}$



II – Algorithme naïf

$s=5, j=1$

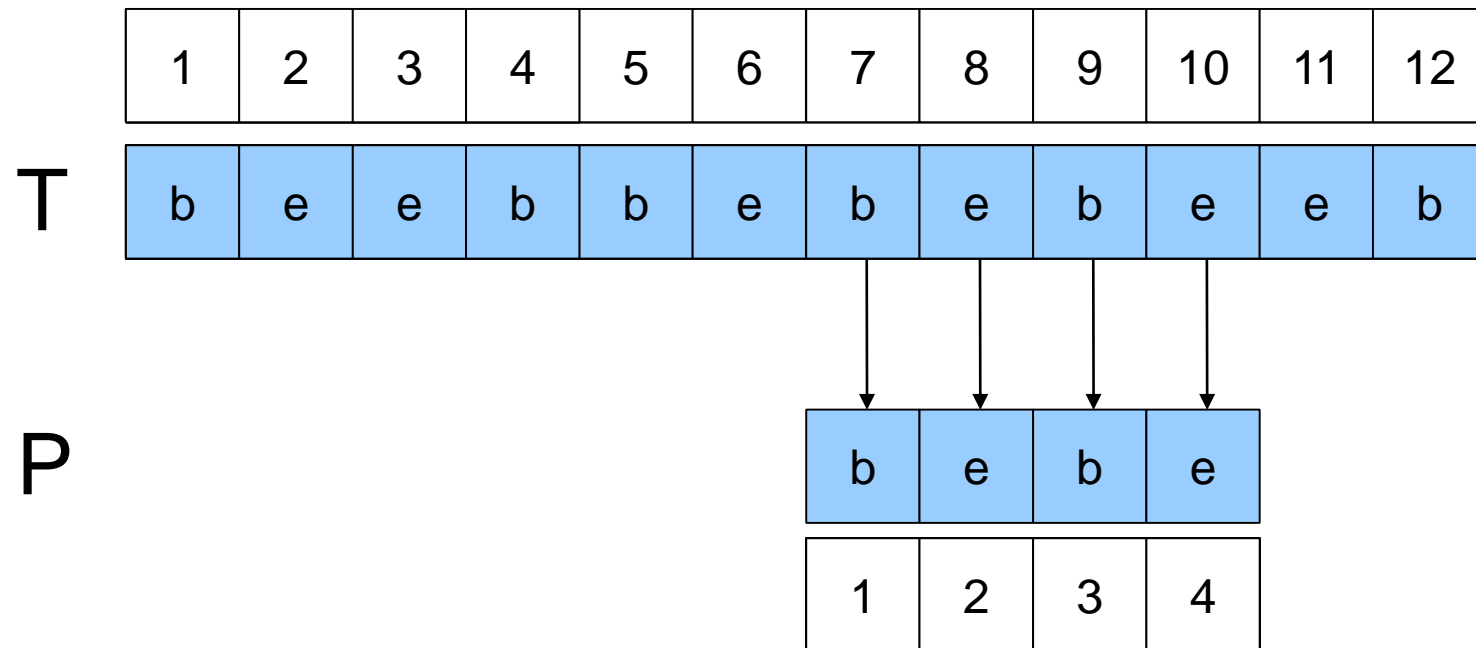
$S=\{4\}$



II – Algorithme naïf

$s=6, j=4$

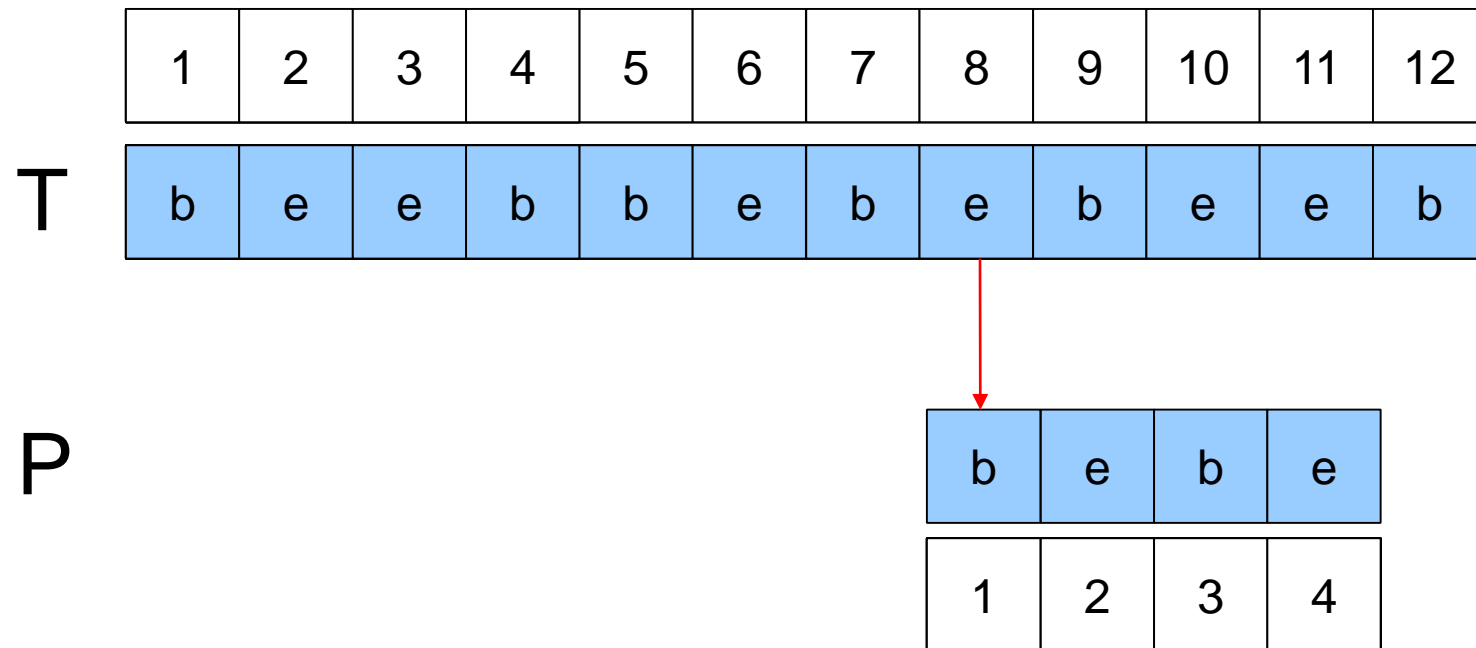
$S=\{4,6\}$



II – Algorithme naïf

$s=7, j=1$

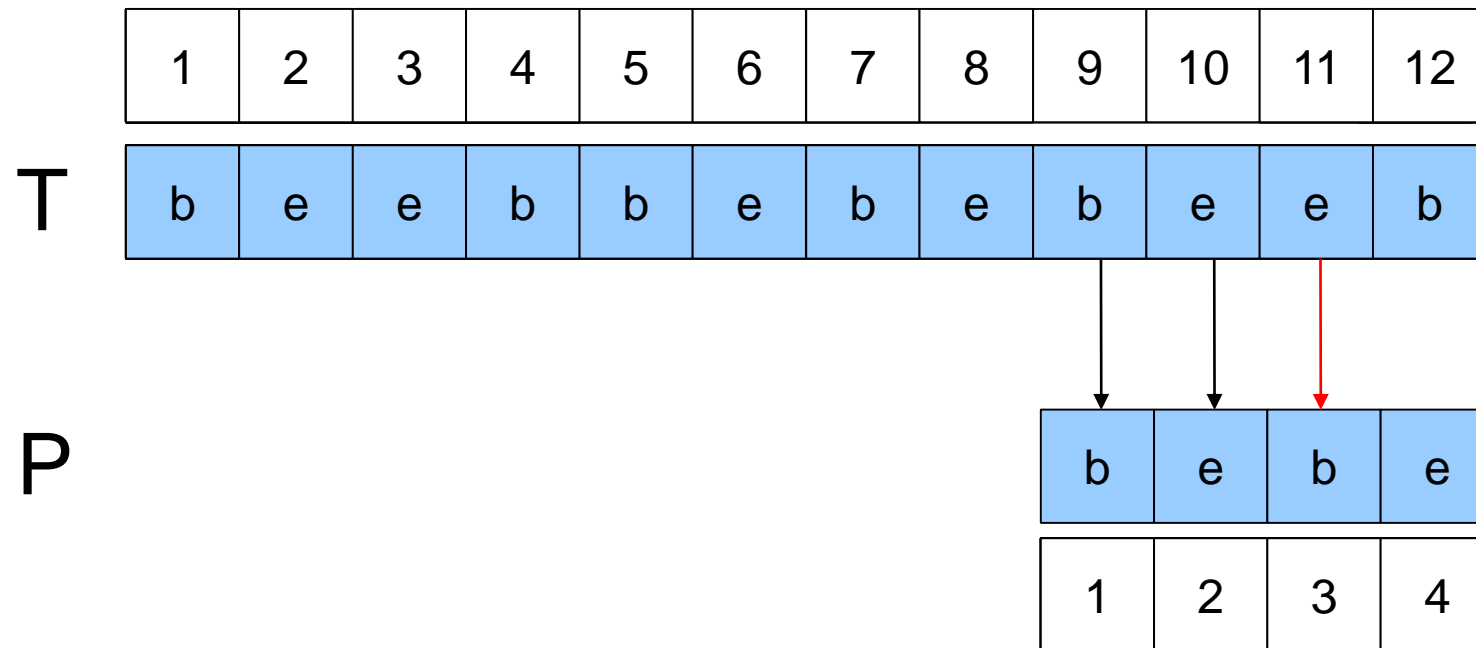
$S=\{4,6\}$



II – Algorithme naïf

$$s=8=12-4=n-m, j=3$$

$$S=\{4,6\}$$

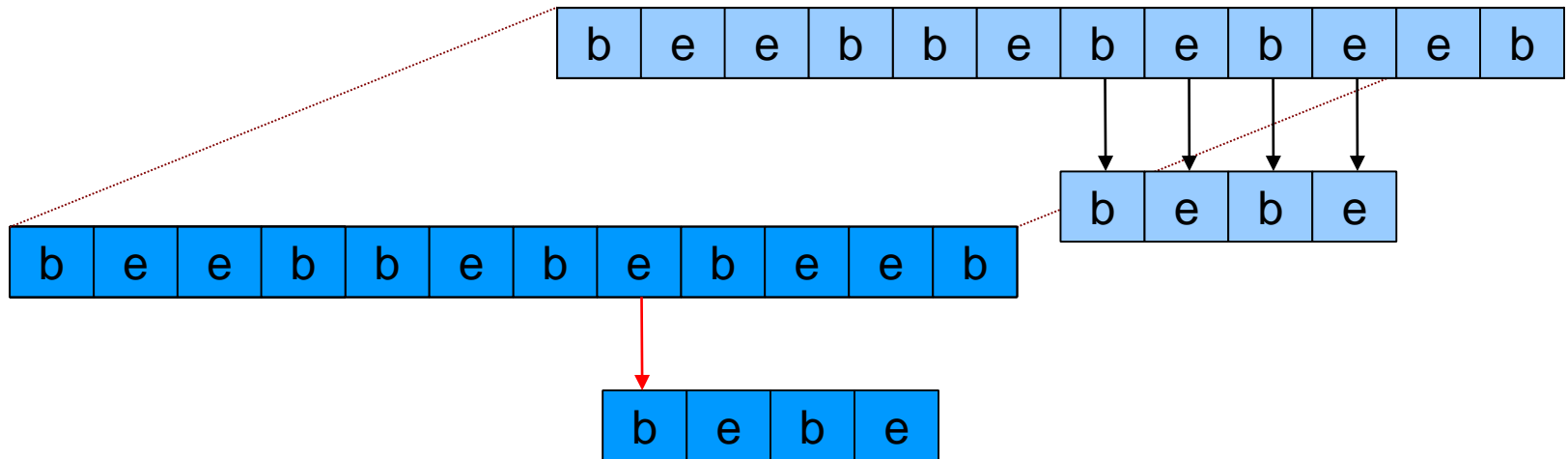


II – Algorithme naïf

Complexité de l'algorithme?

$$O(m(n-m+1))$$

Idée intéressante: réutiliser les résultats et analyser la chaîne de caractère P



III – Rabin-Karp

Objectif: battre l'algorithme naïf de complexité

$$O(m(n-m+1))$$

Analyser la chaîne de caractère:
pétraîtement (*preprocessing*)

Réutiliser les résultats précédent:
accélération

III – Rabin-Karp: formulation 1

Idée:

Pour un alphabet Σ , écrire P sous forme d'un nombre dans la base $d=|\Sigma|$

Analyser la chaîne de caractère:

Trouver une fois la valeur p associée à P

III – Rabin-Karp: formulation 1

Pour l'exemple, on utilise les chiffres 0-9:

Pour un alphabet $\Sigma=\{0, 1, 2, \dots, 8, 9\}$, écrire P sous forme d'un nombre dans la base $d=|\Sigma|=10$

Exemple:

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

$p = 1\ 235$

III – Rabin-Karp: formulation 1

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

$$p = 1\ 235$$

$$p = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 5 \cdot 10^0$$

Pour une longueur quelconque m , la phase de prétraitement peut s'avérer ardue.

III – Rabin-Karp: formulation 1

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

$p = 1\ 235$

$$p = ((1 \cdot 10^1 + 2) \cdot 10^1 + 3) \cdot 10^1 + 5 \cdot 10^0$$

En utilisant l'algorithme de Horner, on réduit le nombre d'opérations pour le prétraitement.

III – Rabin-Karp: formulation 1

Calculer par Horner p de $P[1..m]$

Pour $s=0$ à $n-m$

Calculer par Horner t_s de $T[s+1..s+m]$

Si $t_s = p$

Inclure s dans S

Retourner S

III – Rabin-Karp: formulation 1

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

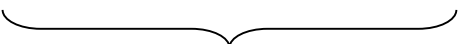
| | | | | |
|---|---|---|---|---|
| P | 1 | 2 | 3 | 5 |
|---|---|---|---|---|

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=0$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$t_s = 1\ 323$

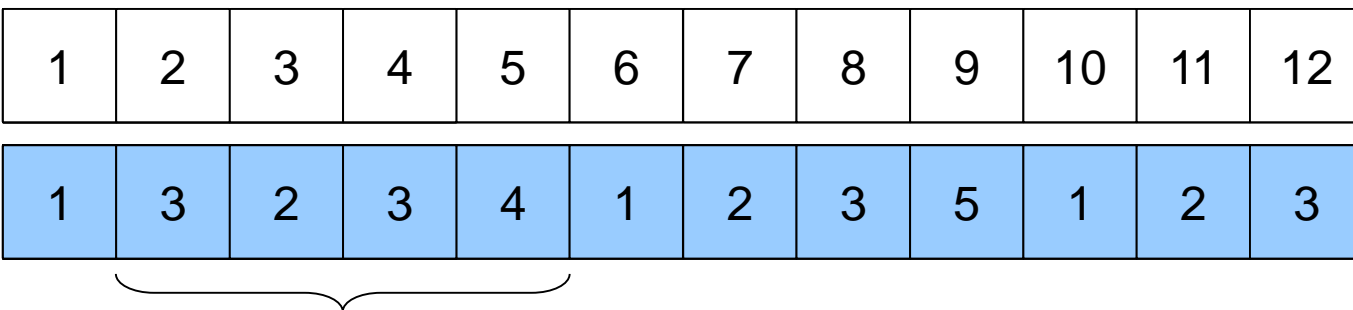
X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=1$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$t_s = 3\ 234$

X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=2$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 2\ 3\ 4\ 1$

X

$p = 1\ 2\ 3\ 5$

III – Rabin-Karp: formulation 1

$s=3$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3\ 4\ 1\ 2$

X

$p = 1\ 2\ 3\ 5$

III – Rabin-Karp: formulation 1

$s=4$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 4\ 1\ 2\ 3$

X

$p = 1\ 2\ 3\ 5$

III – Rabin-Karp: formulation 1

$s=5$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 1\ 235$

✓

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=6$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 2\ 351$

X

$p = 2\ 351$

III – Rabin-Karp: formulation 1

$s=7$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3\ 512$

X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$$s=8=n-m$$

$$S=\{5\}$$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

t_s = 5 123
X

$$p = 1\ 235$$

III – Rabin-Karp: formulation 1

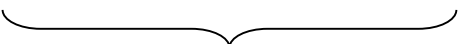
Réutiliser les résultats précédent:
accélération

$$t_{s+1} = 10(t_s - 10^{m-1}T[s+1]) + T[m+s+1]$$

III – Rabin-Karp: formulation 1

$s=0$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$t_s = 1\ 323$

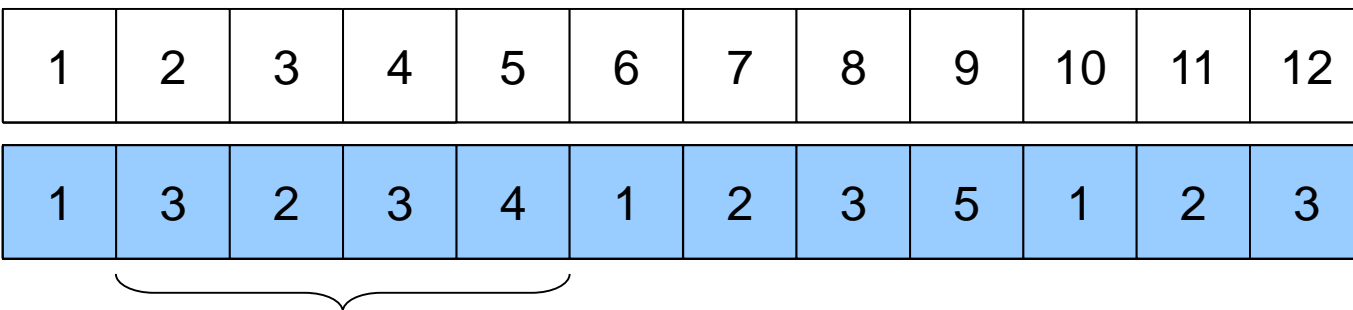
X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

s=1

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$T_s = 10(1\ 323 - 1000 \cdot 1) + 4 = 3\ 234$$

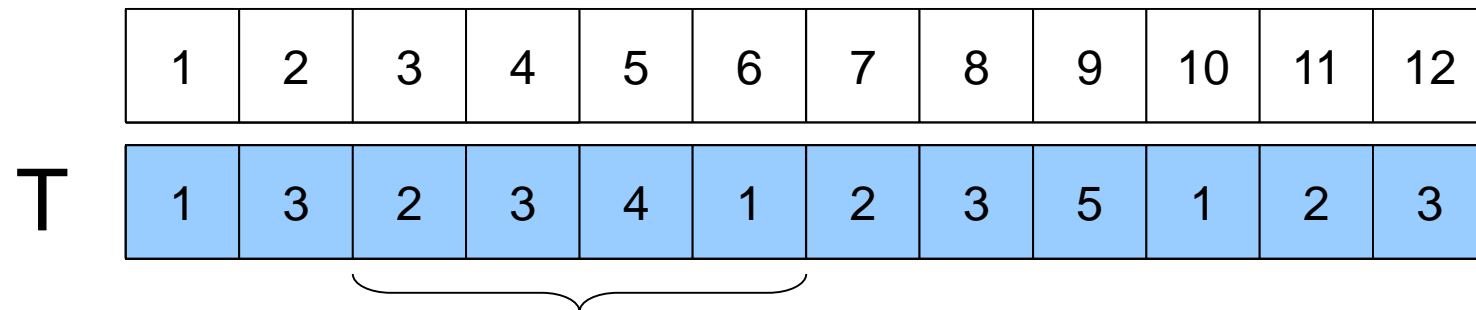
X

p = 1 235

III – Rabin-Karp: formulation 1

$s=2$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 10(3\ 234 - 1000 \cdot 3) + 1 = 2\ 341$$

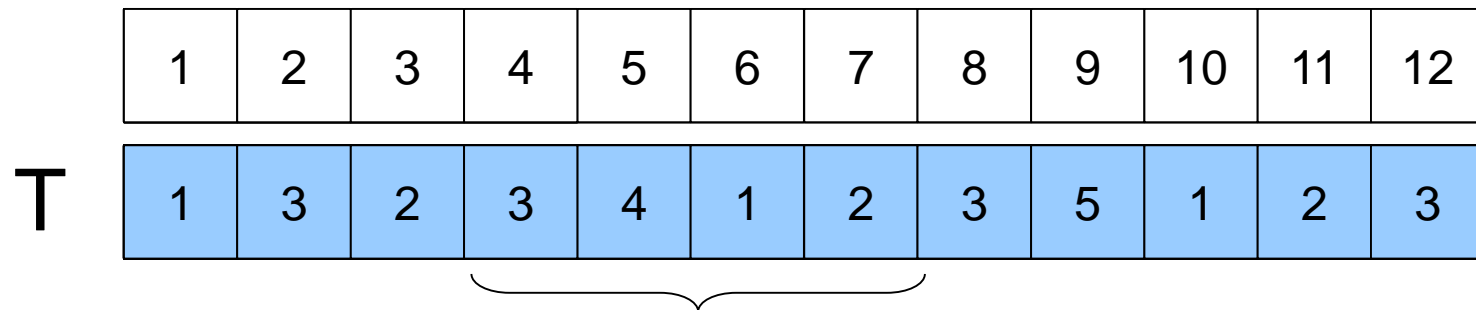
X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=3$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 10(2\ 341 - 1000 \cdot 2) + 2 = 3\ 412$$

X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=4$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|----|----|----|

T

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|

$$t_s = 10(3\ 412 - 1000 \cdot 3) + 3 = 4\ 123$$

X

$p = 1\ 235$

III – Rabin-Karp: formulation 1

$s=5$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$$t_s = 10(4 \ 123 - 1000 \cdot 4) + 5 = 1 \ 235$$

✓

$p = 1 \ 235$

III – Rabin-Karp: formulation 1

$s=6$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$$t_s = 10(1\ 235 - 1000 \cdot 1) + 1 = 2\ 351$$

X

$p = 2\ 351$

III – Rabin-Karp: formulation 1

$s=7$

$S=\{5\}$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|----|----|----|

T

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|---|---|---|

$$t_s = 10(2\ 351 - 1000 \cdot 2) + 2 = 3\ 512$$

X

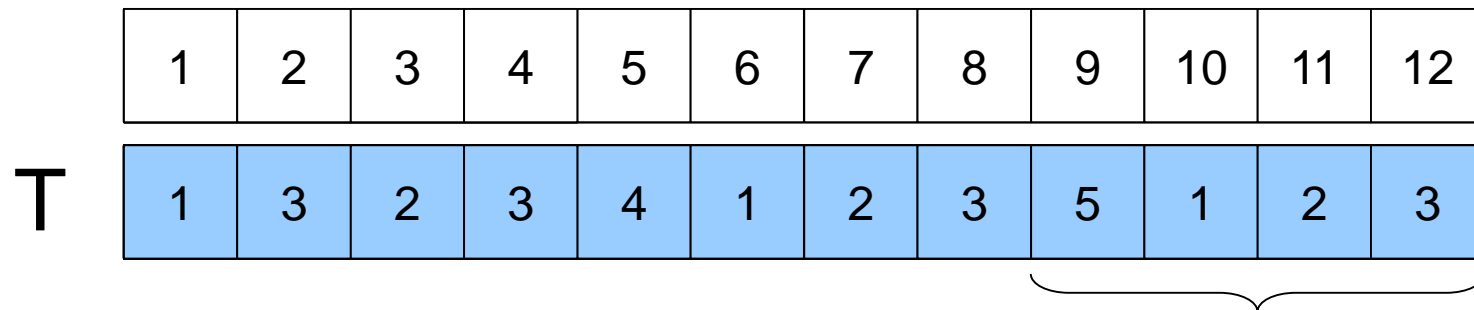
$p = 1\ 235$

III – Rabin-Karp: formulation 1

$$s=8=n-m$$

$$S=\{5\}$$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 10(3\ 512 - 1000 \cdot 3) + 3 = 5\ 123$$

X

$$p = 1\ 235$$

III – Rabin-Karp: formulation 1

Problème:

L'alphabet Σ n'est pas celui des chiffres 0-9 et la base $d=|\Sigma|\neq 10$

ASCII étendu : 256 caractères $d = 256$

UNICODE: 65536 caractères $d = 65536$

Utiliser Horner et la technique itérative sur t_s risque de poser problème quand même. Les nombres vont être grands, et la représentation en virgule flottante risque de donner de faux-positifs.

III – Rabin-Karp: formulation 2

Idée:

Utiliser une écriture modulaire. Pour un alphabet Σ , écrire P sous forme d'un nombre p dans la base $d=|\Sigma|$ modulo q

Choix de q :

On prend q de telle sorte que $d \cdot q$ tiennent dans un mot machine (32 bits) :

Raison: essayer d'exprimer p en fonction de Horner...

III – Rabin-Karp: formulation 2

Pour illustrer l'algorithme, reprenons l'exemple, on utilise les chiffres 0-9: on traduit P sous la forme du nombre p dans la base $d=|\Sigma|=10$ modulo $q = 11$

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

Exemple:

$$p = \{ \{ \{ \{ 1 \cdot 10 + 2 \} \bmod 11 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ \{ \{ 12 \} \bmod 11 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ \{ \{ 1 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ 2 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ 25 \} \bmod 11 = 3$$

III – Rabin-Karp: formulation 2

Pour illustrer l'algorithme, reprenons l'exemple, on utilise les chiffres 0-9: on traduit P sous la forme du nombre p dans la base $d=|\Sigma|=10$ modulo $q = 11$

| | | | | |
|---|---|---|---|---|
| P | 1 | 2 | 3 | 5 |
|---|---|---|---|---|

$$p = 1\ 235 \bmod 11 = 3 \quad !?!$$

Exemple:

$$p = \{ \{ \{ \{ 1 \cdot 10 + 2 \} \bmod 11 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ \{ \{ 12 \} \bmod 11 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ \{ \{ 1 \} \cdot 10 + 3 \} \bmod 11 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ \{ 2 \} \cdot 10 + 5 \} \bmod 11$$

$$p = \{ 25 \} \bmod 11 = 3$$

III – Rabin-Karp: formulation 2

Réutiliser les résultats précédent:
accélération

$$t_{s+1} = \{d(t_s - hT[s+1]) + T[m+s+1]\} \bmod q$$

$$h = d^{m-1} \bmod q$$

III – Rabin-Karp: formulation 2

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

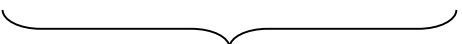
| | | | | |
|---|---|---|---|---|
| P | 1 | 2 | 3 | 5 |
|---|---|---|---|---|

$$p = 1\ 235 \bmod 11 = 3$$

III – Rabin-Karp: formulation 2

$s=0$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 1\ 323 \bmod 11 = 3$$

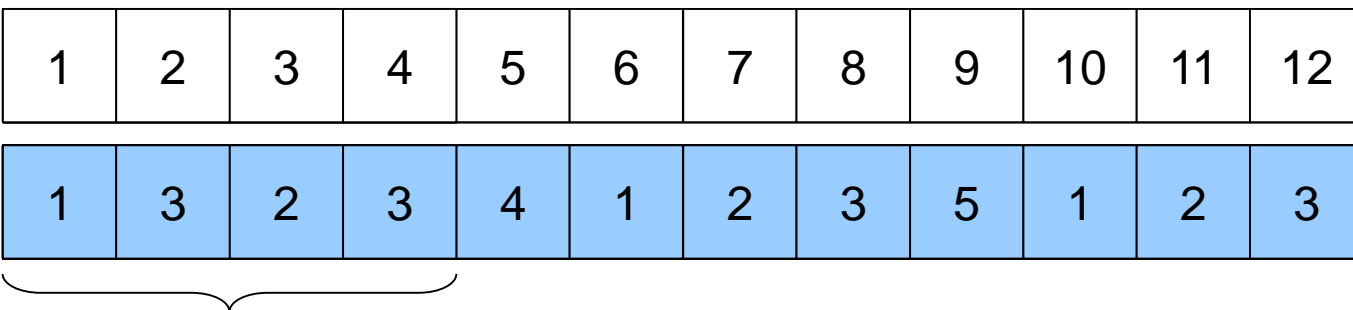
?

$p = 3$

III – Rabin-Karp: formulation 2

$s=0$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 1\ 323 \bmod 11 = 3$$

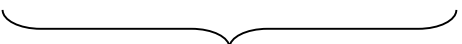
faux-positif

$p = 3$

III – Rabin-Karp: formulation 2

$s=0$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |



$$t_s = 1\ 323 \bmod 11 = 3$$

X

$p = 3$

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

III – Rabin-Karp: formulation 2

$s=1$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

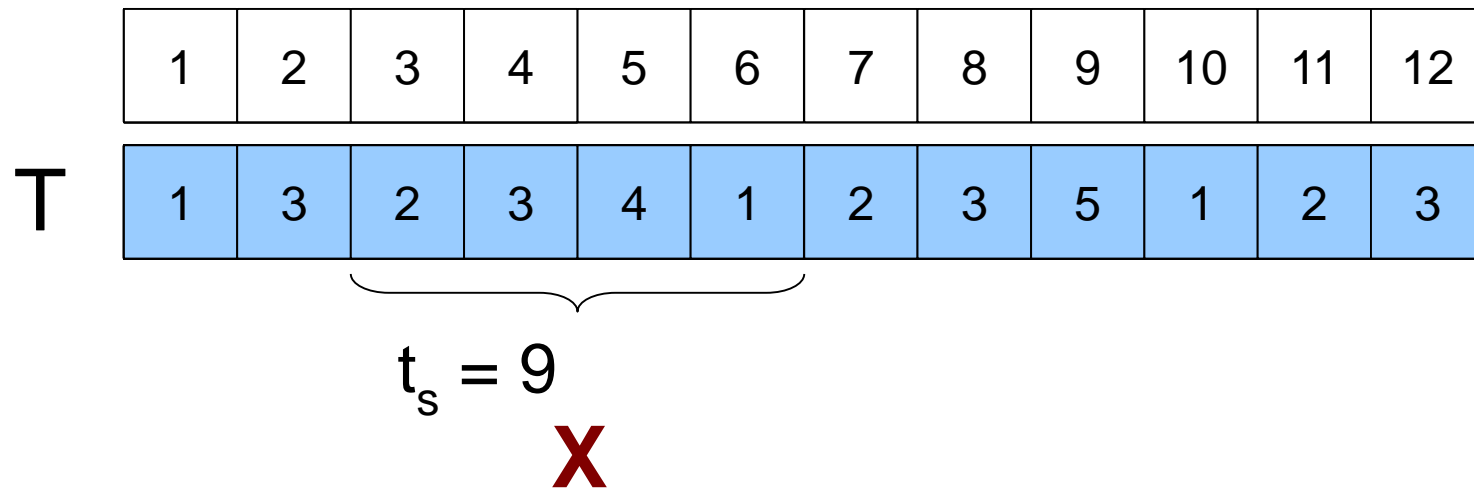
$t_s = 0$

X

$p = 3$

III – Rabin-Karp: formulation 2

$s=2$



$p = 3$

III – Rabin-Karp: formulation 2

$s=3$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 2$

X

$p = 9$

III – Rabin-Karp: formulation 2

$s=4$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 9$
X

$p = 3$

III – Rabin-Karp: formulation 2

$s=5$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3$

?

$p = 3$

III – Rabin-Karp: formulation 2

$s=5$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3$

$p = 3$

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

III – Rabin-Karp: formulation 2

$s=6$

$S=\{5\}$

| | | | | | | | | | | | | |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 8$

X

$p = 3$

III – Rabin-Karp: formulation 2

$s=7$

$S=\{5\}$

| | | | | | | | | | | | | |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3$
?

$p = 3$

III – Rabin-Karp: formulation 2

$s=7$

$S=\{5\}$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

$t_s = 3$

X

$p = 3$

P

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 5 |
|---|---|---|---|

III – Rabin-Karp: formulation 2

$$s=8=n-m$$

$$S=\{5\}$$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| T | 1 | 3 | 2 | 3 | 4 | 1 | 2 | 3 | 5 | 1 | 2 | 3 |

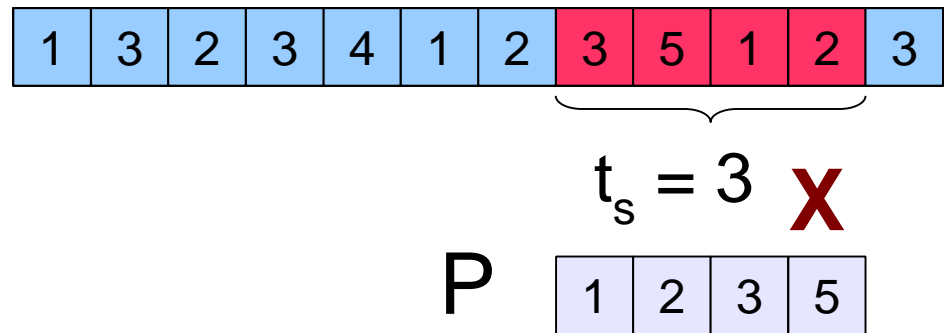
t_s = 5 123
X

$$p = 1\ 235$$

III – Rabin-Karp

Complexité de l'algorithme à cause des faux-positifs

$$O(m(n-m+1))$$



Néanmoins, l'exécution de Rabin Karp est **plus rapide** que l'algorithme naïf.

IV – Automate FSM

Objectif: meilleure complexité

Battre $O(m(n-m+1))$

Meilleure performances que Rabin-Karp

Analyser la chaîne de caractère:

Construire une machine à états (prétraîtement)

Réutiliser les résultats précédent:

L'automate possède une mémoire interne de son état (accélération)

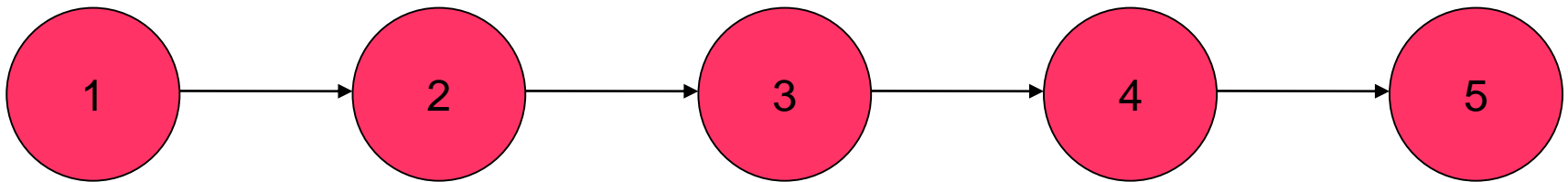
IV – Automate FSM

Construire - Le nombre d'états est égal à $m+1$

Exemple:

P

| | | | |
|---|---|---|---|
| b | e | b | e |
|---|---|---|---|



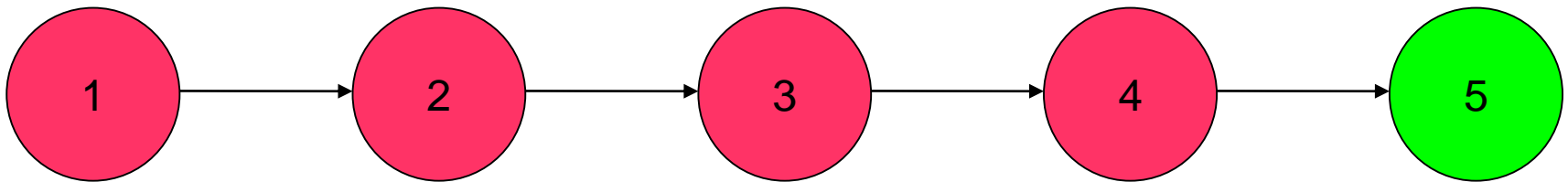
IV – Automate FSM

Construire - Le dernier état valide l'entrée

Exemple:

P

| | | | |
|---|---|---|---|
| b | e | b | e |
|---|---|---|---|



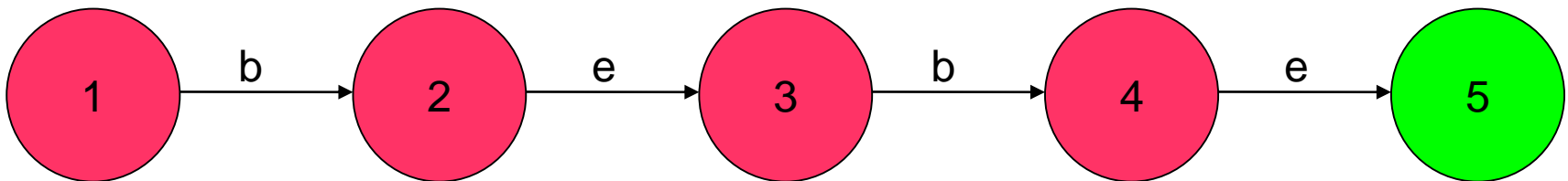
IV – Automate FSM

Construire – Les arcs répondent au unités

Exemple:

P

| | | | |
|---|---|---|---|
| b | e | b | e |
|---|---|---|---|

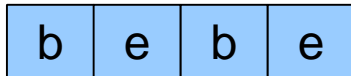


IV – Automate FSM

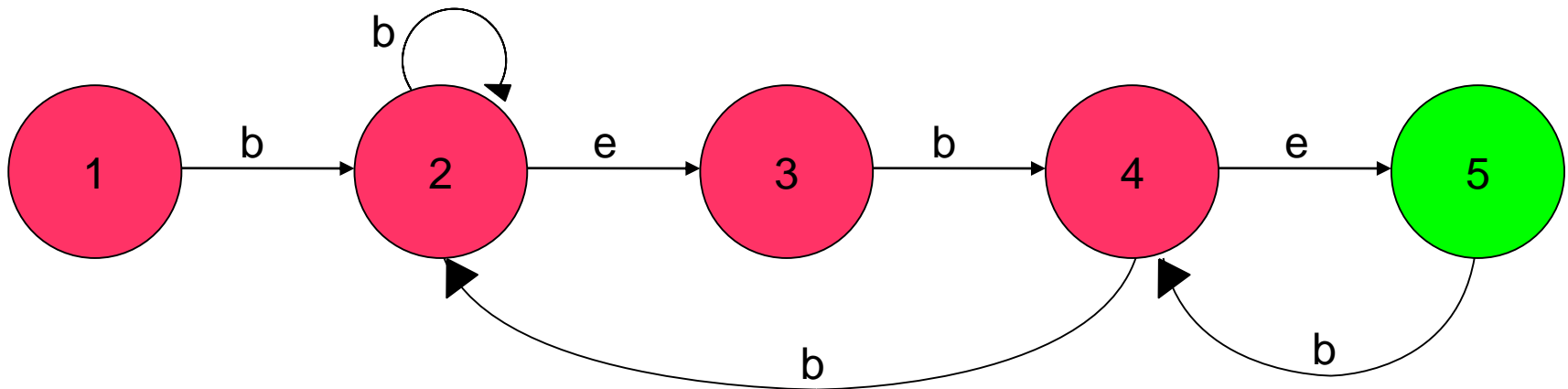
Construire – Ajouter les arcs restants

Exemple:

P

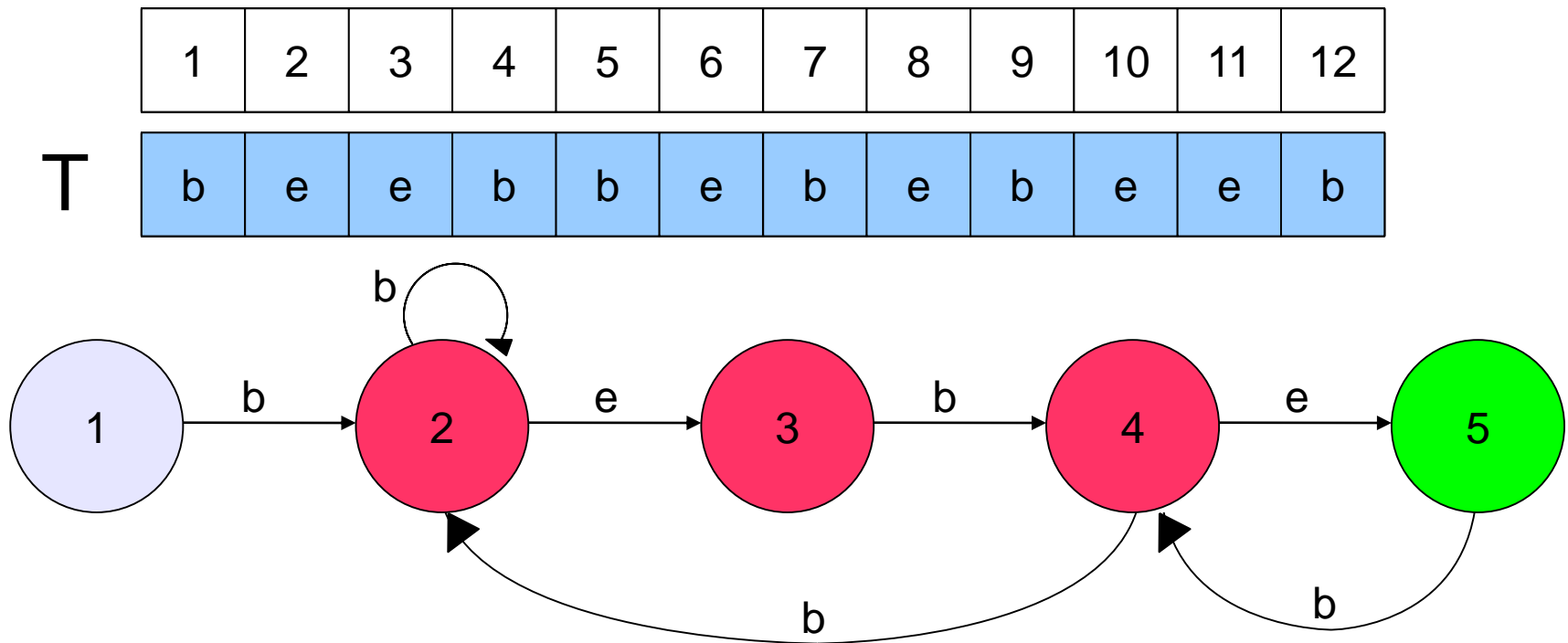


Note: les arcs absents mènent à l'état de départ (1)



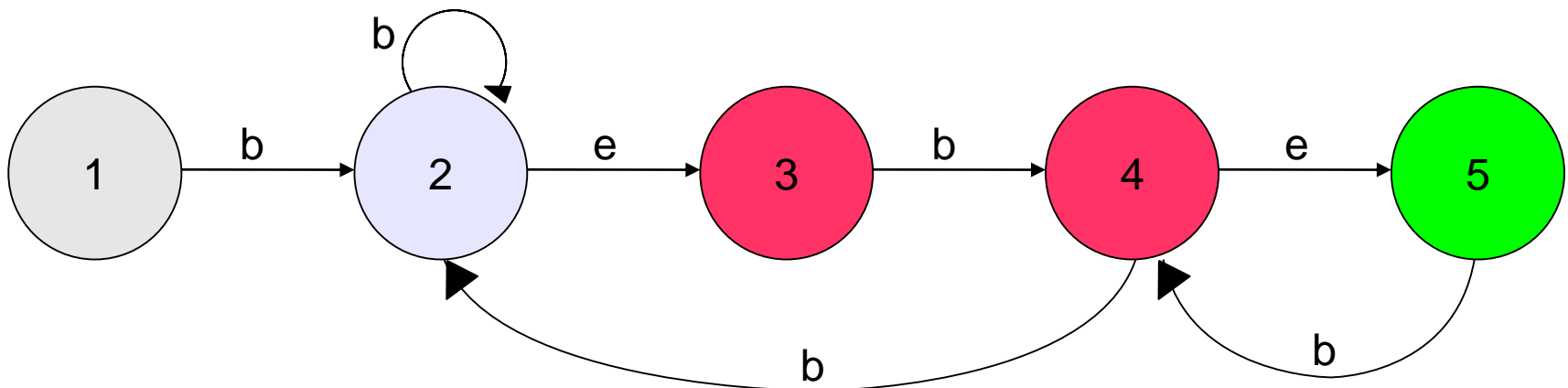
IV – Automate FSM

Traîter – La chaîne T est traversée une fois



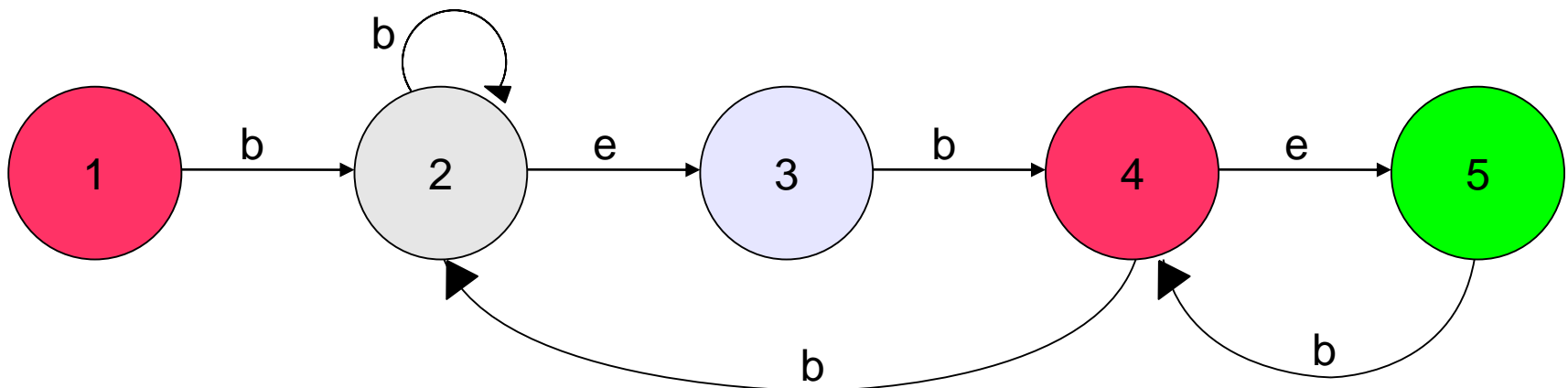
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



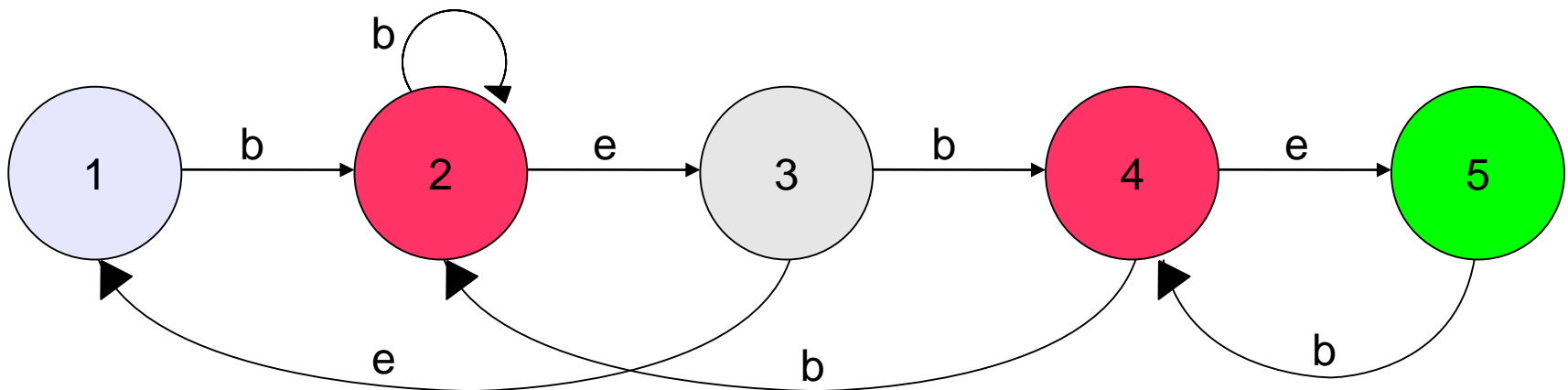
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



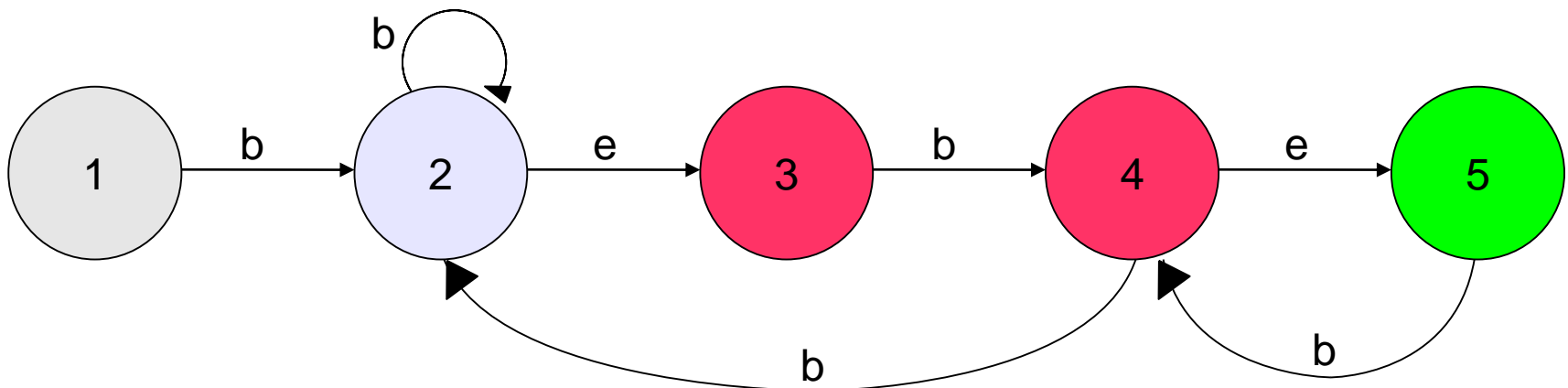
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



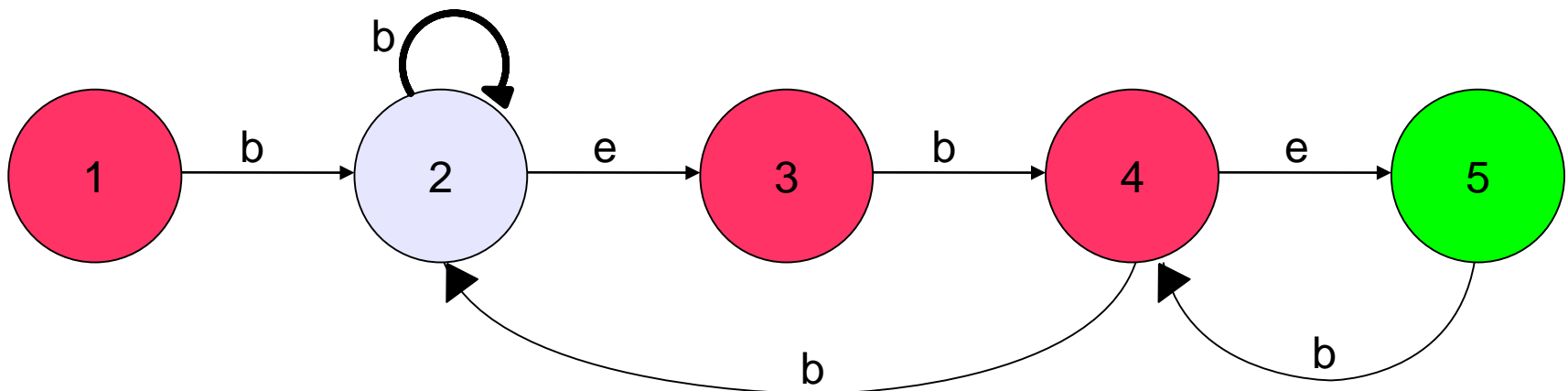
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



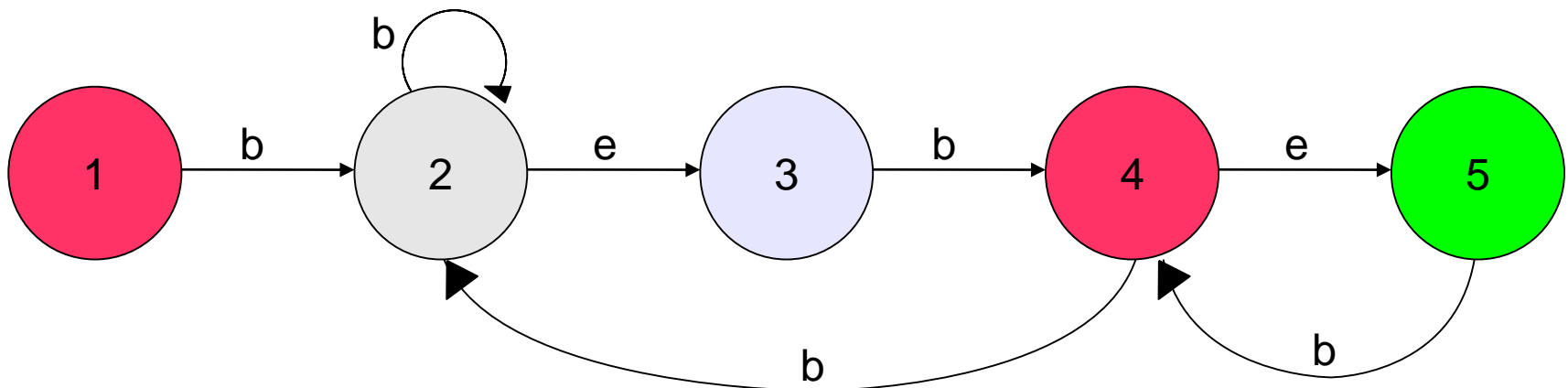
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



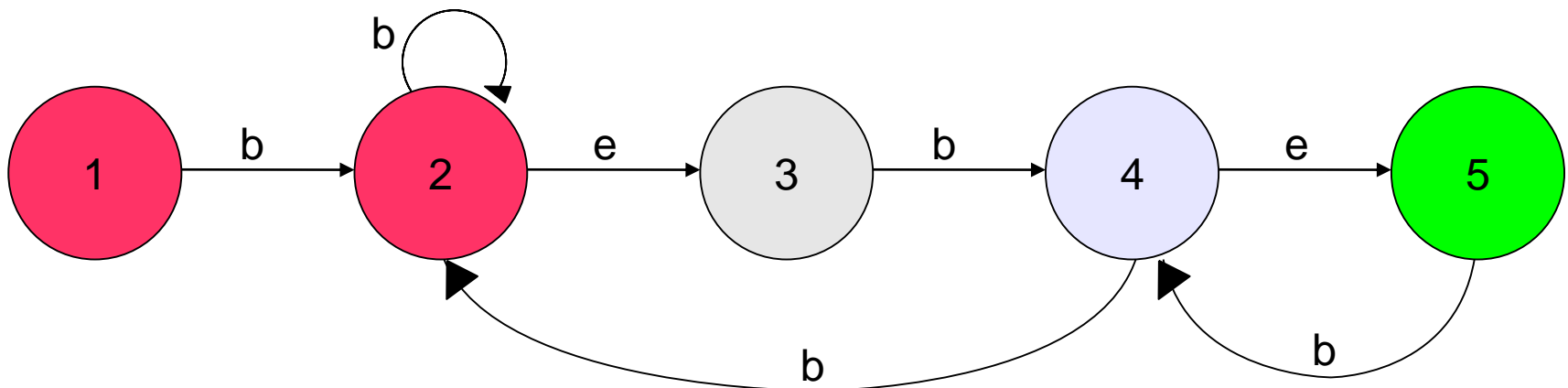
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



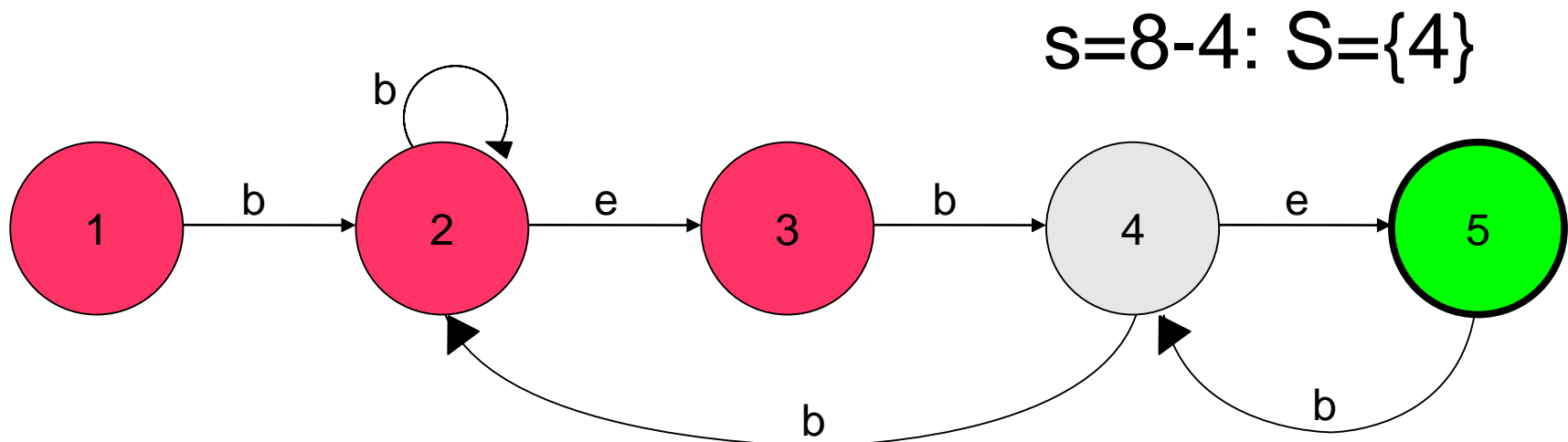
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



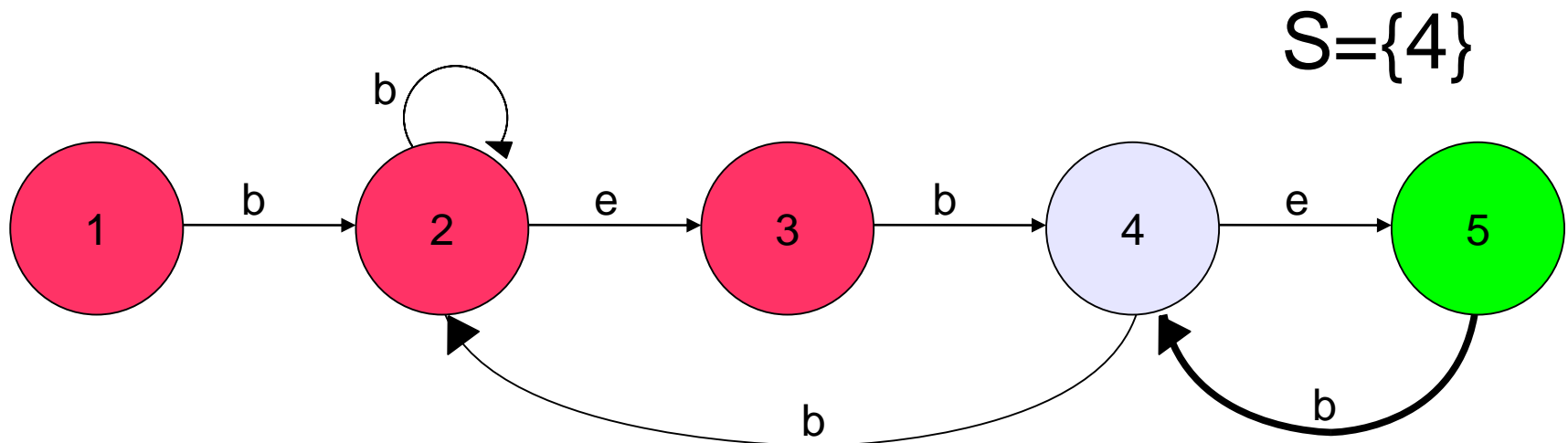
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



IV – Automate FSM

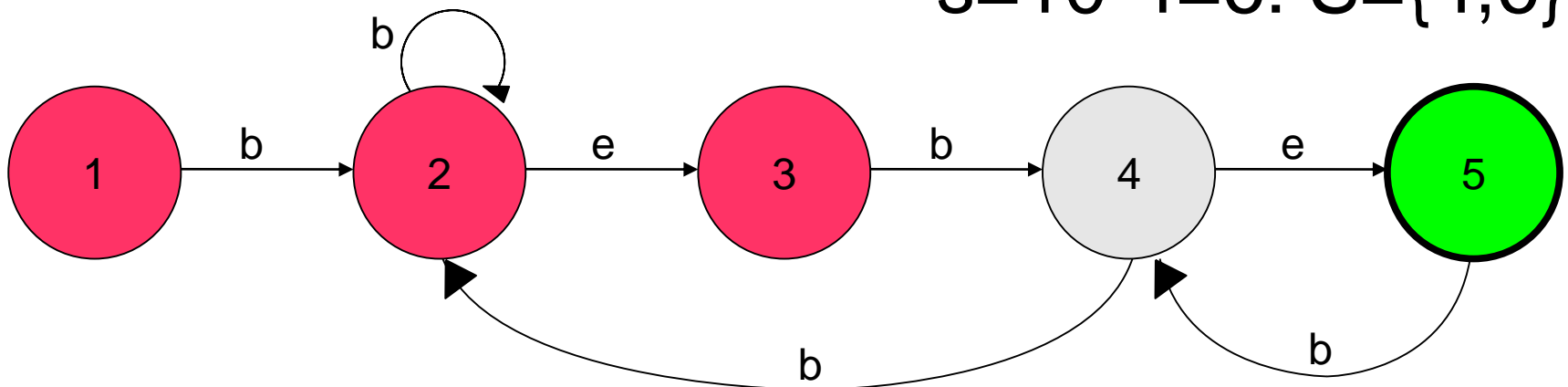
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



IV – Automate FSM

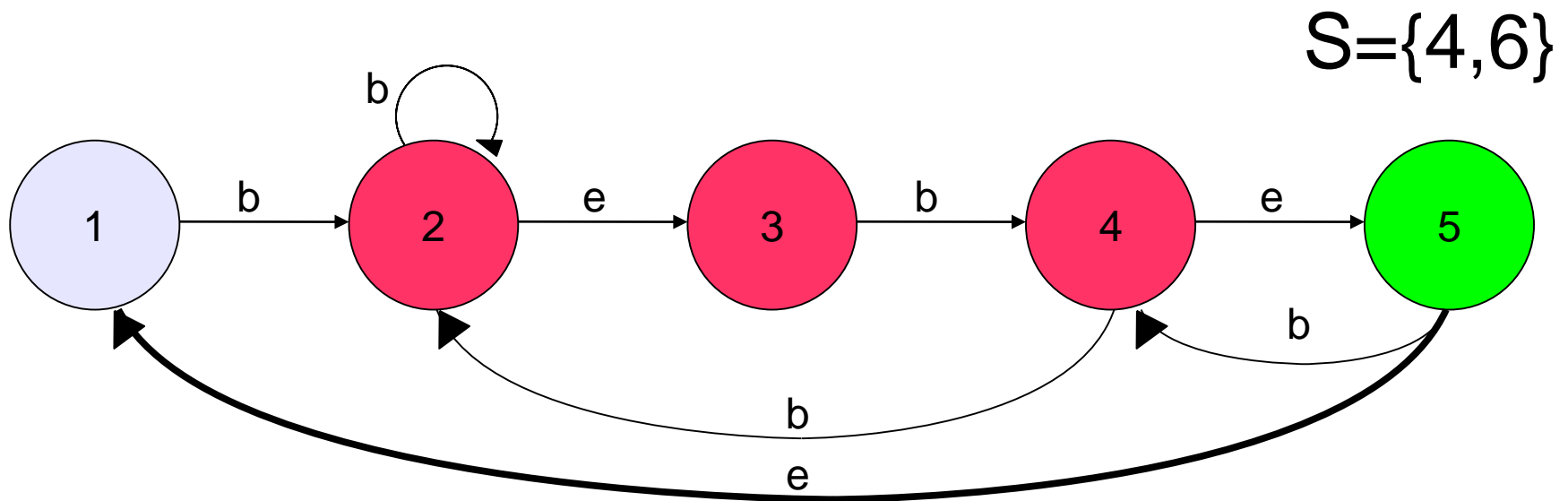
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |

$$s=10-4=6: S=\{4,6\}$$



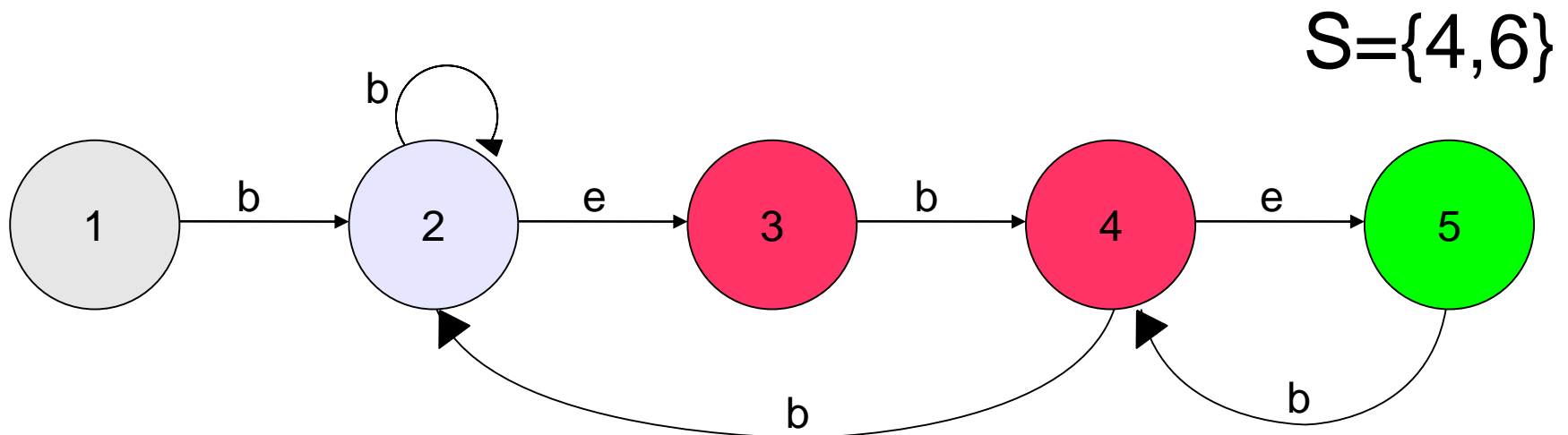
IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



IV – Automate FSM

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| b | e | e | b | b | e | b | e | b | e | e | b |



IV – Automate FSM

Quelques définitions

Ensemble des états $Q = \{q_0, q_1, q_2, \dots\}$

Les états peuvent être représentés par un entier (leur indice)

État initial: 0

État final: m

Fonction post-fixe est notée: \supset

Fonction préfixe est notée: \subset

Le préfixe de P de longueur i est noté P_i ($P_i \subset P$)

Alphabet: Σ

Alphabet réduit aux lettre du patron: Σ_p

Fonction de transition $\delta : Q \times \Sigma_p \rightarrow Q$

Exemple

Rechercher: aabab

$Q = \{q_0, q_1, q_2, q_3, q_4, q_5\}$

$Q_0 = 0$

État final : 5

$\Sigma_p = \{a, b\}$

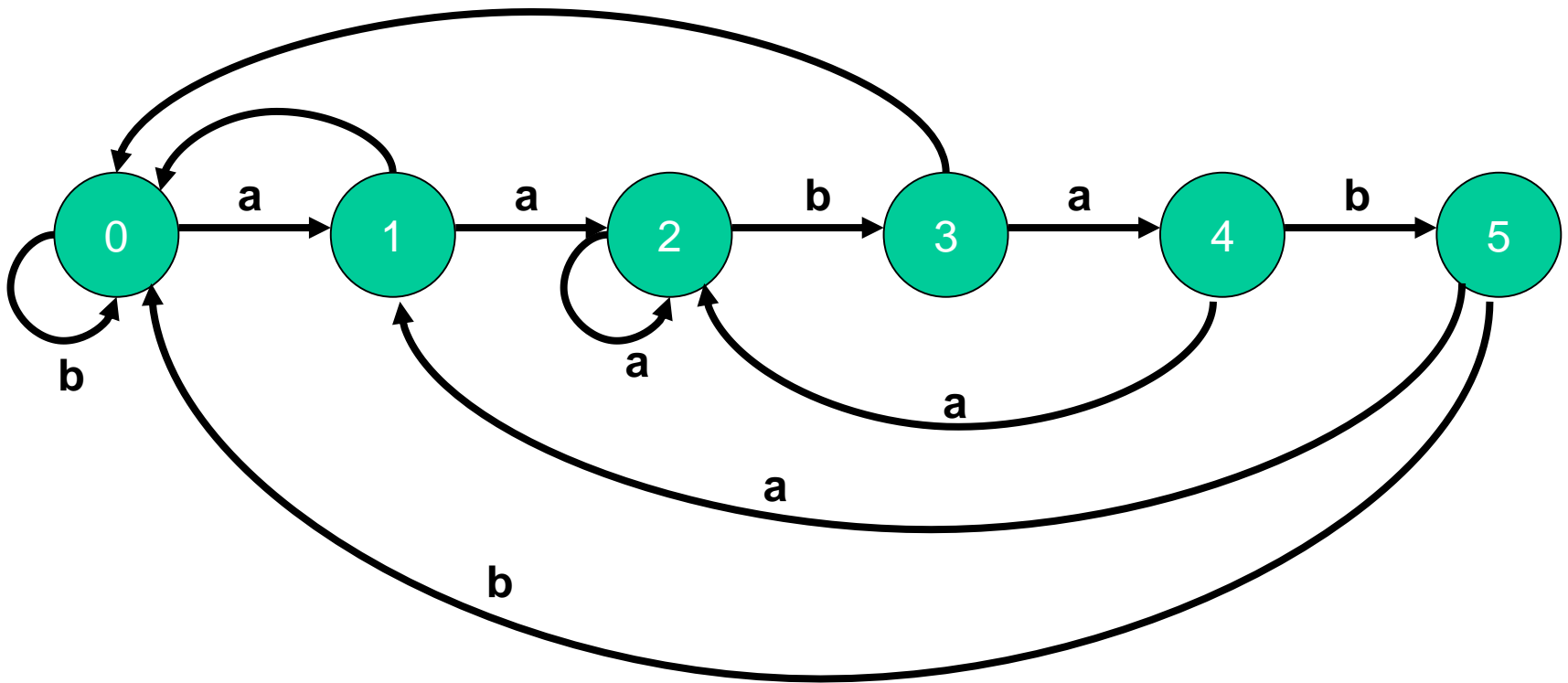
Exemple

- Fonction de transition δ :

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 5 |
| q5 | 1 | 0 |

Patron: aabab

Exemple



Construction

Idée: chaque état a reconnu le préfixe de la chaîne à rechercher qui lui est associée

$q_0 \leftrightarrow \epsilon$

$q_1 \leftrightarrow a$

$q_2 \leftrightarrow aa$

$q_3 \leftrightarrow aab$

$q_4 \leftrightarrow aaba$

$q_5 \leftrightarrow aabab$

Construction

Algorithme

```
Initialiser  $\delta$  à 0
Pour  $q = 0 : m$ 
    Pour chaque caractère  $a$  dans  $\Sigma_P$ 
         $k = \min(m+1, q+2)$ 
        Répéter
             $k = k - 1$ 
            Tant que  $k > 0$  et  $(P_k \supset P_q a)$  est faux
                 $\delta(q, a) = k$ 
        Fin Pour
Fin Pour
```

Exemple de construction

Position: 012345

Chaine: aabab

$P_0 = \varepsilon$

$P_1 = a$

$P_2 = aa$

$P_3 = aab$

$P_4 = aaba$

$P_5 = aabab$

P_i est le préfixe de P de i lettres

Exemple de construction

Algorithme

Patron: aabab

```
Initialiser  $\delta$  à 0
Pour  $q = 0 : m$ 
    Pour chaque caractère  $a$  dans  $\Sigma_P$ 
         $k = \min(m+1, q+2)$ 
        Répéter
             $k = k - 1$ 
            Tant que  $k > 0$  et (  $P_k \supset P_q a$  ) est faux)
                 $\delta(q, a) = k$ 
        Fin Pour
    Fin Pour
```

| | a | b |
|----|---|---|
| q0 | 0 | 0 |
| q1 | 0 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 0, a

1: $m \leftarrow 5$
2: $q \leftarrow 0$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 2)$
 $k \leftarrow 2$
5: $k \leftarrow 1$
6: $a \supset \varepsilon a$?
 $a \supset a$ \checkmark
7: $\delta(0, a) \leftarrow 1$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 0 | 0 |
| q1 | 0 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 0, a

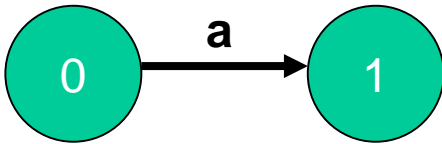
1: $m \leftarrow 5$
2: $q \leftarrow 0$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 2)$
 $k \leftarrow 2$
5: $k \leftarrow 1$
6: $a \supset \varepsilon a$?
 $a \supset a$ \checkmark
7: $\delta(0, a) \leftarrow 1$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 0 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 0, a

$$\delta(0, a) \leftarrow 1$$



Transition 0, b

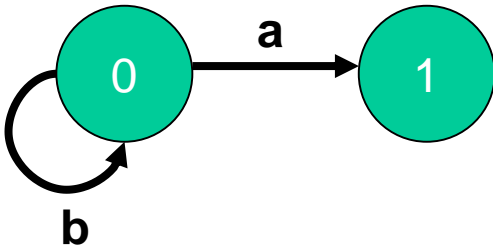
Patron: aabab

```
3: a ← 'b'
4: k ← min(6, 2)
    ← 2
5: k ← 1
6: a ⊃ εb
    a ⊃ b
    ← F
5: k ← 0
6: ε ⊃ εb
    ε ⊃ b
    ← T
7: δ(0, b) ← 0
```

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 0 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 0, b

$$\delta(0, b) \leftarrow 0$$



Transition 1, a

Patron: aabab

2: $q \leftarrow 1$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 3)$
 $\leftarrow 3$
5: $k \leftarrow 2$
6: $aa \supset aa$
 $\leftarrow T$
7: $\delta(1, a) \leftarrow 2$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 0 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 1, a

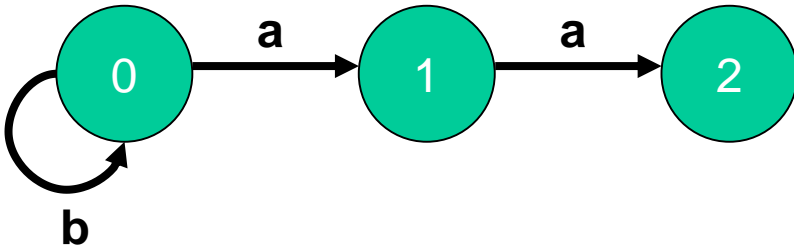
Patron: aabab

2: $q \leftarrow 1$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 3)$
 $\leftarrow 3$
5: $k \leftarrow 2$
6: $aa \supset aa$
 $\leftarrow T$
7: $\delta(1, a) \leftarrow 2$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 1, a

$$\delta(1, a) \leftarrow 2$$



Transition 1, b

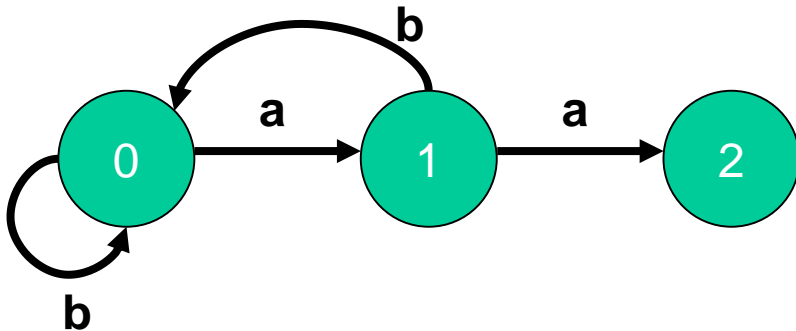
3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 3)$
 $\leftarrow 3$
5: $k \leftarrow 2$
6: $aa \supset ab$
 $\leftarrow F$
5: $k \leftarrow 1$
6: $a \supset ab$
 $\leftarrow F$
5: $k \leftarrow 0$
6: $\varepsilon \supset ab$
 $\leftarrow T$
7: $\delta(1, b) \leftarrow 0$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 1, b

$$\delta(1, b) \leftarrow 0$$



Transition 2, a

Patron: aabab

2: $q \leftarrow 2$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 4)$
 $\leftarrow 4$
5: $k \leftarrow 3$
6: $aab \supset aaa$
 $\leftarrow F$
5: $k \leftarrow 2$
6: $aa \supset aaa$
 $\leftarrow T$
7: $\delta(2, a) \leftarrow 2$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 0 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 2, a

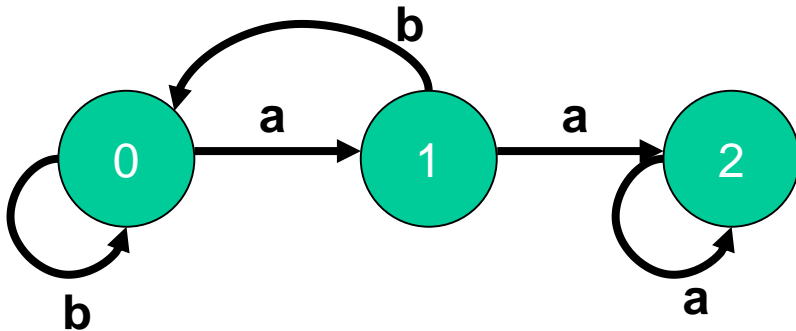
Patron: aabab

2: $q \leftarrow 2$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 4)$
 $\leftarrow 4$
5: $k \leftarrow 3$
6: $aab \supset aaa$
 $\leftarrow F$
5: $k \leftarrow 2$
6: $aa \supset aaa$
 $\leftarrow T$
7: $\delta(2, a) \leftarrow 2$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 2, a

$$\delta(2, a) \leftarrow 2$$



Transition 2, b

Patron: aabab

3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 4)$
 $\leftarrow 4$
5: $k \leftarrow 3$
6: $aab \supset aab$
 $\leftarrow T$
7: $\delta(2, b) \leftarrow 3$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 0 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 2, b

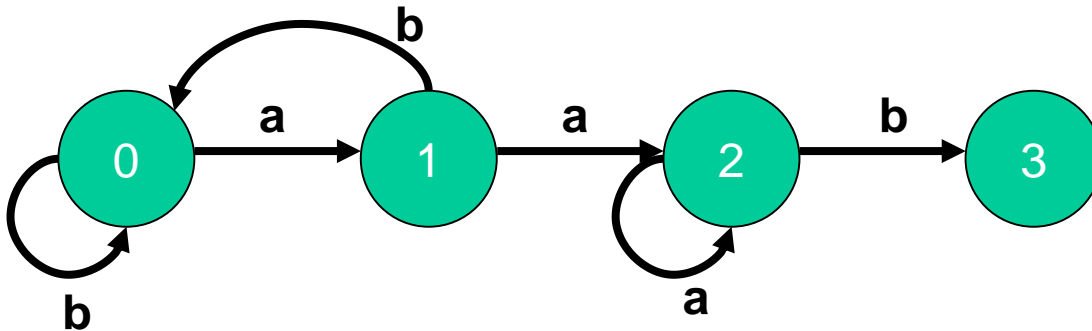
Patron: aabab

3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 4)$
 $\leftarrow 4$
5: $k \leftarrow 3$
6: $aab \supset aab$
 $\leftarrow T$
7: $\delta(2, b) \leftarrow 3$

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 2, b

$$\delta(2, b) \leftarrow 3$$



Transition 3, a

2: $q \leftarrow 3$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 5)$
 $\leftarrow 5$
5: $k \leftarrow 4$
6: $aaba \supset aaba$
 $\leftarrow T$
7: $\delta(3, a) \leftarrow 4$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 0 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 3, a

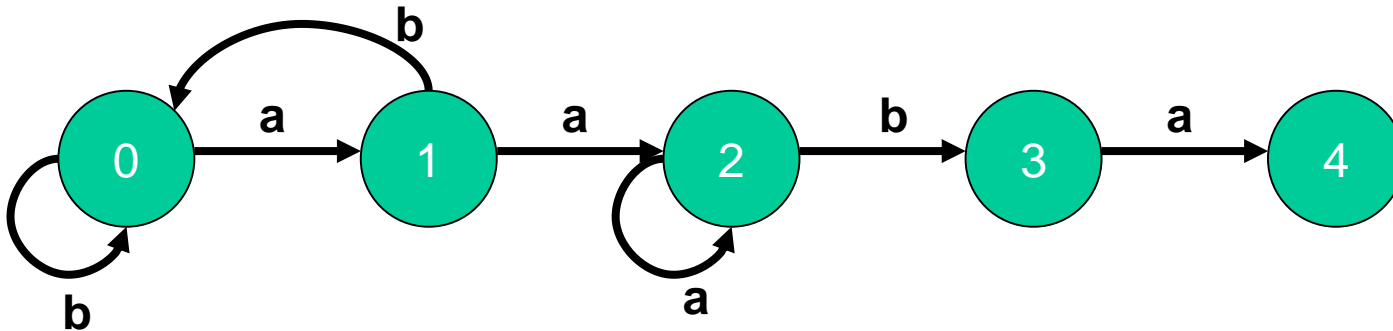
2: $q \leftarrow 3$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 5)$
 $\leftarrow 5$
5: $k \leftarrow 4$
6: $aaba \supset aaba$
 $\leftarrow T$
7: $\delta(3, a) \leftarrow 4$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 3, a

$$\delta(3, a) \leftarrow 4$$



Transition 3, b

3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 5)$
 $\leftarrow 5$
5: $k \leftarrow 4$
6: $aaba \supset aabb$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aabb$
 $\leftarrow F$
5: $k \leftarrow 2$
6: $aa \supset aabb$
 $\leftarrow F$

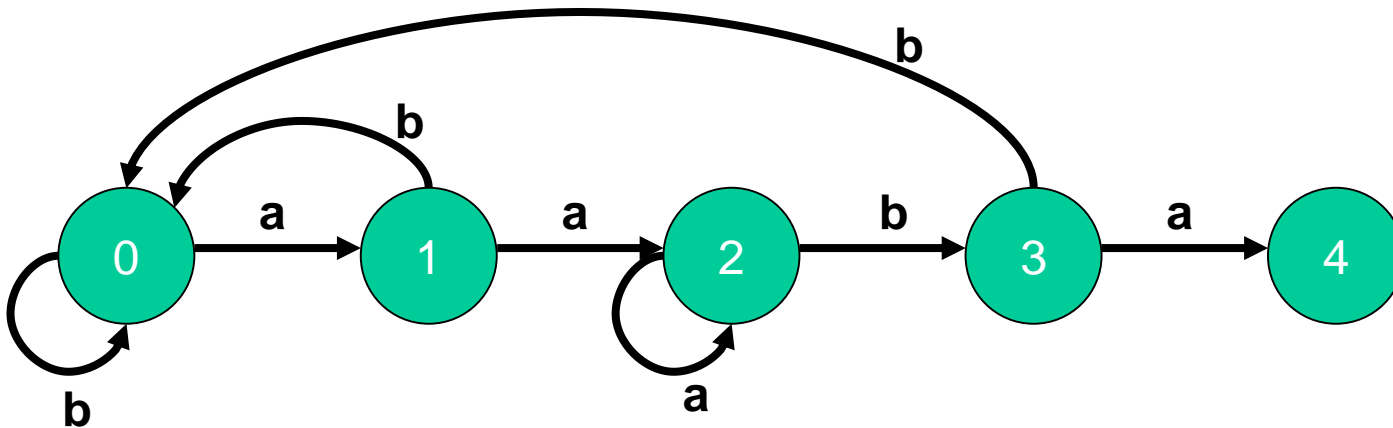
5: $k \leftarrow 1$
6: $a \supset aabb$
 $\leftarrow F$
5: $k \leftarrow 0$
6: $\varepsilon \supset aabb$
 $\leftarrow T$
7: $\delta(3, b) \leftarrow 0$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 3, b

$$\delta(3, b) \leftarrow 0$$



Transition 4, a

2: $q \leftarrow 4$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 6)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $aabab \supset aabaa$
 $\leftarrow F$
5: $k \leftarrow 4$
6: $aaba \supset aabaa$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aabaa$
 $\leftarrow F$

5: $k \leftarrow 2$
6: $aa \supset aabaa$
 $\leftarrow T$
7: $\delta(4, a) \leftarrow 2$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 0 | 0 |
| q5 | 0 | 0 |

Transition 4, a

2: $q \leftarrow 4$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 6)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $aabab \supset aabaa$
 $\leftarrow F$
5: $k \leftarrow 4$
6: $aaba \supset aabaa$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aabaa$
 $\leftarrow F$

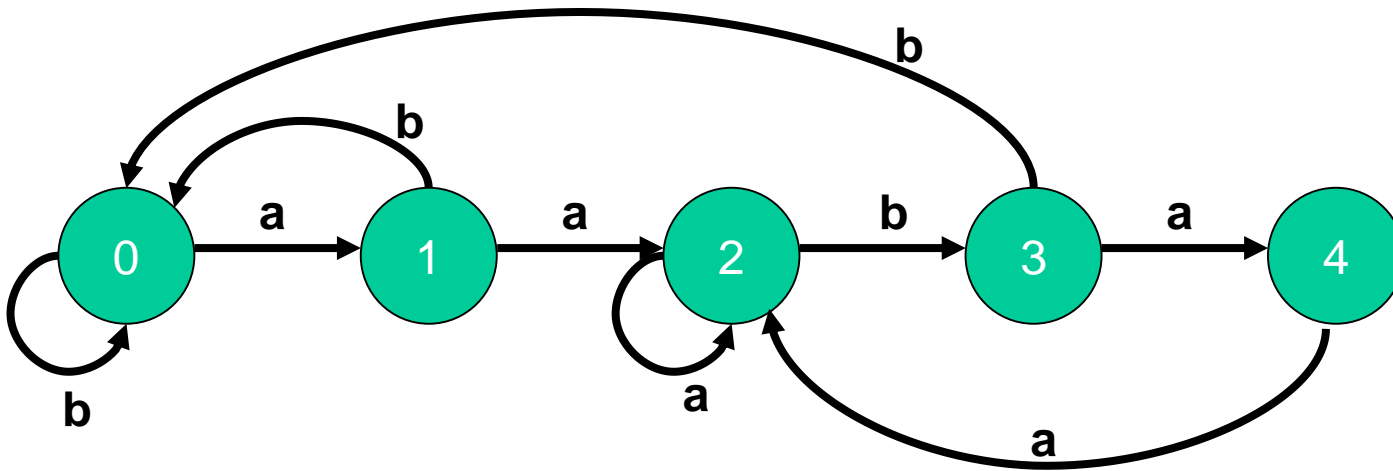
5: $k \leftarrow 2$
6: $aa \supset aabaa$
 $\leftarrow T$
7: $\delta(4, a) \leftarrow 2$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 0 |
| q5 | 0 | 0 |

Transition 4, a

$$\delta(4, a) \leftarrow 2$$



Transition 4, b

3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 6)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $\text{aabab} \supset \text{aabab}$
 $\leftarrow T$
7: $\delta(4, b) \leftarrow 5$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 0 |
| q5 | 0 | 0 |

Transition 4, b

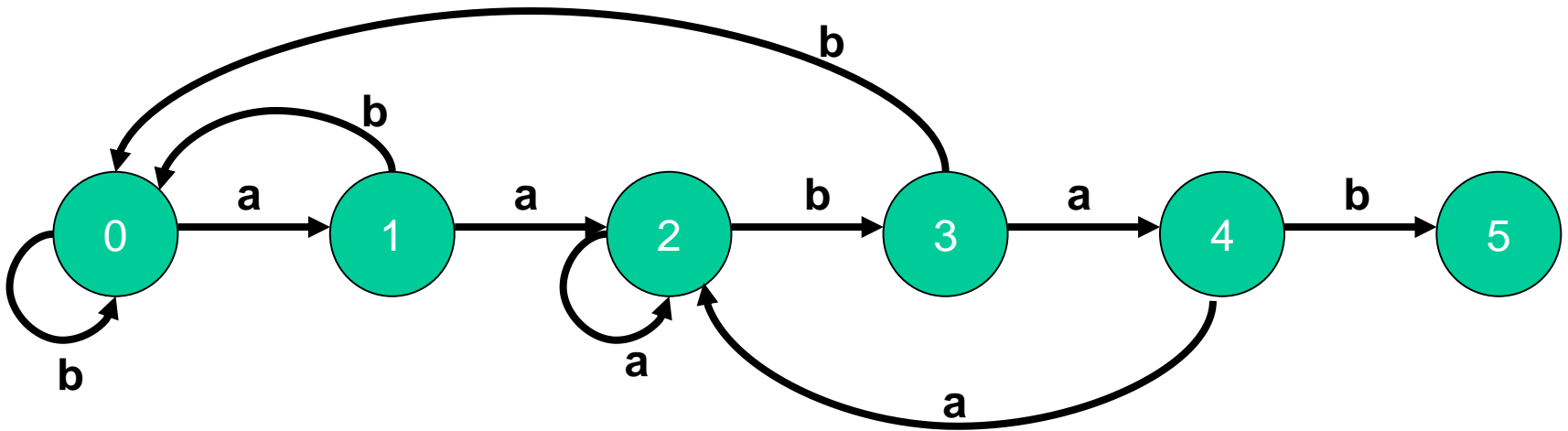
3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 6)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $\text{aabab} \supset \text{aabab}$
 $\leftarrow T$
7: $\delta(4, b) \leftarrow 5$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 5 |
| q5 | 0 | 0 |

Transition 4, b

$$\delta(4, b) \leftarrow 5$$



Transition 5, a

2: $q \leftarrow 5$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 7)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $aabab \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 4$
6: $aaba \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aababa$
 $\leftarrow F$

5: $k \leftarrow 2$
6: $aa \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 1$
6: $a \supset aababa$
 $\leftarrow T$
7: $\delta(5, a) \leftarrow 1$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 5 |
| q5 | 0 | 0 |

Transition 5, a

2: $q \leftarrow 5$
3: $a \leftarrow 'a'$
4: $k \leftarrow \min(6, 7)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $aabab \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 4$
6: $aaba \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aababa$
 $\leftarrow F$

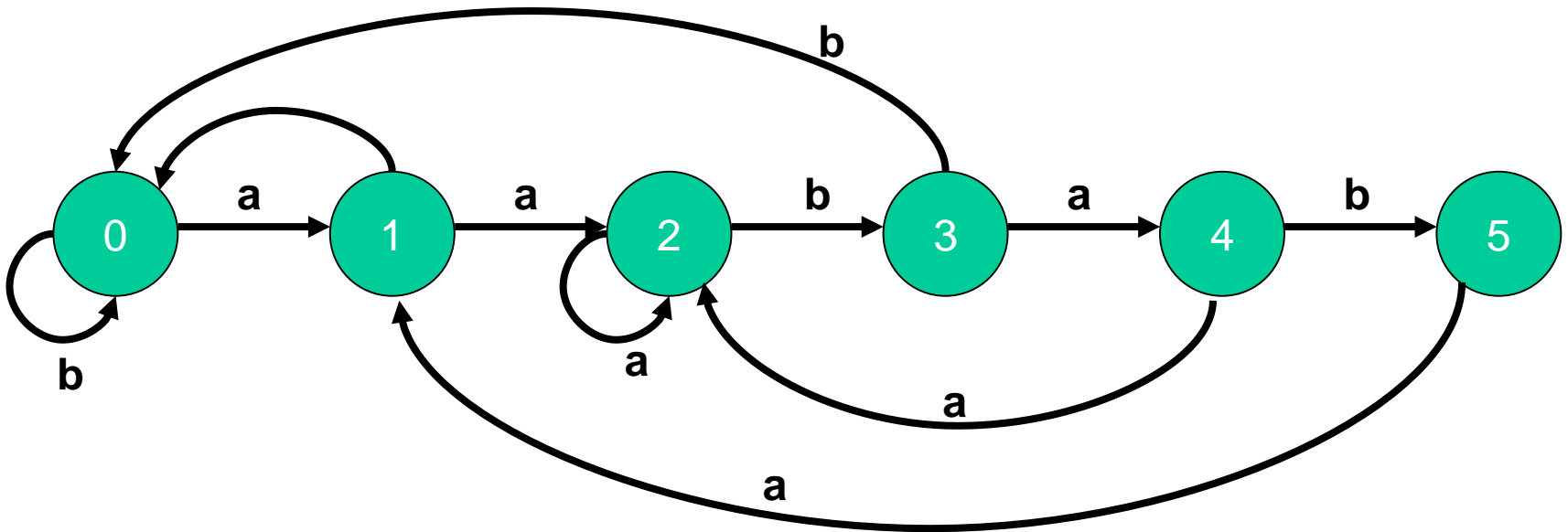
5: $k \leftarrow 2$
6: $aa \supset aababa$
 $\leftarrow F$
5: $k \leftarrow 1$
6: $a \supset aababa$
 $\leftarrow T$
7: $\delta(5, a) \leftarrow 1$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 5 |
| q5 | 1 | 0 |

Transition 5, a

$$\delta(5, a) \leftarrow 1$$



Transition 5, b

3: $a \leftarrow 'b'$
4: $k \leftarrow \min(6, 7)$
 $\leftarrow 6$
5: $k \leftarrow 5$
6: $aabab \supset aababb$
 $\leftarrow F$
5: $k \leftarrow 4$
6: $aaba \supset aababb$
 $\leftarrow F$
5: $k \leftarrow 3$
6: $aab \supset aababb$
 $\leftarrow F$

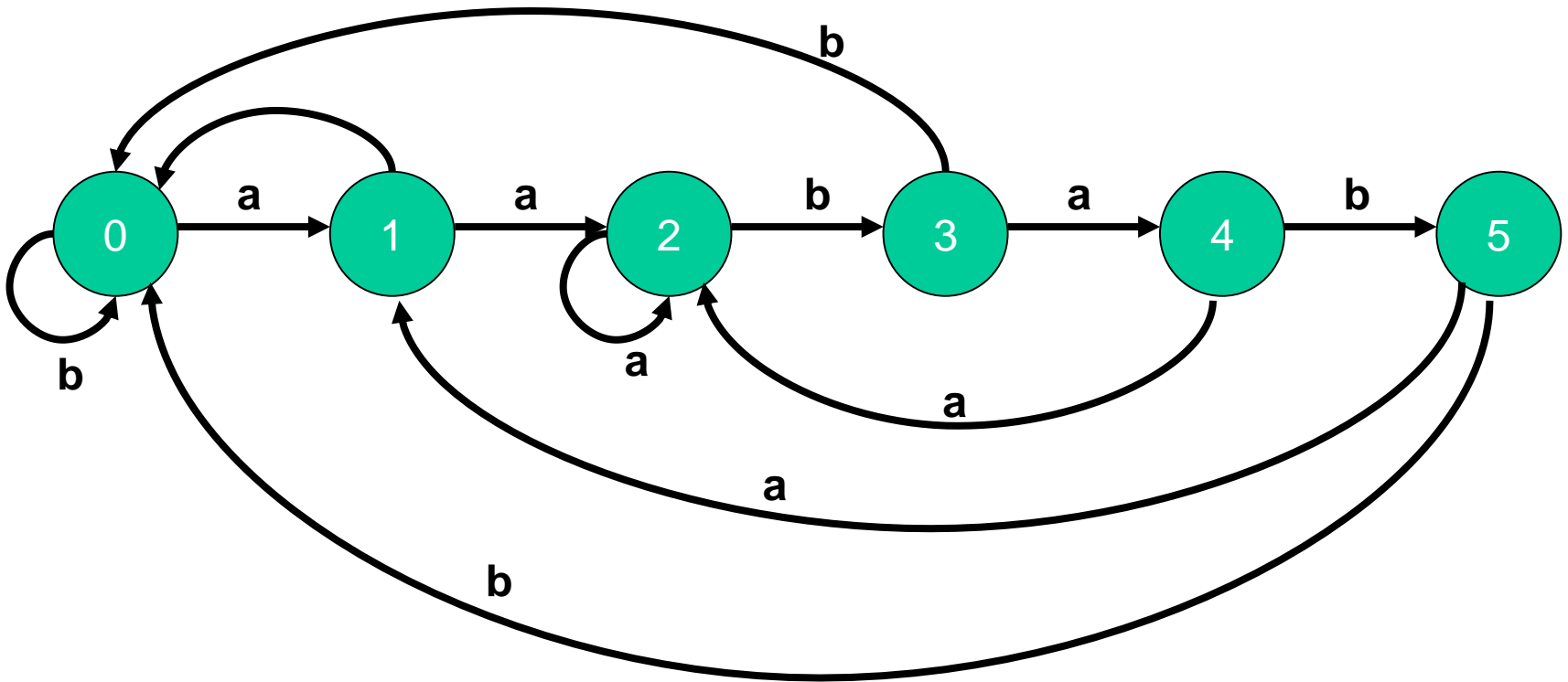
5: $k \leftarrow 2$
6: $aa \supset aababb$
 $\leftarrow F$
5: $k \leftarrow 1$
6: $a \supset aababb$
 $\leftarrow F$
5: $k \leftarrow 0$
6: $\varepsilon \supset aababb$
 $\leftarrow T$
7: $\delta(5, b) \leftarrow 0$

Patron: aabab

| | a | b |
|----|---|---|
| q0 | 1 | 0 |
| q1 | 2 | 0 |
| q2 | 2 | 3 |
| q3 | 4 | 0 |
| q4 | 2 | 5 |
| q5 | 1 | 0 |

Transition 5, b

$$\delta(5, b) \leftarrow 0$$



IV – Automate FSM

Dans le meilleur des cas, l'automate FSM donne une complexité

$$O(n) = \Theta(n)$$

La construction de la machine à états peut être coûteuse: $O(m^3d)$. Ce qui peut devenir handicapant pour certains problèmes de recherche:

$$O(m^3d) + \Theta(n) \text{ vs. } O(m(n-m+1) + m)$$