

Algorithmes de recherche - TD

12 décembre 2018

Table des matières

1	Intros et définitions	2
1.1	Périodicité et bords	2
1.2	Puissance et primitive	3
2	Mots particuliers	4
2.1	Mots de fibonacci	4
2.2	Les mots de DE BRUIJN	5
3	Automates de localisation	6
3.1	Arbre d'un dictionnaire	6
3.2	Localisation de plusieurs mots	6
3.3	Construction de l'automate-dictionnaire	7
3.4	Automate avec fonction de suppleance	8
3.5	Automate avec optimisation de la fonction de suppleance	8
4	Alignements de mots	9
4.1	Comparaison de mots	9
4.2	Alignements	9
4.3	Alignement optimal	10

Table des figures

1	Automate DE BRUIJN ordre 3 avec $A = \{a,b\}$	5
2	Automate $A(X)$ reconnaissant $X = \{aa, abaaa, abab\}$	6
3	Automate $D(X)$ reconnaissant $X = \{aa, abaaa, abab\}$	6
4	Automate $D(X)$ reconnaissant $X = \{ab, babb, bb\}$ sur $A = \{a, b, c\}$. . .	7
5	Automate $DS(X)$ reconnaissant $X = aa, abaaa, abab$ sur $A = \{a, b\}$. . .	8

1 Intros et définitions

Recherche de mots dans un texte, algorithme exact, toutes les solutions et pas d'heuristique.

1.1 Périodicité et bords

Quelles sont les périodes du mot $x = aabaabaa$ et $\text{pér}(x)$ de x ?

$$p = 3 \Rightarrow \forall 0 \leq i \leq 8 - 3 - 1 \Rightarrow i \in [0, 4]$$

$$p = 6 \Rightarrow \forall 0 \leq i \leq 8 - 6 - 1 \Rightarrow i \in [0, 1]$$

$$p = 7 \Rightarrow \forall 0 \leq i \leq 8 - 7 - 1 \Rightarrow i \in [0, 0]$$

$$p = 8 \Rightarrow \text{Admis}$$

$$\text{pér}(x) = 3$$

Quelles sont les bords du mot $x = aabaabaa$?

$\Rightarrow \epsilon, a, aa$ et $aabaa$.

Quelle est la relation entre les notions de bords et périodes ?

Les notions de bords et de périodes sont duales.

Application avec la bord aa .

Le bord correspond à la période 6, qui correspond à $|aabaabaa| - |aa|$.

Quel est le bord du mot $aabaabaa$?

C'est $aabaa$.

Quelles sont les suites des bords et des périodes du mot $x = aabaabaa$?

Suite des bords : $(aabaa, aa, a, \epsilon)$, car LE bord de $aabaabaa$ est $aabaa$ (le bord le plus long)

Suite des périodes : $(3, 6, 7, 8)$, car

$$\begin{aligned} 3 &= |x| - |\text{Bord}(x)| \\ &= |aabaabaa| - |aabaa| \\ &= 8 - 5 \end{aligned}$$

ET

$$\begin{aligned} 6 &= |x| - |\text{Bord}^2(x)| \\ &= |aabaabaa| - |aa| \\ &= 8 - 2 \end{aligned}$$

1.2 Puissance et primitive

Donner un exemple de mots x et y vérifiant $x^m = y^n$ sur un alphabet à 2 lettres ?
Soit $x = ab$ et $y = abab \Rightarrow x^2 = y^1$

Donner un exemple de mot primitif et de non-primitif de longueurs supérieures à 3.
Le mot $abaab$ est primitif.
Le mot $baba = (ba)^2$ n'est pas primitif.

Vérifier la proposition 'Un mot non vide est primitif si et seulement s'il n'a pas un facteur de son carré qu'en tant que préfixe et suffixe', avec les mots $abaab$ et $baba$.
Le mot $abaab$ est primitif car $abaab$ est uniquement un facteur de $abaab$. $abaab$ en tant que préfixe et suffixe.
Le mot $baba$ n'est pas primitif car $baba$ n'est pas uniquement un facteur de $baba$. $baba$ en tant que préfixe et suffixe, par exemple ba . $baba$. ba .

Donner un exemple de mots conjugués.
Avec $A = a, b, x = abbaba, y = abaabb$, on a $u = abb$ et $v = aba$.
Avec $B = A, C, G, T, x = AGTACGTTA, y = ACGTTAAGT$, on a $u = AGT$ et $v = ACGTTA$.
Avec $B^3 = AAA, AAC, \dots, TTT, x = TTTACG, y = ACGTTT$, on a $u = TTT$, $v = ACG$.

Donner un exemple de mot z avec des mots conjugués x et y .
Soit $x = AAC, y = ACA$, on a $z = A$ car $x = z \cdot AC$ et $y = AC \cdot z$

2 Mots particuliers

2.1 Mots de fibonacci

n	F_n	f_n
3	2	ab
4	3	aba
5	5	abaab
6	8	abaababa
7	13	abaababaabaab
8	21	abaababaabaabaabaaba

Propriété remarquable : ???

Démontrer $\phi^n(a) = f_{n+2}$.

$$\begin{aligned}
 \phi^1(a) &= ab = f_3 \\
 \phi^2(a) &= \phi^1(\phi(a)) \\
 &= \phi^1(ab) \\
 &= \phi(a) \cdot \phi(b) \\
 &= ab \cdot a = aba \\
 &= f_4 \\
 \phi^n(a) &= \phi^{n-1}(\phi(a)) \\
 &= \phi^{n-1}(ab) \\
 &= \phi^{n-1}(a) \cdot \phi^{n-1}(b) \\
 &= f_{n+1} \cdot \phi^{n-2}(\phi(b)) \\
 &= f_{n+1} \cdot \phi^{n-2}(a) \\
 &= f_{n+1} \cdot f_n \\
 &= f_{n+2}
 \end{aligned}$$

Démontrer la proposition du palindrome de fibonacci avec $n \geq 3$.

$$\begin{aligned}
 f_3 &= ab, u = \epsilon \Rightarrow \text{palindrome}(\text{trivial}) \\
 f_4 &= aba, u = a \Rightarrow \text{palindrome}(\text{trivial}) \\
 f_5 &= abaab, u = aba \Rightarrow \text{palindromecar}
 \end{aligned}$$

Pour tout $n \geq 5$, on a $f_n = f_{n-1} \cdot f_{n-2} = f_{n-2} \cdot f_{n-3} \cdot f_{n-2}$.

Si n est impair alors, par hypothèse de récurrence alors f_{n-2} est impair et f_{n-3} est pair. Donc $f_{n-2} = u_1 ab$ et $f_{n-3} = u_2 ba$ avec u_1 et u_2 des palindromes.

Donc, $f_n = u_1 ab u_2 ba u_1 ab$.

Mais u_1 et u_2 sont des palindromes, donc $u_1 ab u_2 ba u_1$ est également un palindrome.

En posant $u = u_1 ab u_2 ba u_1$, on a alors $f_n = uab$ où u est un palindrome.

Démonstration de façon similaire si n est pair.

2.2 Les mots de DE BRUIJN

Donner tous les mots de DE BRUIJN d'ordre $k = 1$ dans l'alphabet $A = a, b$.
Les mots ab et ba sont les deux seuls mots de DE BRUIJN d'ordre 1.

Pour $k = 3$:

$A^3 = aaa, aab, \dots, bbb$

Introuvables à la main car trop compliqués. Ex : $aaababbbbaa$ car ses facteurs de longueur 3 sont les 8 mots de A^3 : $aaa, aab, aba, abb, baa, bab, bba, bbb$. Chaque mot n'apparaît qu'une et une seule fois dans ce mot ET tous les mots de A^3 y apparaissent.

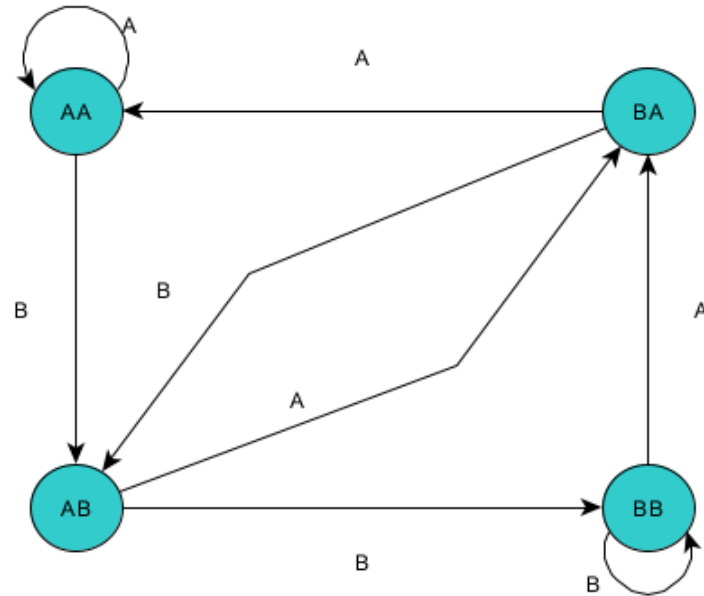


FIGURE 1 – Automate DE BRUIJN ordre 3 avec $A = \{a, b\}$

Exactement 2 flèches sortent de chacun des états, l'une étiquetée par a , l'autre par b , et exactement 2 flèches entrent dans chacun des états, toutes 2 étiquetées par la même lettre. Il faut passer par toutes les flèches en commençant et terminant par le même état.

Longueur d'un mot de DE BRUIJN d'ordre $k = f(k)$?
 $2^k + (k - 1)$

3 Automates de localisation

3.1 Arbre d'un dictionnaire

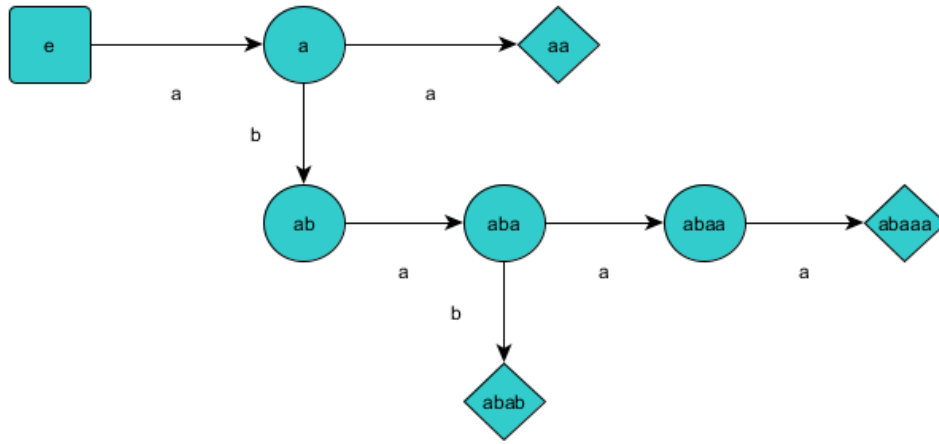


FIGURE 2 – Automate $A(X)$ reconnaissant $X = \{aa, abaaa, abab\}$

3.2 Localisation de plusieurs mots

$h(b) = \epsilon \Rightarrow$ Tout ce qui se finit par b va vers ϵ
 $h(a) = a$
 $h(aa) = aa$
 $h(ab) = ab$
 $h(aaa) = aa$
 $h(aab) = ab$
 $h(aba) = aba$
 $h(abb) = \epsilon$

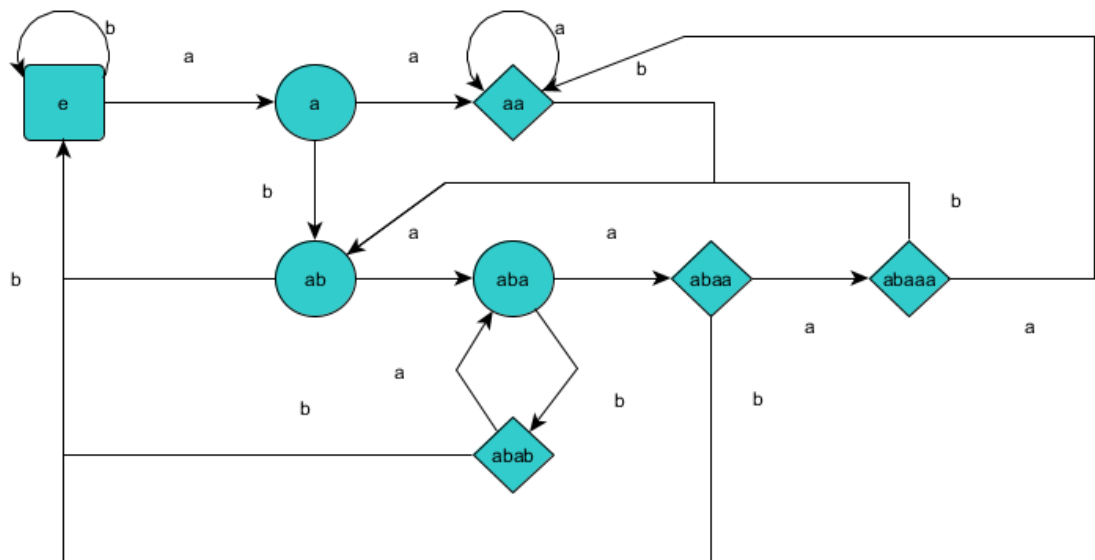


FIGURE 3 – Automate $D(X)$ reconnaissant $X = \{aa, abaaa, abab\}$

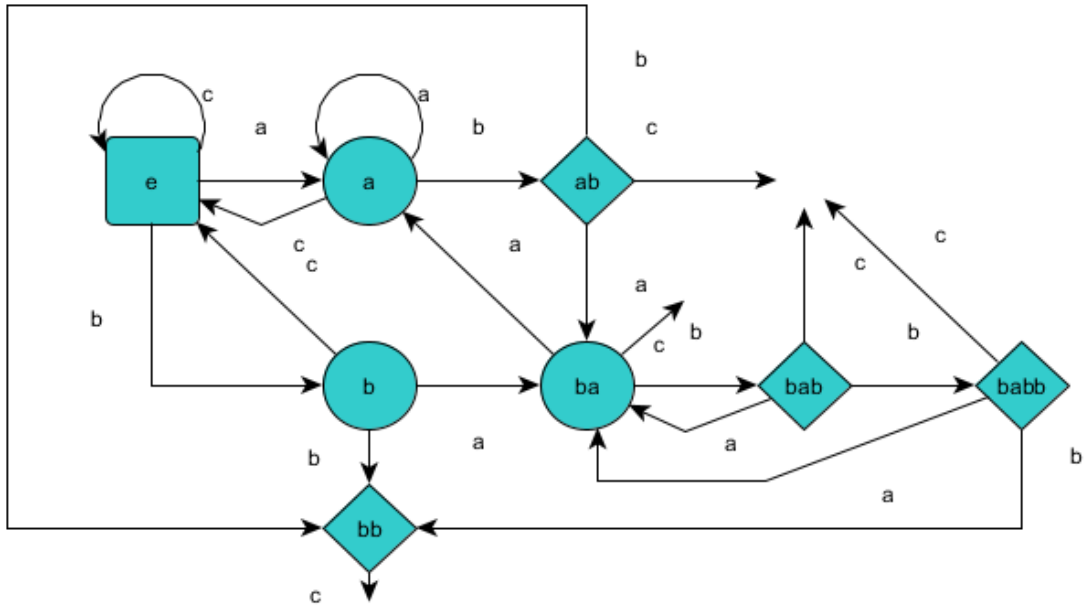


FIGURE 4 – Automate $D(X)$ reconnaissant $X = \{ab, babb, bb\}$ sur $A = \{a, b, c\}$

Sur ce dernier automate, les flèches étiquetées 'c' vont toutes vers l'état ϵ .

3.3 Construction de l'automate-dictionnaire

Localisation des mots $X = ab, babb, bb$ sur $A = a, b, c$ sur le texte $y = cbabba$. Utilisation de l'algorithme LOCALISATION(X, y) sur l'automate $D(X)$ fait précédemment.

j	$y[j]$	Etat r	Mot
		ϵ	
0	c	ϵ	
1	b	b	
2	a	ba	
3	b	bab	Occurence de ab
4	b	babb	Occurence de babb et bb
5	a	ba	

La colonne Mot contient la liste des mots reconnus, si l'état r appartient aux états terminaux de $D(X)$.

3.4 Automate avec fonction de suppleance

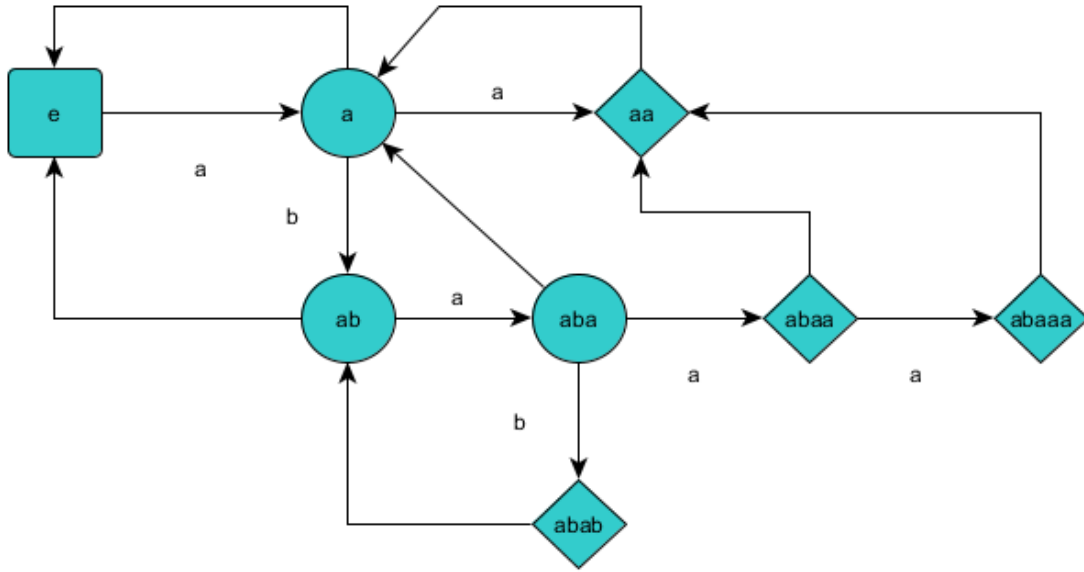


FIGURE 5 – Automate DS(X) reconnaissant $X = aa, abaaa, abab$ sur $A = \{a, b\}$

3.5 Automate avec optimisation de la fonction de suppleance

Avec l'automate DS(X) (page 8), donner des exemples d'ensembles $Suivant(u)$ avec $u = abaa$ (état abaa).

$$\begin{aligned}
 Suivant(u) &= \{a\} \\
 f(u) &= aa \\
 Suivant(f(u)) &= Suivant(aa) = \emptyset \\
 &\Rightarrow \text{car } \{a\} \Leftrightarrow \{a\} \cup \emptyset \\
 f(f(u)) &= f(aa) = a \\
 Suivant(f(f(u))) &= Suivant(a) = \{a, b\} \\
 &\Rightarrow \text{fin car } Suivant(a) \notin Suivant(aa) \\
 &\Rightarrow g : \text{état abaa} \rightarrow \text{état a}
 \end{aligned}$$

Exemple avec $u = aba$ (état 4).

$$\begin{aligned}
 Suivant(u) &= \{a, b\} \\
 f(u) &= a \\
 Suivant(a) &= \{a, b\} \\
 f(f(u)) &= \epsilon \\
 Suivant(\epsilon) &= \{a\} \\
 &\Rightarrow g : \text{état aba} \rightarrow \text{état } \epsilon
 \end{aligned}$$

4 Alignements de mots

4.1 Comparaison de mots

Démontrer que la distance préfixe est bien une distance.

Positivité : $d_{prf}(u, v) \geq 0$ car $lpc(u, v) \leq \min(|u|, |v|)$

Séparation : Si $u = v$ alors $lpc(u, v) = |u| = |v|$ et $d_{prf}(u, v) = 0$ et réciproquement.

Symétrie : $d_{prf}(u, v) = d_{prf}(v, u)$: évident

Inégalité triangulaire : $d_{prf}(u, v) \leq d_{prf}(u, w) + d_{prf}(w, v)$ pour tout $w \in A^*$
car

$$\begin{aligned} d_{prf}(u, w) + d_{prf}(w, v) &= |u| + |w| - 2 * lpc(u, w) + |v| + |w| - 2 * lpc(v, w) \\ &= |u| + |v| - 2 * \min(|lpc(u, w)|, |lpc(v, w)|) \\ &\quad + 2 * |w| - 2 * \max(|lpc(u, w)|, |lpc(v, w)|) \\ &\geq |u| + |v| - 2 * |lpc(u, v)| + 2(|w| - \max(|lpc(u, w)|, |lpc(v, w)|)) \geq |u| + |v| - 2 * |lpc(u, v)| \end{aligned}$$

Donner la suite des opérations élémentaires pour passer de $x = ACGA$ à $y = ATGCTA$ en supposant pour toutes lettres $a, b \in A$ les coûts suivants :

- $Sub(a, a) = 0$ et $Sub(a, b) = 1$ pour $a \neq b$
- $D(a) = Ins(a) = 1$

Opération	Mot résultant	Coût
Mot initial ACGA	ACGA	X
Sub(A,A)	A CGA	0
Sub(C,T)	AT G A	1
Sub(G,G)	AT G A	0
Ins(C)	ATG C A	1
Ins(T)	ATGCT A	1
Sub(A,A)	ATGCT A	0
Mot final	ATGCTA	X

4.2 Alignements

L'alignement global c'est un alignement qui cherche à aligner, pour deux mots, la première lettre de x avec la première de y, et la dernière de x avec la dernière de y.

Attention x est en haut et y est bas!!!

Donner les paires alignées et leurs coûts pour passer de $x = ACGA$ à $y = ATGCTA$ Avec les mêmes coûts que précédemment.

Opération	Paire alignée	Coût
Sub(A,A)	(A,A)	0
Sub(C,T)	(C,T)	1
Sub(G,G)	(G,G)	0
Ins(C)	(-,C)	1
Ins(T)	(-,T)	1
Sub(A,A)	(A,A)	0

$$\begin{pmatrix} A & C & G & - & - & A \\ A & T & G & C & T & A \end{pmatrix}$$

Donner le graphe d'édition $G(\text{ACGA}, \text{ATGCTA})$ de $x = \text{ACGA}$ à $y = \text{ATGCTA}$.
Donner l'alignement optimal.

G	y	A	T	G	C	T	A
x	(-1,-1)						
A		(0,0)					
C			(1,1)				
G				(2,2)			
T					(2,3)		
						(2,4)	
							(3,5)

Normalement, il faut tout compléter mais j'ai la flemme. Dans le tableau :

- Délétion \rightarrow case verticale
- Insertion \rightarrow case à droite
- Substitution \rightarrow case en diagonale

4.3 Alignement optimal

Calcul de la distance d'édition entre les 2 mots $x = \text{ACGA}$ et $y = \text{ATGCTA}$ sachant :

- Matrice de substitution Sub avec $\text{Sub}(a,b) = 1$ ($a \neq b$) et $\text{Sub}(a,a) = 0$
- Vecteurs de délétion et d'insertion avec $\text{Dél}(a) = \text{Ins}(a) = 1$

T	j	-1	0	1	2	3	4	5
i	/	y[j]	A	T	G	C	T	A
-1	x[i]	0	1	2	3	4	5	6
0	A	1	0	1	2	3	4	5
1	C	2	1	1	2	2	3	4
2	G	3	2	2	1	2	3	4
3	A	4	3	3	2	2	3	3