

Solutionnaire du travail pratique #2

Réponses

Note: les réponses aux différentes questions ne sont pas toujours uniques.

1. (a)

Texte	Triplet
caabbcaabcba	$\langle 0, 0, c \rangle$
c aabbcaabcba	$\langle 0, 0, a \rangle$
ca abbcaabcba	$\langle 1, 1, b \rangle$
caab bcaabcba	$\langle 1, 1, c \rangle$
caabbc aabcba	$\langle 5, 3, c \rangle$
caabbcaabc ba	$\langle 2, 1, a \rangle$
caabbcaabcba	

(b)

Texte	Message
caabbcaabcba	$\langle c \rangle$
c aabbcaabcba	$\langle a \rangle$
ca abbcaabcba	$\langle a \rangle$
caab bcaabcba	$\langle b \rangle$
caab bcaabcba	$\langle b \rangle$
caabbc aabcba	$\langle 5, 4 \rangle$
caabbcaab cba	$\langle c \rangle$
caabbcaabc ba	$\langle b \rangle$
caabbcaabc ba	$\langle a \rangle$
caabbcaabcba	

- (c) Ici, l'état du texte à compresser, les couples et les entrées du dictionnaire sont donnés de manière compacte. Les étapes de traitement commencent au #1. Lors de l'étape # i , seules les entrées #0 à # $i - 1$ sont présentes dans le dictionnaire.

#	Texte	Couple	Entrée
0			ϵ
1	caabbcaabcba	$\langle 0, c \rangle$	c
2	c aabbcaabcba	$\langle 0, a \rangle$	a
3	ca abbcaabcba	$\langle 2, b \rangle$	ab
4	caab bcaabcba	$\langle 0, b \rangle$	b
5	caabb caabcba	$\langle 1, a \rangle$	ca
6	caabbca abcba	$\langle 3, c \rangle$	abc
7	caabbcaabc ba	$\langle 4, a \rangle$	ba
8	caabbcaabcba		

- (d) À nouveau, l'état du texte à compresser, les index et les entrées du dictionnaire sont donnés de manière compacte. Les étapes de traitement commencent au #3. Lors de l'étape # i , seules les entrées #0 à # $i - 1$ sont présentes dans le dictionnaire.

#	Texte	Index	Entrée
0			a
1			b
2			c
3	caabbcaabcba	2	ca
4	c aabbcaabcba	0	aa
5	ca abbcaabcba	0	ab
6	caab bcaabcba	1	bb
7	caab bcaabcba	1	bc
8	caabb caabcba	3	caa
9	caabbca abcba	5	abc
10	caabbcaab cba	2	cb
11	caabbcaabc ba	1	ba
12	caabbcaabcba	0	a?

(e)

Triplet	Texte
$\langle 0, 0, c \rangle$	c
$\langle 0, 0, r \rangle$	cr
$\langle 0, 0, a \rangle$	cra
$\langle 0, 0, n \rangle$	cran
$\langle 2, 1, c \rangle$	cranac
$\langle 5, 3, e \rangle$	cranacrane
$\langle 5, 1, a \rangle$	cranacraneca
$\langle 6, 1, n \rangle$	cranacranecarn
$\langle 0, 0, e \rangle$	cranacranecarne

(f)

Message	Texte
$\langle r \rangle$	r
$\langle e \rangle$	re
$\langle p \rangle$	rep
$\langle e \rangle$	repe
$\langle t \rangle$	repet
$\langle i \rangle$	repeti
$\langle 2, 2 \rangle$	repetiti
$\langle v \rangle$	repetitiv
$\langle 4, 2 \rangle$	repetitivit
$\langle e \rangle$	repetitivite

(g) Au moment de décoder le couple $\#i$, seules les entrées $\#0$ à $\#i - 1$ sont présentes dans le dictionnaire.

#	Couple	Entrée	Texte
0		ϵ	
1	$\langle 0, c \rangle$	c	c
2	$\langle 0, a \rangle$	a	ca
3	$\langle 0, t \rangle$	t	cat
4	$\langle 2, l \rangle$	al	catal
5	$\langle 2, n \rangle$	an	catalan
6	$\langle 0, i \rangle$	i	catalani
7	$\langle 2, n \rangle$	an	catalanian

- (h) Au moment de décoder l'index $\#i$, seules les entrées $\#0$ à $\#i - 1$ sont présentes dans le dictionnaire.

#	Index	Entrée	Texte	Ent. préc.
0		a		
1		e		a
2		l		e
3		m		l
4		n		m
5		t		n
6	3	m?	m	t
7	1	e?	me	me
8	4	n?	men	en
9	5	t?	ment	nt
10	0	a?	menta	ta
11	2	l?	mental	al
12	1	e?	mentale	le
13	6	me?	mentaleme	em
14	8	nt?	mentalement	men

- (i) Lorsque plusieurs plus longues copies sont disponibles, les différents choix sont séparés par des barres obliques.

Texte	Message(s)
cabaabcaabbcbbaa	$\langle c \rangle$
c abaabcaabbcbbaa	$\langle a \rangle$
ca baabcaabbcbbaa	$\langle b \rangle$
cab aabcaabbcbbaa	$\langle a \rangle$
caba abcaabbcbbaa	$\langle 3, 2 \rangle$
cabaab caabbcbbaa	$\langle 6, 2 \rangle$
cabaabca abbcbbaa	$\langle 4, 2 \rangle / \langle 7, 2 \rangle$
cabaabcaab bcbbaa	$\langle 5, 2 \rangle$
cabaabcaabbc bbaa	$\langle 3, 2 \rangle$
cabaabcaabbcbb aa	$\langle 7, 2 \rangle / \langle 11, 2 \rangle$
cabaabcaabbcbbaa	

(j) Chaque façon de traiter l'entrée est donnée séparément.

Texte	Msg.	Texte	Msg.	Texte	Msg.	Texte	Msg.
chouchou	⟨c⟩	chouchou	⟨c⟩	chouchou	⟨c⟩	chouchou	⟨c⟩
çhouchou	⟨h⟩	çhouchou	⟨h⟩	çhouchou	⟨h⟩	çhouchou	⟨h⟩
chouchou	⟨o⟩	chouchou	⟨o⟩	chouchou	⟨o⟩	chouchou	⟨o⟩
chouchou	⟨u⟩	chouchou	⟨u⟩	chouchou	⟨u⟩	chouchou	⟨u⟩
chou chou	⟨4, 4⟩	chou chou	⟨4, 3⟩	chou chou	⟨4, 2⟩	chou chou	⟨4, 2⟩
chouchou		chouchou	⟨u⟩	chouchou	⟨4, 2⟩	chouchou	⟨o⟩
						chouchou	⟨u⟩
						chouchou	

Texte	Msg.	Texte	Msg.	Texte	Msg.	Texte	Msg.
chouchou	⟨c⟩	chouchou	⟨c⟩	chouchou	⟨c⟩	chouchou	⟨c⟩
çhouchou	⟨h⟩	çhouchou	⟨h⟩	çhouchou	⟨h⟩	çhouchou	⟨h⟩
chouchou	⟨o⟩	chouchou	⟨o⟩	chouchou	⟨o⟩	chouchou	⟨o⟩
chouchou	⟨u⟩	chouchou	⟨u⟩	chouchou	⟨u⟩	chouchou	⟨u⟩
chou chou	⟨c⟩	chou chou	⟨c⟩	chou chou	⟨c⟩	chou chou	⟨c⟩
chouchou	⟨4, 3⟩	chouchou	⟨4, 2⟩	chouchou	⟨h⟩	chouchou	⟨h⟩
		chouchou	⟨u⟩	chouchou	⟨4, 2⟩	chouchou	⟨o⟩
						chouchou	⟨u⟩
						chouchou	

2. Dans les réponses suivantes, nous distinguons les deux instances du contexte — présentes dans les modèles d'ordre 0 et -1 en les notant $-^0$ et $-^1$, respectivement.

- (a) Encodage de ⟨Esc⟩ dans le contexte **bac**; prob. $\frac{1}{2}$.
 Encodage de ⟨Esc⟩ dans le contexte **ac**; prob. $\frac{1}{2}$.
 Encodage de ⟨Esc⟩ dans le contexte **c**; prob. $\frac{1}{3}$.
 Encodage de **b** dans le contexte $-^0$; prob. $\frac{2}{9}$.
 Incrémentement de la fréquence de **b** dans les contextes **bac**, **ac**, **c** et $-^0$.
- (b) Encodage de **c** dans le contexte **bac**; prob. $\frac{1}{2}$.
 Incrémentement de la fréquence de **c** dans les contextes **bac**, **ac**, **c** et $-^0$.
 Encodage de ⟨Esc⟩ dans le contexte **acc**; prob. $\frac{1}{2}$.
 Encodage de ⟨Esc⟩ dans le contexte **cc**; prob. $\frac{1}{2}$.
 Encodage de ⟨Esc⟩ dans le contexte **c**; prob. $\frac{1}{4}$.
 Encodage de ⟨Esc⟩ dans le contexte $-^0$; prob. $\frac{1}{10}$.
 Encodage de **d** dans le contexte $-^1$; prob. $\frac{1}{4}$.
 Incrémentement de la fréquence de **d** dans les contextes **acc**, **cc**, **c** et $-^0$.

- (c) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **bac**; prob. $\frac{1}{2}$.
 Encodage de $\langle \text{Esc} \rangle$ dans le contexte **ac**; prob. $\frac{1}{2}$.
 Encodage de **a** dans le contexte **c**; prob. $\frac{1}{3}$.
 Incrémentation de la fréquence de **a** dans les contextes **bac**, **ac**, **c** et $—^0$.
 Création du contexte **aca** et encodage automatique de $\langle \text{Esc} \rangle$; prob. 1.
 Encodage de **b** dans le contexte **ca**; prob. $\frac{1}{2}$.
 Incrémentation de la fréquence de **b** dans les contextes **aca**, **ca**, **a** et $—^0$.
- (d) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **bac**; prob. $\frac{1}{2}$.
 Encodage de $\langle \text{Esc} \rangle$ dans le contexte **ac**; prob. 1, car exclusion de **c**.
 Encodage de $\langle \text{Esc} \rangle$ dans le contexte **c**; prob. $\frac{1}{2}$, car exclusion de **c**.
 Encodage de **b** dans le contexte $—^0$; prob. $\frac{2}{3}$, car exclusion de **a** et **c**.
 Incrémentation de la fréquence de **b** dans les contextes **bac**, **ac**, **c** et $—^0$.

3. (a) La matrice des rotations de la chaîne et la version triée de la matrice sont les suivantes.

Matrice des rotations:

contamination
 ontaminationc
 ntaminationco
 taminationcon
 aminationcont
 minationconta
 inationcontam
 nationcontami
 ationcontamin
 tioncontamina
 ioncontaminat
 oncontaminati
 ncontaminatio

Matrice des rotations triée:

aminationcont
 ationcontamin
 contamination
 inationcontam
 ioncontaminat
 minationconta
 nationcontami
 ncontaminatio
 ntaminationco
 oncontaminati
 ontaminationc
 taminationcon
 tioncontamina

Le résultat de la Transformée de Burrows-Wheeler est

$\langle \text{tnnmtaiooicna}, 2 \rangle$.

(b) On sait que la matrice triée des rotations a la forme suivante.

```

· ... c
· ... c
· ... e
· ... a
· ... t
· ... u
· ... a
· ... e
· ... h

```

En faisant une rotation vers la droite de toutes les rangées par une position, on obtient une autre matrice des rotations, laquelle n'est pas (nécessairement) triée. Bien sûr, on peut trier cette matrice et obtenir une nouvelle matrice qui, elle aussi, est un extrait de la matrice triée des rotations.

```

c ... ·          a ... ·
c ... ·          a ... ·
e ... ·          c ... ·
a ... ·          c ... ·
t ... ·          e ... ·
u ... ·          e ... ·
a ... ·          h ... ·
e ... ·          t ... ·
h ... ·          u ... ·

```

⇒

Puisqu'on connaît le contenu de la dernière colonne, on peut l'inscrire dans l'extrait que l'on vient d'obtenir. On peut à nouveau effectuer une rotation et, ensuite, trier les rangées.

```

a ... c          ca ... ·          ac ... ·
a ... c          ca ... ·          ah ... ·
c ... e          ec ... ·          ca ... ·
c ... a          ac ... ·          ca ... ·
e ... t          te ... ·          ec ... ·
e ... u          ue ... ·          et ... ·
h ... a          ah ... ·          hu ... ·
t ... e          et ... ·          te ... ·
u ... h          hu ... ·          ue ... ·

```

⇒

On peut refaire ces opérations pour déterminer le contenu de la troisième colonne.

ac ... c		cac		aca
ah ... c		cah		ahu
ca ... e		eca		cac
ca ... a		aca		cah
ec ... t	⇒	tec	⇒	eca
et ... u		uet		ete
hu ... a		ahu		hue
te ... e		ete		tec
ue ... h		hue		uet

Encore une fois.

aca ... c		caca		acah
ahu ... c		cahu		ahue
cac ... e		ecac		caca
cah ... a		acah		cahu
eca ... t	⇒	teca	⇒	ecac
ete ... u		uete		etec
hue ... a		ahue		huet
tec ... e		etec		teca
uet ... h		huet		uete

Normalement, on continuerait à effectuer ces deux opérations à répétition et on finirait par obtenir la matrice triée complète. Cependant, une personne réalise vite que la liste de préfixes de longueur 4 permet de les chaîner les uns à la suite des autres pour découvrir la chaîne originale. C'est-à-dire que les trois dernières lettres d'un préfixe détermine sans ambiguïté le prochain préfixe. Comme le rang est 2, on sait que la chaîne originale commence par **caca**. Les trois dernières lettres déterminent sans ambiguïté le prochain préfixe, qui est **acah**, ce qui signifie que la chaîne originale commence par **cacah**. En continuant ainsi, on reconstitue facilement **cacahuete**.

(c) Voici une trace montrant les listes d'ordonnancement successives ainsi que les symboles à encoder et les index obtenus.

a	c	i	d	i	f	i	e	e
↪	↪	↪	↪	↪	↪	↪	↪	↪
0	2	8	4	1	6	1	6	0

0: a	0: a	0: c	0: i	0: d	0: i	0: f	0: i	0: e	0: e
1: b	1: b	1: a	1: c	1: i	1: d	1: i	1: f	1: i	1: i
2: c	2: c	2: b	2: a	2: c	2: c	2: d	2: d	2: f	2: f
3: d	3: d	3: d	3: b	3: a	3: a	3: c	3: c	3: d	3: d
4: e	4: e	4: e	4: d	4: b	4: b	4: a	4: a	4: c	4: c
5: f	5: f	5: f	5: e	5: e	5: e	5: b	5: b	5: a	5: a
6: g	6: g	6: g	6: f	6: f	6: f	6: e	6: e	6: b	6: b
7: h	7: h	7: h	7: g						
8: i	8: i	8: i	8: h						
9: j									

(d)

3	4	4	8	3	3	0
↪	↪	↪	↪	↪	↪	↪
d	e	c	i	d	e	e

0: a	0: d	0: e	0: c	0: i	0: d	0: e	0: e
1: b	1: a	1: d	1: e	1: c	1: i	1: d	1: d
2: c	2: b	2: a	2: d	2: e	2: c	2: i	2: i
3: d	3: c	3: b	3: a	3: d	3: e	3: c	3: c
4: e	4: e	4: c	4: b	4: a	4: a	4: a	4: a
5: f	5: f	5: f	5: f	5: b	5: b	5: b	5: b
6: g	6: g	6: g	6: g	6: f	6: f	6: f	6: f
7: h	7: h	7: h	7: h	7: g	7: g	7: g	7: g
8: i	8: i	8: i	8: i	8: h	8: h	8: h	8: h
9: j							

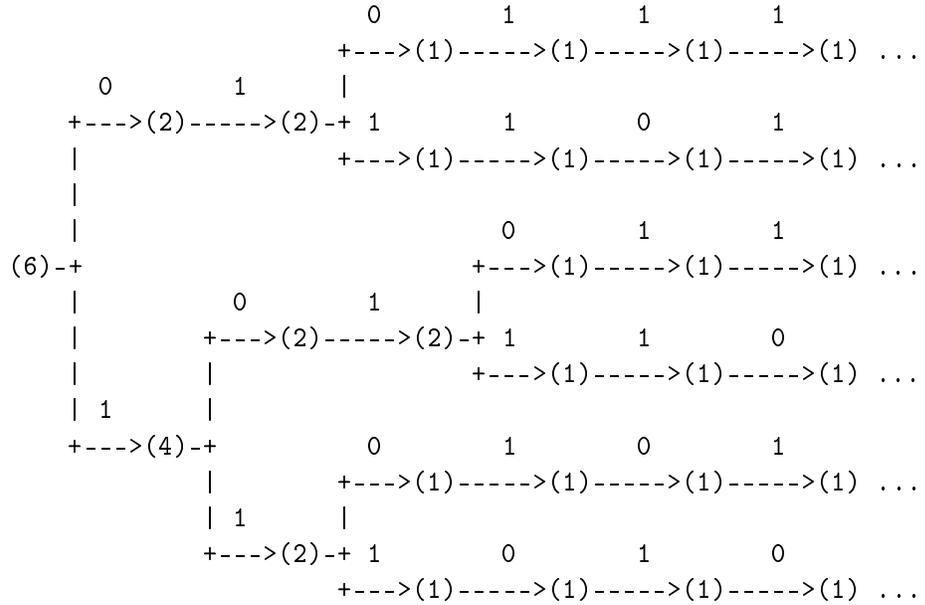
4. (a)

Length	Substrings					
0	6 × ε					
1	2 × 0		4 × 1			
2	2 × 01		2 × 10		2 × 11	
3	1 × 010	1 × 011	2 × 101		1 × 110	1 × 111
4	1 × 0101	1 × 0111	1 × 1010	1 × 1011	1 × 1101	1 × 1110
5	1 × 01011	1 × 01110	1 × 10101	1 × 10111	1 × 11010	1 × 11101
6	1 × 010111	1 × 011101	1 × 101011	1 × 101110	1 × 110101	1 × 111010

(b)

$$0 = \max(0, C_{01} - C_{11}) \leq C_{010} \leq \min(C_{01}, C_{10}) = 2$$

(c)



(d)

