

Solutionnaire du travail pratique #2

Réponses

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

- (a)

Texte	Triplet
acbcbadbcdbaacdbbcacbdda	$\langle 0, 0, a \rangle$
a cbcbadbcdbaacdbbcacbdda	$\langle 0, 0, c \rangle$
ac bcbadbcdbaacdbbcacbdda	$\langle 0, 0, b \rangle$
acb cbadbcdbaacdbbcacbdda	$\langle 2, 2, a \rangle$
acbcba dbcdbaacdbbcacbdda	$\langle 0, 0, d \rangle$
acbcbad bcdbaacdbbcacbdda	$\langle 5, 2, d \rangle$
acbcbadbcd baacdbbcacbdda	$\langle 6, 2, a \rangle$
acbcbadbcdbaa cddbbcacbdda	$\langle 5, 3, b \rangle$
acbcbadbcdbaacd bb cacbdda	$\langle 4, 1, a \rangle$
acbcbadbcdbaacd bbca cbdda	$\langle 16, 2, d \rangle$
acbcbadbcdbaacd bbcacbd da	$\langle 1, 1, a \rangle$
acbcbadbcdbaacd bbcacbdda	

(b)

Texte	Message
acbcbadbcdbaacdbbcacbdda	$\langle a \rangle$
a c c b a d b c d b a a c d b b c a c b d d a	$\langle c \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle b \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 2, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle a \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle d \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 5, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 3, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle a \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 12, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 5, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 9, 2 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle 18, 3 \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle d \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle d \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	$\langle a \rangle$
a c b c b a d b c d b a a c d b b c a c b d d a	

- (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	acbcbadbcdbaacdbbcacbdda	$\langle 0, a \rangle$	a
2	a cbcbadbcdbaacdbbcacbdda	$\langle 0, c \rangle$	c
3	ac bcbadbcdbaacdbbcacbdda	$\langle 0, b \rangle$	b
4	acb cbadbcdbaacdbbcacbdda	$\langle 2, b \rangle$	cb
5	acbc badbcdbaacdbbcacbdda	$\langle 1, d \rangle$	ad
6	acbcba bcdbaacdbbcacbdda	$\langle 3, c \rangle$	bc
7	acbcbadbc dbaacdbbcacbdda	$\langle 0, d \rangle$	d
8	acbcbadbcd baacdbbcacbdda	$\langle 3, a \rangle$	ba
9	acbcbadbcdba acdbbcacbdda	$\langle 1, c \rangle$	ac
10	acbcbadbcdbaac dbbbcacbdda	$\langle 7, b \rangle$	db
11	acbcbadbcdbaacd bbcacbdda	$\langle 6, a \rangle$	bca
12	acbcbadbcdbaacdbbc acbdda	$\langle 4, d \rangle$	cbd
13	acbcbadbcdbaacdbbcac dbda	$\langle 7, a \rangle$	da
14	acbcbadbcdbaacdbbcac bdda		

- (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 #4. 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			d
4	acbcbadbcdbaacdbbcacbdda	0	ac
5	a cbcbadbcdbaacdbbcacbdda	2	cb
6	ac bcbadbcdbaacdbbcacbdda	1	bc
7	acbc badbcdbaacdbbcacbdda	5	cba
8	acbcba dbcdbaacdbbcacbdda	0	ad
9	acbcba dbcdbaacdbbcacbdda	3	db
10	acbcba dbcdbaacdbbcacbdda	6	bcd
11	acbcba dbcdbaacdbbcacbdda	9	dba
12	acbcba dbcdbaacdbbcacbdda	0	aa
13	acbcba dbcdbaacdbbcacbdda	4	acd
14	acbcba dbcdbaacdbbcacbdda	9	dbb
15	acbcba dbcdbaacdbbcacbdda	6	bca
16	acbcba dbcdbaacdbbcacbdda	4	acb
17	acbcba dbcdbaacdbbcacbdda	1	bd
18	acbcba dbcdbaacdbbcacbdda	3	dd
19	acbcba dbcdbaacdbbcacbdda	3	da
20	acbcba dbcdbaacdbbcacbdda	0	a?
21	acbcba dbcdbaacdbbcacbdda		

(e)

Triplet	Texte
$\langle 0, 0, a \rangle$	a
$\langle 0, 0, c \rangle$	ac
$\langle 0, 0, d \rangle$	acd
$\langle 0, 0, b \rangle$	acdb
$\langle 1, 2, a \rangle$	acdbbba
$\langle 6, 1, a \rangle$	acdbbbaca
$\langle 0, 0, c \rangle$	acdbbbacac

(f)

Message	Texte
$\langle a \rangle$	a
$\langle b \rangle$	ab
$\langle b \rangle$	abb
$\langle c \rangle$	abbc
$\langle d \rangle$	abbcd
$\langle c \rangle$	abbcdcd
$\langle 6, 2 \rangle$	abbcdcab
$\langle 5, 2 \rangle$	abbcdcabcd

(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	$\langle 0, b \rangle$	b	b
2	$\langle 0, a \rangle$	a	ba
3	$\langle 0, d \rangle$	d	bad
4	$\langle 1, c \rangle$	bc	badbc
5	$\langle 2, b \rangle$	ab	badbcab
6	$\langle 0, c \rangle$	c	badbcabc
7	$\langle 3, b \rangle$	db	badbcabcdb

- (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		b		a
2		c		b
3		d		c
4	1	b?	b	d
5	2	c?	bc	bc
6	4	bc?	bcbcb	cb
7	3	d?	bcbcb	bcd
8	3	d?	bcbcb	dd
9	4	bc?	bcbcb	db
10	5	cb?	bcbcb	bcc
11	9	bcc?	bcbcb	cbb

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

Texte	Message(s)
lebardubarbarebarbu	$\langle 1 \rangle$
lebardubarbarebarbu	$\langle e \rangle$
lebardubarbarebarbu	$\langle b \rangle$
lebardubarbarebarbu	$\langle a \rangle$
lebardubarbarebarbu	$\langle r \rangle$
lebardubarbarebarbu	$\langle d \rangle$
lebardubarbarebarbu	$\langle u \rangle$
lebardubarbarebarbu	$\langle 5, 3 \rangle$
lebardubarbarebarbu	$\langle 3, 3 \rangle / \langle 8, 3 \rangle$
lebardubarbarebarbu	$\langle 12, 4 \rangle$
lebardubarbarebarbu	$\langle b \rangle$
lebardubarbarebarbu	$\langle u \rangle$
lebardubarbarebarbu	

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

Texte	Msg.	Texte	Msg.	Texte	Msg.	Texte	Msg.
raisinraide	<r>	raisinraide	<r>	raisinraide	<r>	raisinraide	<r>
ṛaisinraide	<a>	ṛaisinraide	<a>	ṛaisinraide	<a>	ṛaisinraide	<a>
rạisinraide	<i>	rạisinraide	<i>	rạisinraide	<i>	rạisinraide	<i>
raịsinraide	<s>	raịsinraide	<s>	raịsinraide	<s>	raịsinraide	<s>
raiṣinraide	<i>	raiṣinraide	<i>	raiṣinraide	<i>	raiṣinraide	<i>
raisiṇraide	<n>	raisiṇraide	<n>	raisiṇraide	<n>	raisiṇraide	<n>
raisinraide	<6,3>	raisinraide	<6,2>	raisinraide	<r>	raisinraide	<r>
raisinraide	<d>	raisinraide	<i>	raisinraide	<6,2>	raisinraide	<a>
raisinraide	<e>	raisinraide	<d>	raisinraide	<d>	raisinraide	<i>
raisinraide		raisinraide	<e>	raisinraide	<e>	raisinraide	<d>
raisinraide		raisinraide		raisinraide		raisinraide	<e>

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 **a** dans le contexte **am**; prob. $1/2$.
Incrémentation de la fréquence de **a** dans les contextes **am**, **m** et $—^0$.
- (b) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **am**; prob. $1/2$.
Encodage de $\langle \text{Esc} \rangle$ dans le contexte **m**; prob. $1/2$.
Encodage de $\langle \text{Esc} \rangle$ dans le contexte $—^0$; prob. $1/15$.
Encodage de **i** dans le contexte $—^{-1}$; prob. ~~$1/26$~~ $1/22$ (voir la note plus bas).
Incrémentation de la fréquence de **i** dans les contextes **am**, **m** et $—^0$ (avec création préalable de l'entrée pour **i** là où elle était inexistante).
- (c) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **am**; prob. $1/2$.
Encodage de $\langle \text{Esc} \rangle$ dans le contexte **m**; prob. $1/2$.
Encodage de **o** dans le contexte $—^0$; prob. $3/15$.
Incrémentation de la fréquence de **o** dans les contextes **am**, **m** et $—^0$ (avec création préalable de l'entrée pour **o** là où elle était inexistante).
- (d) (a) Encodage de **a** dans le contexte **am**; prob. $1/2$.
Incrémentation de la fréquence de **a** dans les contextes **am**, **m** et $—^0$.
- (b) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **am**; prob. $1/2$.
Encodage de $\langle \text{Esc} \rangle$ dans le contexte **m**; prob. 1 (car **a** est exclu).
Encodage de $\langle \text{Esc} \rangle$ dans le contexte $—^0$; prob. $1/11$ (car **a** est exclu).
Encodage de **i** dans le contexte $—^{-1}$; prob. ~~$1/20$~~ $1/16$ (car **a**, **h**, **k**, **m**, **o** et **y** sont exclus).
Incrémentation de la fréquence de **i** dans les contextes **am**, **m** et $—^0$ (avec création préalable de l'entrée pour **i** là où elle était inexistante).
- (c) Encodage de $\langle \text{Esc} \rangle$ dans le contexte **am**; prob. $1/2$.
Encodage de $\langle \text{Esc} \rangle$ dans le contexte **m**; prob. 1 (car **a** est exclu).
Encodage de **o** dans le contexte $—^0$; prob. $3/11$ (car **a** est exclu).
Incrémentation de la fréquence de **o** dans les contextes **am**, **m** et $—^0$ (avec création préalable de l'entrée pour **o** là où elle était inexistante).

Note: On m'a signalé que, dans l'énoncé, il manquait quatre lettres dans le contexte d'ordre -1 . En effet, il n'y a que 22 des 26 lettres qui sont énumérées. C'est mon erreur. J'accepterai autant une réponse basée sur le contexte tel que présenté dans l'énoncé qu'une réponse basée sur l'hypothèse qu'il y a 26 lettres dans l'alphabet latin.

3. (a) La matrice des rotations de la chaîne et la version triée de la matrice sont les suivantes. Notez qu'on considère le caractère de soulignement comme venant avant les lettres, dans l'ordre lexicographique.

Matrice des rotations:

```
well_is_this_well_well
ell_is_this_well_wellw
ll_is_this_well_wellwe
l_is_this_well_wellwel
_is_this_well_wellwell
is_this_well_wellwell_
s_this_well_wellwell_i
_this_well_wellwell_is
this_well_wellwell_is_
his_well_wellwell_is_t
is_well_wellwell_is_th
s_well_wellwell_is_thi
_well_wellwell_is_this
well_wellwell_is_this_
ell_wellwell_is_this_w
ll_wellwell_is_this_we
l_wellwell_is_this_wel
_wellwell_is_this_well
wellwell_is_this_well_
ellwell_is_this_well_w
llwell_is_this_well_we
llwell_is_this_well_wel
```

Matrice des rotations triée:

```
_is_this_well_wellwell
_this_well_wellwell_is
_well_wellwell_is_this
_wellwell_is_this_well
ell_is_this_well_wellw
ell_wellwell_is_this_w
ellwell_is_this_well_w
his_well_wellwell_is_t
is_this_well_wellwell_
is_well_wellwell_is_th
l_is_this_well_wellwel
l_wellwell_is_this_wel
ll_is_this_well_wellwe
ll_wellwell_is_this_we
llwell_is_this_well_we
llwell_is_this_well_wel
s_this_well_wellwell_i
s_well_wellwell_is_thi
this_well_wellwell_is_
well_is_this_well_well
well_wellwell_is_this_
wellwell_is_this_well_
```

Le résultat de la Transformée de Burrows-Wheeler est

$\langle \text{lsslwwwt_hlleelii_l_}, 19 \rangle$.

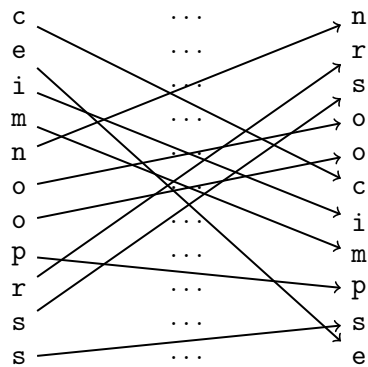
(b) On sait que la matrice triée des rotations a la forme suivante.

·	...	n
·	...	r
·	...	s
·	...	o
·	...	o
·	...	c
·	...	i
·	...	m
·	...	p
·	...	s
·	...	e

On sait aussi que la première colonne contient une version triée de la dernière colonne.

c	...	n
e	...	r
i	...	s
m	...	o
n	...	o
o	...	c
o	...	i
p	...	m
r	...	p
s	...	s
s	...	e

De plus, on sait que c'est un tri stable qui envoie les caractères de la dernière colonne vers leurs équivalents dans la première colonne. On peut donc tracer le graphe montrant à quel caractère de la dernière colonne chacun des caractères de la première colonne correspond.



En commençant à la ligne numérotée 0 (le rang), on peut facilement suivre les flèches et passer aux voisins de droite (en bouclant de la dernière colonne à la première colonne) et récupérer **compression**.

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

```

n   r   s   o   o   c   i   m   p   s   e
↳   ↳   ↳   ↳   ↳   ↳   ↳   ↳   ↳   ↳   ↳
4   7   8   7   0   4   6   7   8   5   8

```

0: c	0: n	0: r	0: s	0: o	0: o	0: c	0: i	0: m	0: p	0: s	0: e
1: e	1: c	1: n	1: r	1: s	1: s	1: o	1: c	1: i	1: m	1: p	1: s
2: i	2: e	2: c	2: n	2: r	2: r	2: s	2: o	2: c	2: i	2: m	2: p
3: m	3: i	3: e	3: c	3: n	3: n	3: r	3: s	3: o	3: c	3: i	3: m
4: n	4: m	4: i	4: e	4: c	4: c	4: n	4: r	4: s	4: o	4: c	4: i
5: o	5: o	5: m	5: i	5: e	5: e	5: e	5: n	5: r	5: s	5: o	5: c
6: p	6: p	6: o	6: m	6: i	6: i	6: i	6: e	6: n	6: r	6: r	6: o
7: r	7: r	7: p	7: o	7: m	7: m	7: m	7: m	7: e	7: n	7: n	7: r
8: s	8: s	8: s	8: p	8: p	8: p	8: p	8: p	8: p	8: e	8: e	8: n

4. Voici les erreurs moyennes que j'ai obtenues en utilisant les divers prédicteurs.

Prédicteur	Erreur moyenne
1	39.53
2	4.90
3	6.55
4	7.66
5	5.22
6	5.04
7	4.47
8	4.74

Ceci signifie que le prédicteur #7 est le plus performant pour l'image fournie et il devrait minimiser la taille du fichier compressé.

5. (a) On a 6 séquences à émettre.
- 2 pixels blancs: 0111
 - 2 pixels noirs: 11
 - 2 pixels blancs: 0111
 - 2 pixels noirs: 11
 - 3 pixels blancs: 1000
 - 5 pixels noirs: 0011
- (b) On a 5 étapes à effectuer. Supposons qu'on numérote les colonnes à partir de 0.
- i. a_0 pointe sur le pixel imaginaire blanc avant la ligne, $a_0 = -1$. Le premier changement se produit à $a_1 = 2$. Le deuxième changement se produit à $a_2 = 4$. Dans la ligne de référence, le premier passage à un pixel noir à droite de a_0 (la couleur opposée de celle du pixel pointé par a_0) se situe à $b_1 = 1$. Le changement suivant se produit ensuite à $b_2 = 4$.
Le mode vertical est adopté car a_1 pointe avant b_2 et $|a_1 - b_1| \leq 3$. On encode $a_1 - b_1 = 1$, ce qui cause l'émission du mot de code 011.
On termine en avançant a_0 jusqu'au niveau de a_1 .
 - ii. $a_0 = 2$ pointe sur le premier pixel noir de la ligne à encoder. Le prochain changement se produit à $a_1 = 4$. Le changement suivant se produit à $a_2 = 6$. On a $b_1 = 4$ qui pointe sur le premier pixel blanc de la ligne de référence à droite de a_0 . Enfin, $b_2 = 10$.
Le mode vertical est adopté car $a_1 \leq b_2$ et $|a_1 - b_1| \leq 3$. On encode $a_1 - b_1 = 0$ par le mot de code 1.
On avance a_0 jusqu'à a_1 .
 - iii. On a $a_0 = 4$, $a_1 = 6$ et $a_2 = 8$. Sur la ligne de référence, on a $b_1 = 10$ et $b_2 = 12$.
Le mode horizontal est adopté car $a_1 \leq b_2$ mais $|a_1 - b_1| > 3$. On envoie la longueur des deux séquences avec le code HM. Il s'agit d'une séquence blanche suivie d'une séquence noire. Les mots de code 0111 et 11 sont émis.
On avance a_0 jusqu'à a_2 .
 - iv. On a $a_0 = 8$, $a_1 = 11$ et $a_2 = 16$. On a aussi $b_1 = 10$ et $b_2 = 12$.
Le mode vertical est adopté car $a_1 \leq b_2$ et $|a_1 - b_1| \leq 3$. On encode $a_1 - b_1 = 1$ par le mot de code 011.
On avance a_0 jusqu'à a_1 .
 - v. On a $a_0 = 11$, $a_1 = 16$ et $a_2 = +\infty$ (disons). On a aussi $b_1 = 12$ et $b_2 = 16$.
Le mode horizontal est adopté car $a_1 \leq b_2$ mais $|a_1 - b_1| > 3$. On enverrait normalement la longueur des deux prochaines séquences avec le code HM. Toutefois, l'envoi de la première va suffire à compléter la ligne. Comme il s'agit d'une séquence noire, le mot de code 0011 est émis.
On avance a_0 jusqu'à a_1 et on a terminé l'encodage de la ligne.
6. En réalité, il n'y avait pas de question 6 dans le TP. Toutefois, je tiens à souligner que j'avais oublié de mettre des questions sur **l'énumération de sous-chaînes**. Cette matière est **quand même** sujette à examen!