

Applications de la transformée de Burrows-Wheeler

Mikaël Salson, Thierry Lecroq, Martine Léonard,
Laurent Mouchard

LITIS, Université de Rouen

Journée des doctorants de l'école doctorale SPMII
24 février 2009



Recherche de données

But : Trouver tous les emplacements d'un *motif* dans un texte.

Recherche de données

But : Trouver tous les emplacements d'un *motif* dans un texte.

Solutions :

- Lire tout le texte.
- Utiliser une *structure d'indexation*.

Cette mesure était financée par des annulations de mêmes montants en autorisations d'engagement et en crédits de paiement. Les annulations les plus importantes ont porté sur les missions **Recherche et enseignement supérieur** (50,11 M€ en CP), **Immigration, asile et intégration** (36 M€ en CP) et **Ecologie, développement et aménagement durables** (33,31 M€ en CP) ;

- le décret en cours de publication à la date du présent rapport ouvre, sur le budget général, 1.294,36 M€ en autorisations d'engagement et 1.253,16 M€ en crédits de paiement, au titre des opérations militaires extérieures (259,21 M€ en AE et en CP), des aides à l'emploi (277,40 M€ en AE et 251,40M€ en CP), de la prime de Noël des bénéficiaires du RMI et de l'hébergement d'urgence et des personnes en grande précarité (pour un total de 428,80 M€ en AE et en CP) et des engagements financiers de l'Etat (150,00 M€ en AE et en CP).

Par ailleurs, le décret remédie à diverses insuffisances de crédits de personnel (pour un total de 122,2 M€ en AE et en CP sur le budget général et de 34,6 M€ en AE et en CP sur le budget annexe *Contrôle et exploitation aériens*).

En contrepartie, sont annulés des montants similaires sur vingt-quatre missions. Quatre d'entre elles supportent 60 % du total des annulations : **Recherche et enseignement supérieur**, **Défense**, **Ecologie, développement et aménagement durables** et **Gestion des finances et des ressources humaines**. A elle seule, la mission **Recherche et enseignement supérieur** aura donc supporté 22 % du total des annulations de crédits opérées en gestion 2008 (386 M€).

Rechercher : recherche

Caractéristiques

- Parcours du texte pour rechercher les termes choisis.

Avantages

- Pas d'espace supplémentaire.
- Ne dépend pas du texte.
- Recherche exhaustive.

Inconvénient

- Recherche pouvant être très longue.

Cette mesure était financée par des annulations de mêmes montants en autorisations d'engagement et en crédits de paiement. Les annulations les plus importantes ont porté sur les missions **Recherche et enseignement supérieur** (50,11 M€ en CP), **Immigration, asile et intégration** (36 M€ en CP) et **Ecologie, développement et aménagement durables** (33,31 M€ en CP) ;

- le décret en cours de publication à la date du présent rapport ouvre, sur le budget général, 1.294,36 M€ en autorisations d'engagement et 1.253,16 M€ en crédits de paiement, au titre des opérations militaires extérieures (259,21 M€ en AE et en CP), des aides à l'emploi (277,40 M€ en AE et 251,40M€ en CP), de la prime de Noël des bénéficiaires du RMI et de l'hébergement d'urgence et des personnes en grande précarité (pour un total de 428,80 M€ en AE et en CP) et des engagements financiers de l'Etat (150,00 M€ en AE et en CP).

Par ailleurs, le décret remédie à diverses insuffisances de crédits de personnel (pour un total de 122,2 M€ en AE et en CP sur le budget général et de 34,6 M€ en AE et en CP sur le budget annexe *Contrôle et exploitation aériens*).

En contrepartie, sont annulés des montants similaires sur vingt-quatre missions. Quatre d'entre elles supportent 60 % du total des annulations : **Recherche et enseignement supérieur**, **Défense**, **Ecologie, développement et aménagement durables** et **Gestion des finances et des ressources humaines**. A elle seule, la mission **Recherche et enseignement supérieur** aura donc supporté 22 % du total des annulations de crédits opérées en gestion 2008 (386 M€).

Rechercher : recherche

Caractéristiques

- Parcours du texte pour rechercher les termes choisis.

Avantages

- Pas d'espace supplémentaire.
- Ne dépend pas du texte.
- Recherche exhaustive.

Inconvénient

- Recherche pouvant être très longue.

Cette mesure était financée par des annulations de mêmes montants en autorisations d'engagement et en crédits de paiement. Les annulations les plus importantes ont porté sur les missions **Recherche et enseignement supérieur** (50,11 M€ en CP), **Immigration, asile et intégration** (36 M€ en CP) et **Ecologie, développement et aménagement durables** (33,31 M€ en CP) ;

- le décret en cours de publication à la date du présent rapport ouvre, sur le budget général, 1.294,36 M€ en autorisations d'engagement et 1.253,16 M€ en crédits de paiement, au titre des opérations militaires extérieures (259,21 M€ en AE et en CP), des aides à l'emploi (277,40 M€ en AE et 251,40M€ en CP), de la prime de Noël des bénéficiaires du RMI et de l'hébergement d'urgence et des personnes en grande précarité (pour un total de 428,80 M€ en AE et en CP) et des engagements financiers de l'Etat (150,00 M€ en AE et en CP).

Par ailleurs, le décret remédie à diverses insuffisances de crédits de personnel (pour un total de 122,2 M€ en AE et en CP sur le budget général et de 34,6 M€ en AE et en CP sur le budget annexe *Contrôle et exploitation aériens*).

En contrepartie, sont annulés des montants similaires sur vingt-quatre missions. Quatre d'entre elles supportent 60 % du total des annulations : **Recherche et enseignement supérieur**, **Défense**, **Ecologie, développement et aménagement durables** et **Gestion des finances et des ressources humaines**. A elle seule, la mission **Recherche et enseignement supérieur** aura donc supporté 22 % du total des annulations de crédits opérées en gestion 2008 (386 M€).

Rechercher : recherche Précédent Suivant

Caractéristiques

- Parcours du texte pour rechercher les termes choisis.

Avantages

- Pas d'espace supplémentaire.
- Ne dépend pas du texte.
- Recherche exhaustive.

Inconvénient

- Recherche pouvant être très longue.



Caractéristiques de l'index

- Mots clés choisis.
- Pages sélectionnées.

Avantages

- Petite taille.
- Mise à jour relativement facile.
- Temps de recherche peu dépendant de la taille du texte.

Inconvénient

- Non exhaustif.



Caractéristiques de l'index

- Mots clés choisis.
- Pages sélectionnées.

Avantages

- Petite taille.
- Mise à jour relativement facile.
- Temps de recherche peu dépendant de la taille du texte.

Inconvénient

- Non exhaustif.



Caractéristiques de l'index

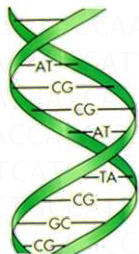
- Mots clés choisis.
- Pages sélectionnées.

Avantages

- Petite taille.
- Mise à jour relativement facile.
- Temps de recherche peu dépendant de la taille du texte.

Inconvénient

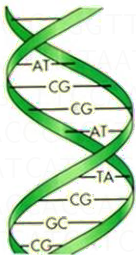
- Non exhaustif.



Chromosome 1, chez l'Homme

> 220 millions de paires de bases

Soit près de **50000 pages** avec 80 caractères par ligne et 56 lignes par page.



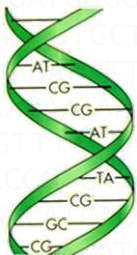
Chromosome 1, chez l'Homme

> 220 millions de paires de bases

Soit près de **50000 pages** avec 80 caractères par ligne et 56 lignes par page.

Problème

On veut répondre rapidement à toute requête du type « *Où sont toutes les occurrences de ACCAAGCG ?* ».



Chromosome 1, chez l'Homme

> 220 millions de paires de bases

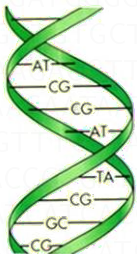
Soit près de **50000 pages** avec 80 caractères par ligne et 56 lignes par page.

Problème

On veut répondre rapidement à toute requête du type « *Où sont toutes les occurrences de ACCAAGCG ?* ».

Indexer comme un livre ?

Seules informations identifiables :
annotations ($\simeq 30\%$ du chromosome).



Chromosome 1, chez l'Homme

> 220 millions de paires de bases

Soit près de **50000 pages** avec 80 caractères par ligne et 56 lignes par page.

Problème

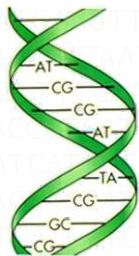
On veut répondre rapidement à toute requête du type « *Où sont toutes les occurrences de ACCAAGCG ?* ».

Indexer comme un livre ?

Seules informations identifiables :
annotations ($\simeq 30\%$ du chromosome).

Parcourir le texte ?

Beaucoup trop long !



Chromosome 1, chez l'Homme

> 220 millions de paires de bases

Soit près de **50000 pages** avec 80 caractères par ligne et 56 lignes par page.

Solution

- Pré-traitement : indexation de **tout** le chromosome.
- Index de taille plus importante que le texte !

But : Stocker une même information en utilisant moins de place.

But : Stocker une même information en utilisant moins de place.



Compression



But : Stocker une même information en utilisant moins de place.

Solution : Transformée de Burrows-Wheeler → *bzip2*

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires	Permutations circulaires
0	GAGACGAT\$	
1	AGACGAT\$G	
2	GACGAT\$GA	
3	ACGAT\$GAG	
4	CGAT\$GAGA	
5	GAT\$GAGAC	
6	AT\$GAGACG	
7	T\$GAGACGA	
8	\$GAGACGAT	

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires	Permutations circulaires
0	GAGACGAT\$	
1	AGACGAT\$G	
2	GACGAT\$GA	
3	ACGAT\$GAG	
4	CGAT\$GAGA	
5	GAT\$GAGAC	
6	AT\$GAGACG	
7	T\$GAGACGA	
8	\$GAGACGAT	

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G		
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G		
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

i	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G		
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA		
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA	6	AT\$GAGACG
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA	6	AT\$GAGACG
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA	6	AT\$GAGACG
3	ACGAT\$GAG		
4	CGAT\$GAGA		
5	GAT\$GAGAC		
6	AT\$GAGACG		
7	T\$GAGACGA		
8	\$GAGACGAT		

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ & G & A & G & A & C & G & A & T & \$ \end{matrix}$$

<i>i</i>	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GAG
1	AGACGAT\$G	1	AGACGAT\$G
2	GACGAT\$GA	6	AT\$GAGACG
3	ACGAT\$GAG	4	CGAT\$GAGA
4	CGAT\$GAGA	2	GACGAT\$GA
5	GAT\$GAGAC	0	GAGACGAT\$
6	AT\$GAGACG	5	GAT\$GAGAC
7	T\$GAGACGA	7	T\$GAGACGA
8	\$GAGACGAT	8	\$GAGACGAT

Transformée de Burrows-Wheeler

Tri des permutations circulaires

$T = \begin{matrix} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \text{G} & \text{A} & \text{G} & \text{A} & \text{C} & \text{G} & \text{A} & \text{T} & \$ \end{matrix}$

i	Permutations circulaires		Permutations circulaires
0	GAGACGAT\$	3	ACGAT\$GA G
1	AGACGAT\$G	1	AGACGAT\$ G
2	GACGAT\$GA	6	AT\$GAGAC G
3	ACGAT\$GAG	4	CGAT\$GAGA A
4	CGAT\$GAGA	2	GACGAT\$GA A
5	GAT\$GAGAC	0	GAGACGAT\$ T
6	AT\$GAGACG	5	GAT\$GAGAC C
7	T\$GAGACGA	7	T\$GAGACGA A
8	\$GAGACGAT	8	\$GAGACGAT T

$TBW(T) = \text{GGGAA\$CAT}$

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

TBW

neteenrrrrssshhhha____a_uhuisscccccccheo
aas\$aueen___sss_aaanqqe

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

TBW

neteenrrrrssshhhha____a_uhuisscccccccheo
aas\$aueen___sss_aaanqqe

Espace utilisé & temps de recherche

Pour une séquence génomique de 100 Mo, recherche d'un "mot" de 25 lettres.

	Parcours du texte	Index	Index compressé
Espace (Mo.)	100	500	de 39 à 63
Recherche (ms.)	300	0,005	de 3 à 0,08

0	1	2	3	4	5	6	7	8
G	A	G	A	C	G	A	T	\$

	0	1	2	3	4	5	6	7	8
	G	A	G	A	C	G	A	T	\$
Construction					↓	coûteuse			

	0	1	2	3	4	5	6	7	8
	G	A	G	A	C	G	A	T	\$
Construction	↓ coûteuse								
i	TBW								
0	G								
1	G								
2	G								
3	A								
4	A								
5	\$								
6	C								
7	A								
8	T								

0 1 2 3 4 5 6 7 8
G A G A C G A T \$

Construction | coûteuse
↓

i	TBW
0	G
1	G
2	G
3	A
4	A
5	\$
6	C
7	A
8	T

0 1 2 3 4 5 6 7 8 9
G A G A T C G A T \$

0 1 2 3 4 5 6 7 8
G A G A C G A T \$

Construction | coûteuse
↓

i	TBW
0	G
1	G
2	G
3	A
4	A
5	\$
6	C
7	A
8	T

0 1 2 3 4 5 6 7 8 9
G A G A T C G A T \$

Construction | coûteuse
↓

	0	1	2	3	4	5	6	7	8
	G	A	G	A	C	G	A	T	\$
Construction					↓	coûteuse			
<i>i</i>	TBW								
0	G								
1	G								
2	G								
3	A								
4	A								
5	\$								
6	C								
7	A								
8	T								

	0	1	2	3	4	5	6	7	8	9	
	G	A	G	A	T	C	G	A	T	\$	
Construction						coûteuse					
<i>i</i>	<i>L</i>										
0	G										
1	G										
2	G										
3	T										
4	\$										
5	C										
6	A										
7	A										
8	A										
9	T										

0 1 2 3 4 5 6 7 8
G A G A C G A T \$

Construction | coûteuse

i	TBW
0	G
1	G
2	G
3	A
4	A
5	\$
6	C
7	A
8	T

Mise à jour

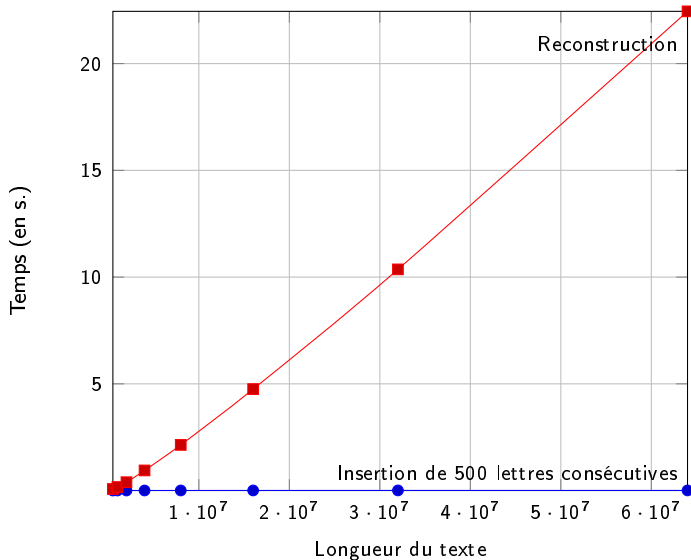
économique?

0 1 2 3 4 5 6 7 8 9
G A G A T C G A T \$

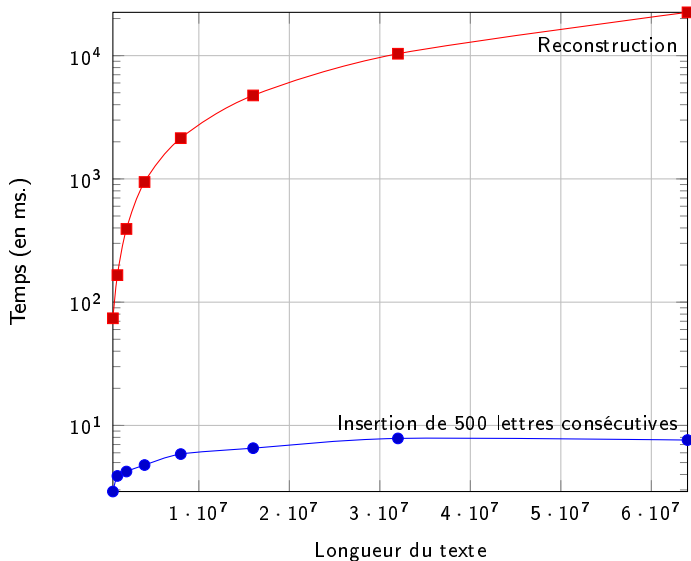
Construction | coûteuse

i	L
0	G
1	G
2	G
3	T
4	\$
5	C
6	A
7	A
8	A
9	T

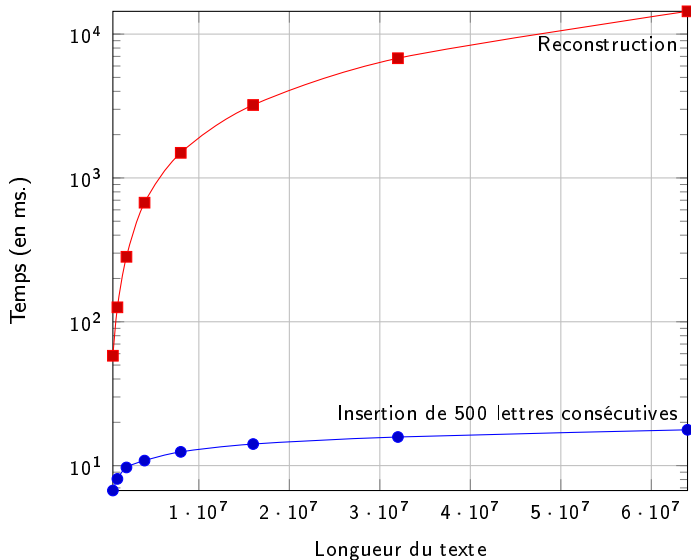
ADN



ADN



Données XML



Résultats

- Mise à jour efficace de la transformée de Burrows-Wheeler
- Structure d'indexation efficace en espace, en temps et qu'il est possible de mettre à jour.

Applications

- Indexation et mise à jour de séquences génomiques, de pages Wikipedia (modifications fréquentes).

Publications



M. Salson, T. Lecoq, M. Léonard, and L. Mouchard (2009). A Four-Stage Algorithm for Updating a Burrows-Wheeler Transform. *Theoretical Computer Science*. À paraître.



M. Salson, T. Lecoq, M. Léonard, and L. Mouchard (2009). Dynamic extended suffix arrays. *Journal of Discrete Algorithms*. À paraître.

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

TBW

neteenrrrrssshhhha____a_uhuisscccccccheo
aas\$aueen___sss_aaanqqe

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

TBW

neteenrrrrssshhhha____a_uhuisscccccccheo
aas\$aueen____sss_aaanqqe

Texte

que_chaque_chasseur_sachant_chasser_sache
_chasser_sans_son_chien\$

TBW

neteenrrrrssshhhha____a_uhuisscccccccheo
aas\$aueen____sss_aaanoqe

	<i>L</i>
hant_...	c
haque...	c
hasse...	c
hasse...	c
hasse...	c
he_ch...	c
hien\$...	c

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Plus petite permutation qui se **termine** par un T.

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	→ Plus petite permutation qui se termine par un T.
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	→ Plus petite permutation qui commence par un T.
0	TCGGTCGC\$	

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	→ Plus petite permutation qui se termine par un T.
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	→ Plus petite permutation qui commence par un T.
0	TCGGTCGC\$	

Transformée de Burrows-Wheeler (II)

SA	F	L
8	\$TCGGTCG	C
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un C → la précédente est la première à commencer par C.

C

Transformée de Burrows-Wheeler (II)

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un C → la précédente est la première à commencer par C.

C

Transformée de Burrows-Wheeler (II)

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un G → la précédente est la première à commencer par un G.

GC

Transformée de Burrows-Wheeler (II)

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un G → la précédente est la première à commencer par un G.

GC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGT	C
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un C → la précédente est la deuxième à commencer par un C.

CGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un C → la précédente est la deuxième à commencer par un C.

CGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un T → la précédente est la première à commencer par un T.

TCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un T → la précédente est la première à commencer par un T.

TCGC

Transformée de Burrows-Wheeler (II)

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCG G	
0	TCGGTCGC\$	

Troisième permutation qui se termine par un G → la précédente est la troisième à commencer par un T.

GTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Troisième permutation qui se termine par un G → la précédente est la troisième à commencer par un T.

GTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un G → la précédente est la deuxième à commencer par un G.

GTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un G → la précédente est la deuxième à commencer par un G.

GGTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$T	C
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Troisième permutation qui se termine par un C → la précédente est la troisième à commencer par un C.

C

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Troisième permutation qui se termine par un C → la précédente est la troisième à commencer par un C.

CGGTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un T → la précédente est la seconde à commencer par un T.

TCGGTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Deuxième permutation qui se termine par un T → la précédente est la seconde à commencer par un T.

TCGGTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un \$ → la précédente est la première à commencer par un \$.

TCGGTCGC

SA	F	L
8	\$TCGGTCGC	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	

Première permutation qui se termine par un \$ → la précédente est la première à commencer par un \$.

TCGGTCGC

SA	F	L
8	<u>\$TCGGTCGC</u>	
7	C\$TCGGTCG	
5	CGC\$TCGGT	
1	CGGTCGC\$T	
6	GC\$TCGGTC	
2	GGTCGC\$TC	
3	GTCGC\$TCG	
4	TCGC\$TCGG	
0	TCGGTCGC\$	