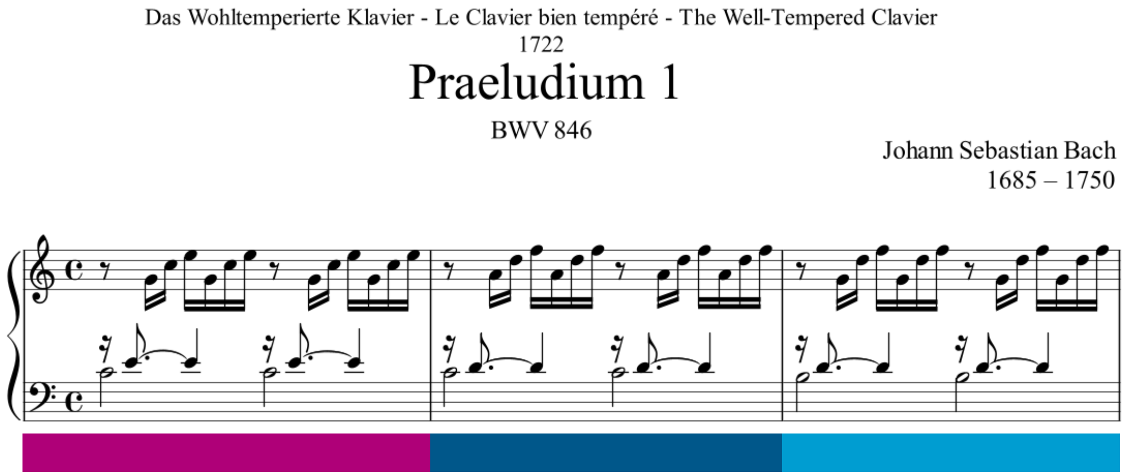
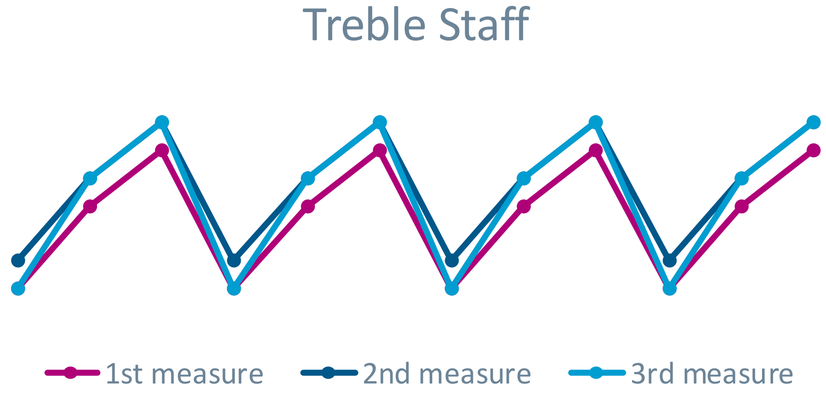
複数文字列に対するデカルト木円形パターン照合の索引




Treble Staff

Eric Osterkamp and Dominik Köppl

”Extending the Burrows-Wheeler Transform for Cartesian Tree Matching and Constructing It. ” CPM 2025: 26:1-26:17

背景

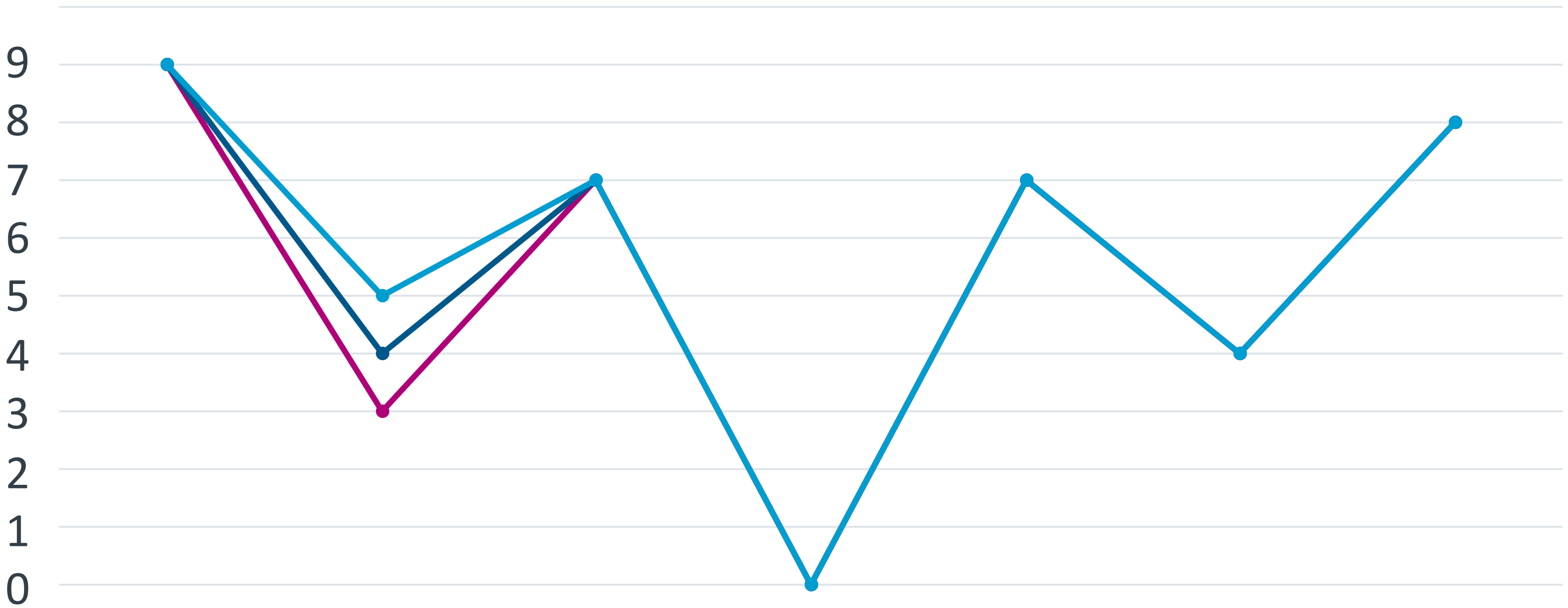
- 株価チャートの反転パターン

Inverse Head & Shoulder [Fu+ ‘07]

背景

- 順序保存照合 (Order-Preserving Matching) [Kim+ ‘14] の時には厳しすぎる

Inverse Head & Shoulder [Fu+ ‘07]



9370748 ≠ 9470748 ≠ 9570748

デカルト木照合

[Park+ ‘20]

	1	2	3	4	5	6
$T =$	b	a	n	a	n	a

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]

$T =$

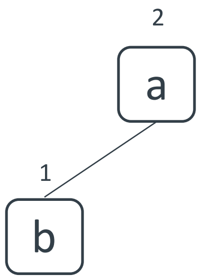
1	2	3	4	5	6
b	a	n	a	n	a

2
a

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]



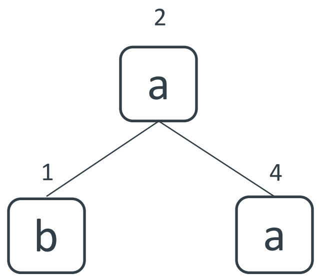
$T =$

1	2	3	4	5	6
b	a	n	a	n	a

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]



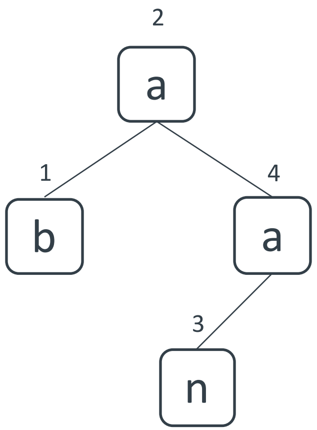
$T =$

1	2	3	4	5	6
b	a	n	a	n	a

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]



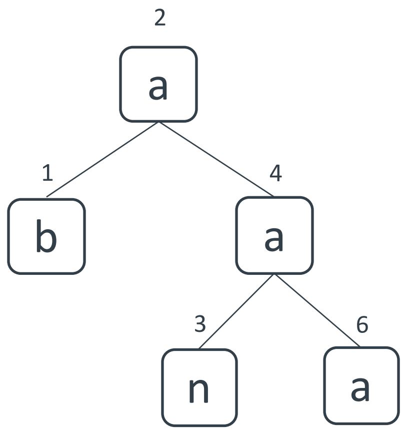
T =

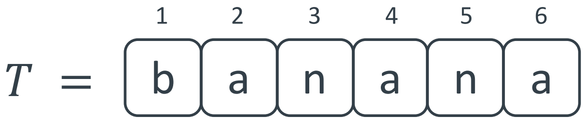
1	2	3	4	5	6
b	a	n	a	n	a

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]

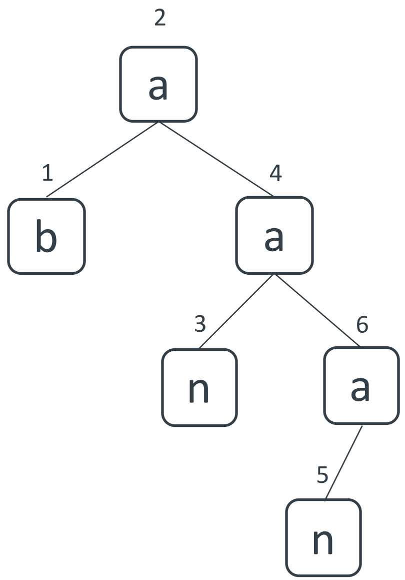




T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]



$T =$

1	2	3	4	5	6
b	a	n	a	n	a

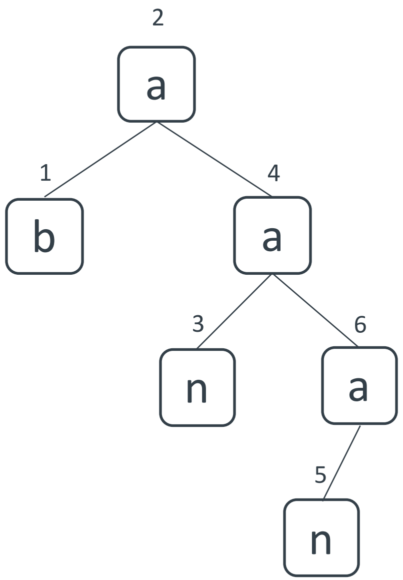
T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\text{ct}} S$)

デカルト木照合

[Park+ ‘20]

$T =$

1	2	3	4	5	6
b	a	n	a	n	a

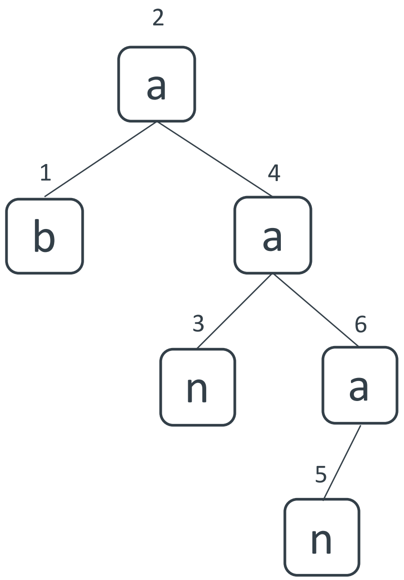


$\mathrm{ct}(T)$

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]



$\mathrm{ct}(T)$

	1	2	3	4	5	6
$T =$	b	a	n	a	n	a
$S =$	p	a	p	a	y	a

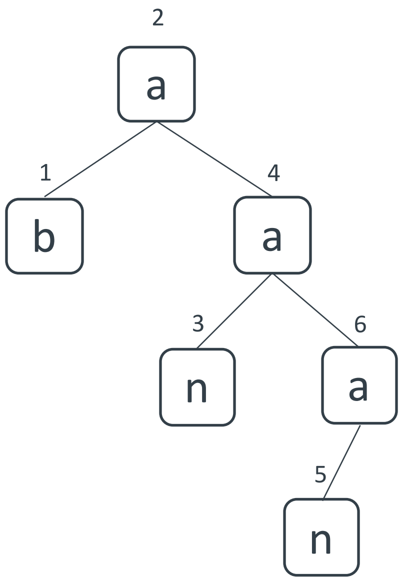
T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

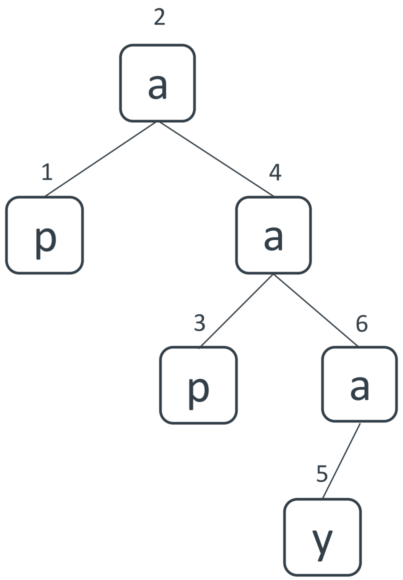
[Park+ ‘20]

	1	2	3	4	5	6
$T =$	b	a	n	a	n	a

	1	2	3	4	5	6
$S =$	p	a	p	a	y	a



$\mathrm{ct}(T)$

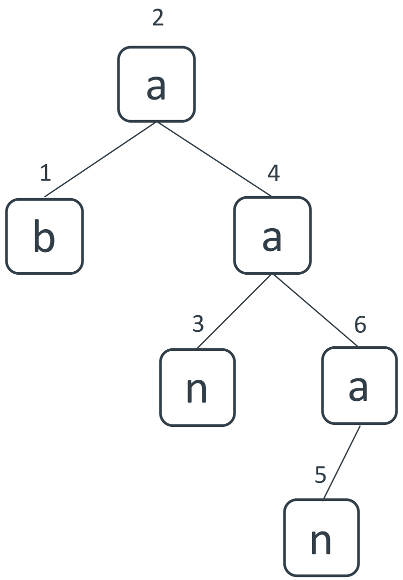


$\mathrm{ct}(S)$

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

[Park+ ‘20]

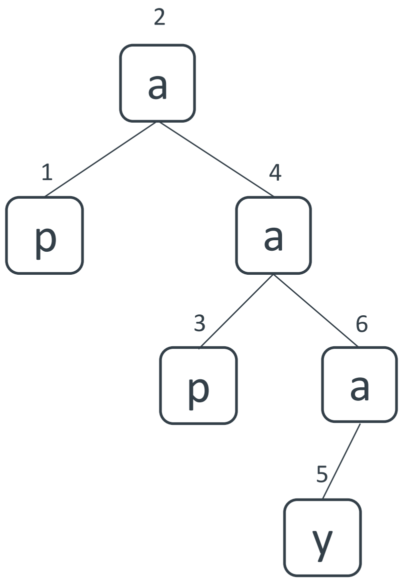


$\mathrm{ct}(T)$

	1	2	3	4	5	6
T =	b	a	n	a	n	a

	1	2	3	4	5	6
S =	p	a	p	a	y	a

$$T =_{\mathrm{ct}} S$$

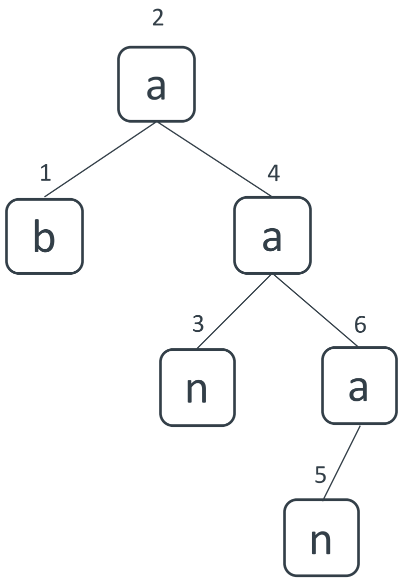


$\mathrm{ct}(S)$

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

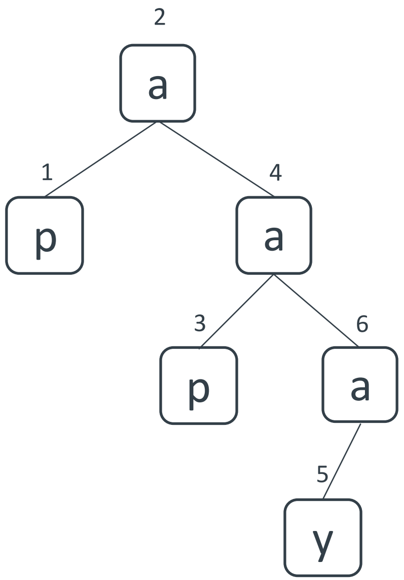
[Park+ ‘20]



ct(T)

	1	2	3	4	5	6
T =	b	a	n	a	n	a
S =	p	a	p	a	y	a
R =	l	y	c	h	e	e

$T =_{\mathrm{ct}} S$

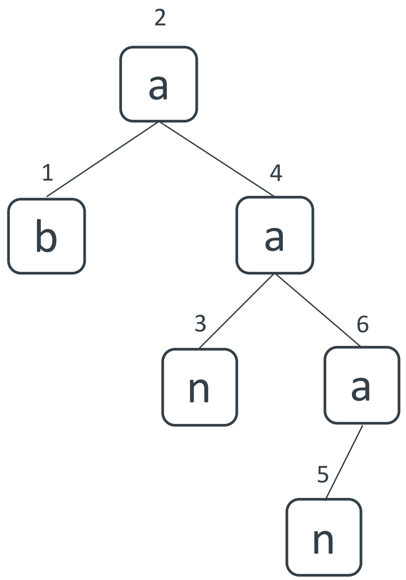


ct(S)

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合

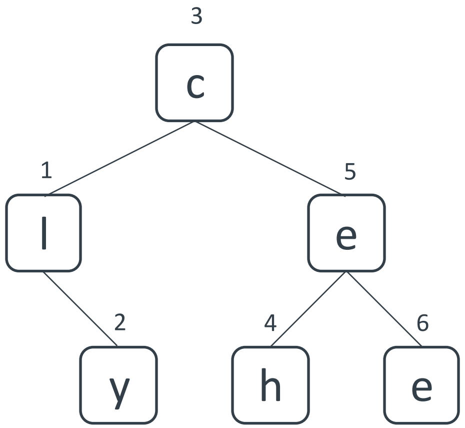
[Park+ ‘20]



$\text{ct}(T)$

	1	2	3	4	5	6
$T =$	b	a	n	a	n	a
$S =$	p	a	p	a	y	a
$R =$	l	y	c	h	e	e

$$T =_{\text{ct}} S$$

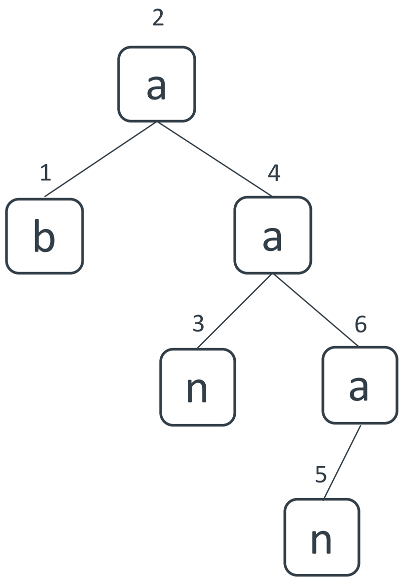


$\text{ct}(R)$

T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\text{ct}} S$)

デカルト木照合

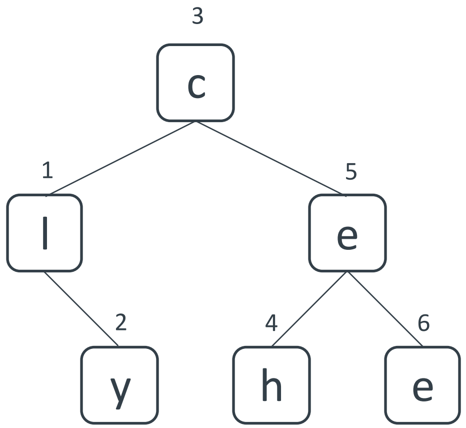
[Park+ ‘20]



ct(T)

	1	2	3	4	5	6
T =	b	a	n	a	n	a
S =	p	a	p	a	y	a
R =	l	y	c	h	e	e

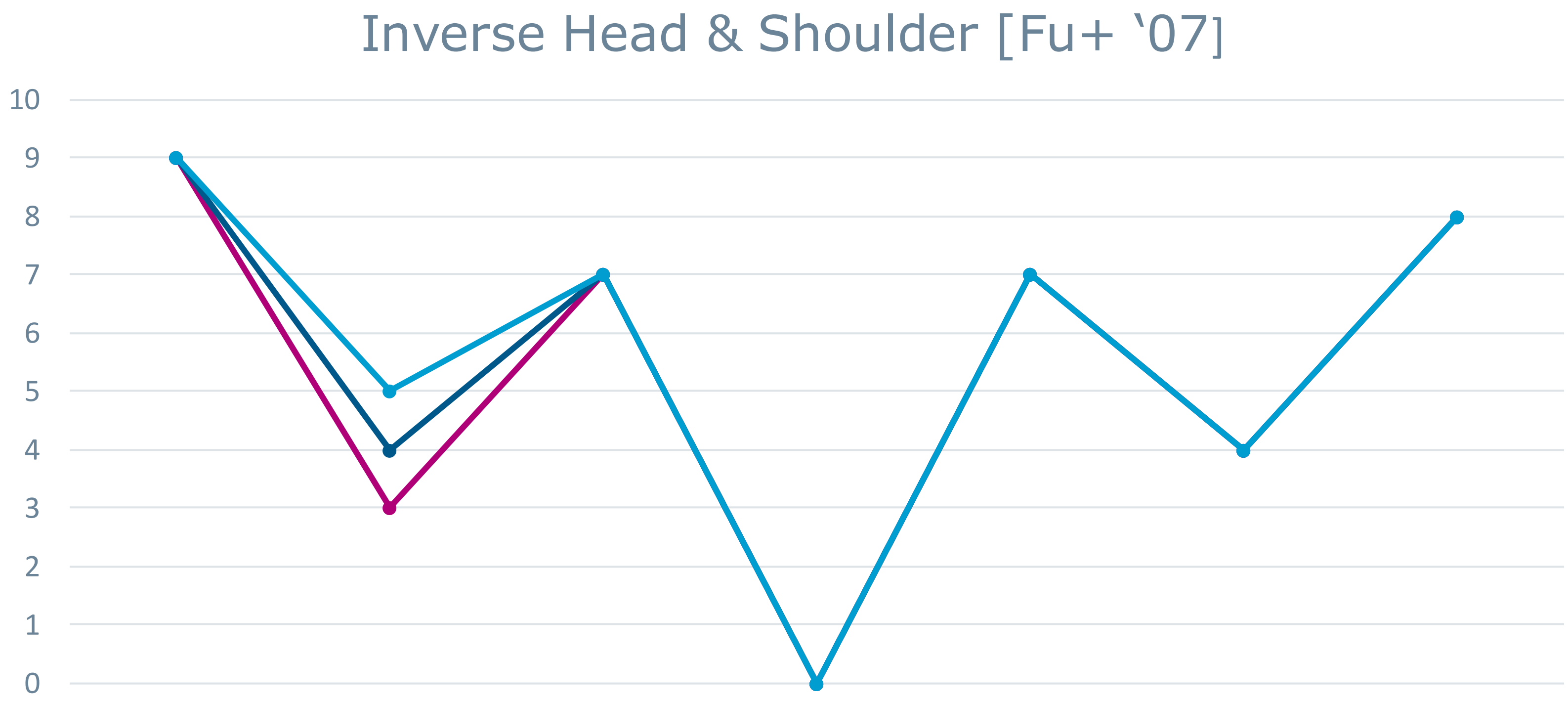
$$T =_{\mathrm{ct}} S \neq_{\mathrm{ct}} R$$



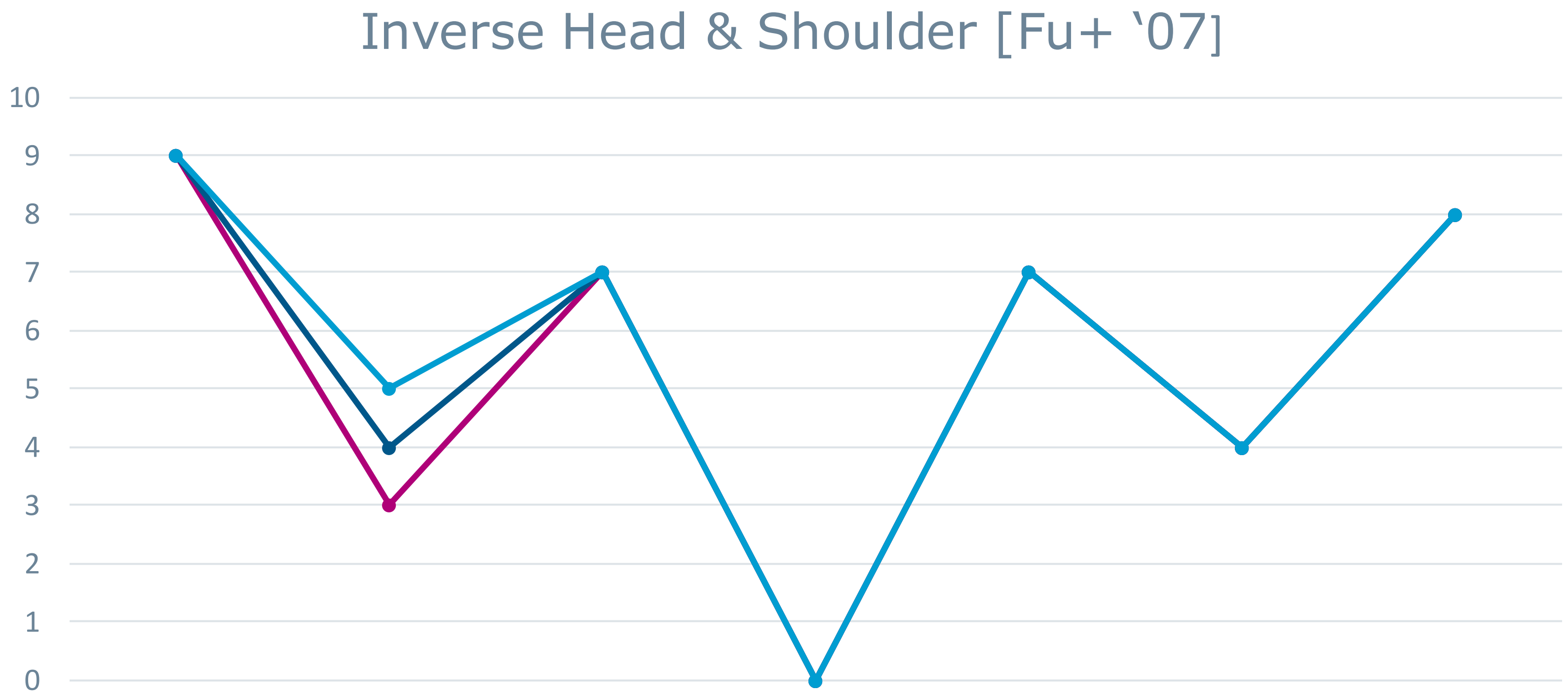
ct(R)

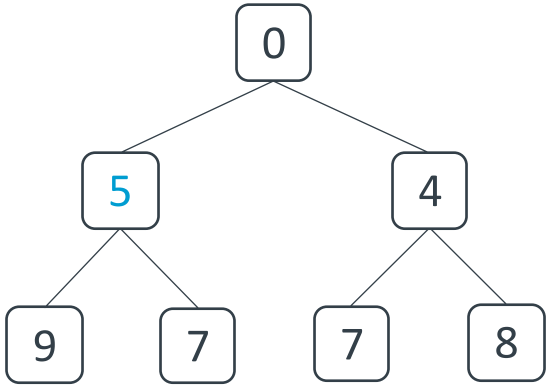
T と S のデカルト木は一致する場合、T と S は ct マッチすると言う($T =_{\mathrm{ct}} S$)

デカルト木照合



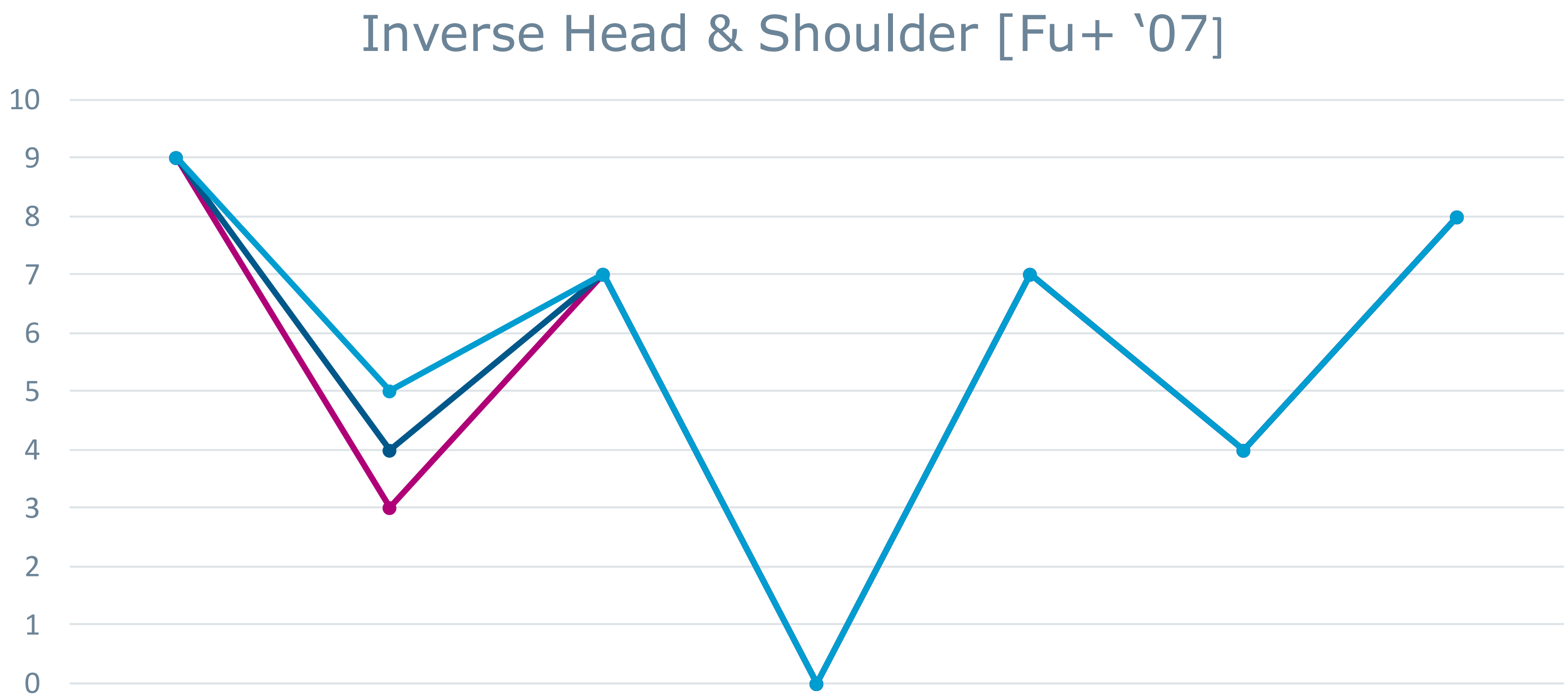
デカルト木照合

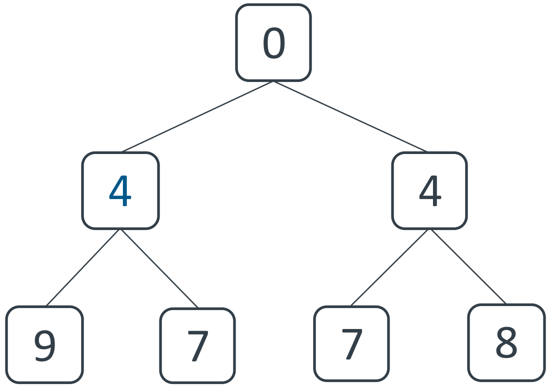




ct(9570748)

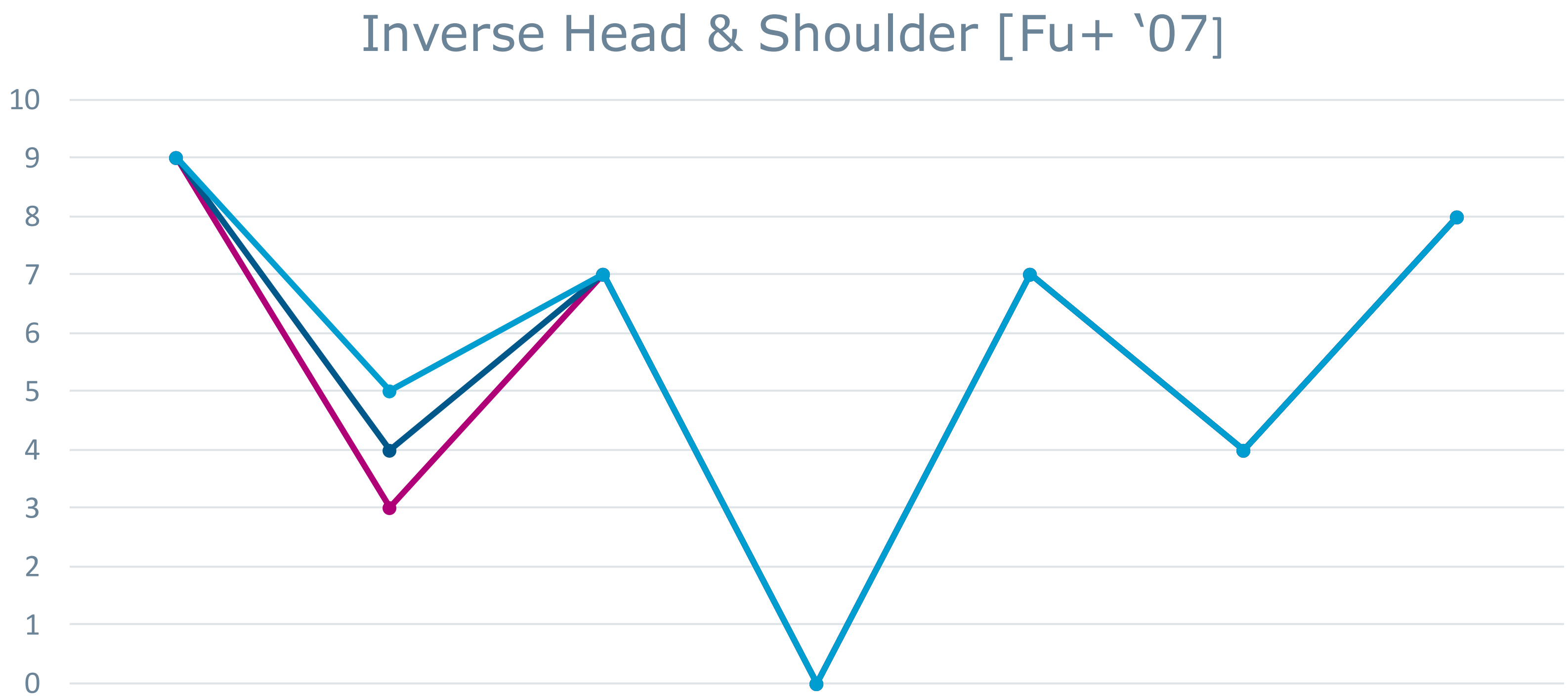
デカルト木照合

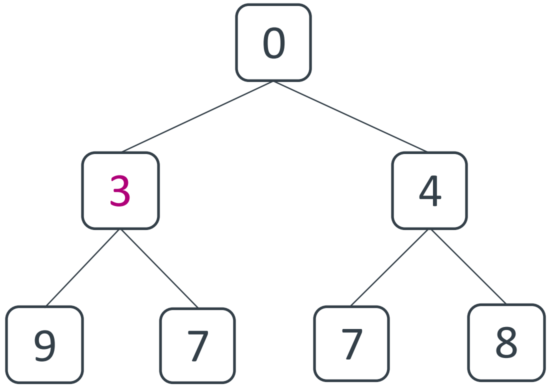




ct(9470748)

デカルト木照合





ct(9370748)

デカルト木照合

- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる

デカルト木照合

- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる

$P =$

k	i	w	i

$T =$

1	2	3	4	5	6
b	a	n	a	n	a

デカルト木照合

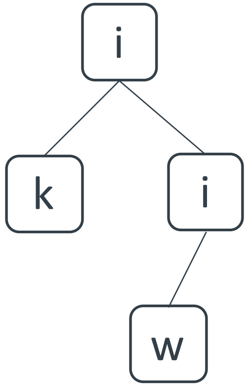
- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる

P =

k	i	w	i

T =

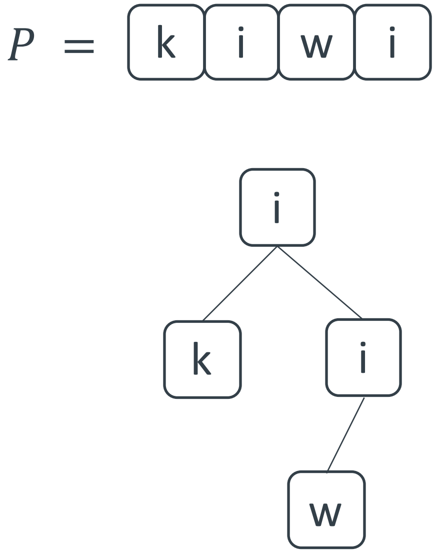
1	2	3	4	5	6
b	a	n	a	n	a



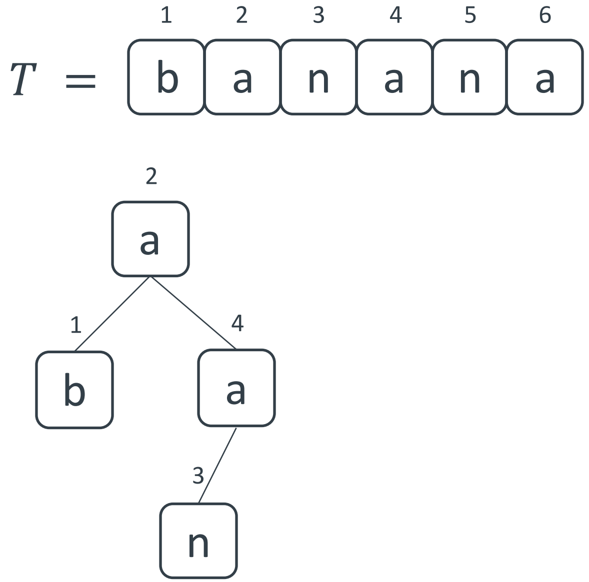
$\mathrm{ct}(P)$

デカルト木照合

- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる



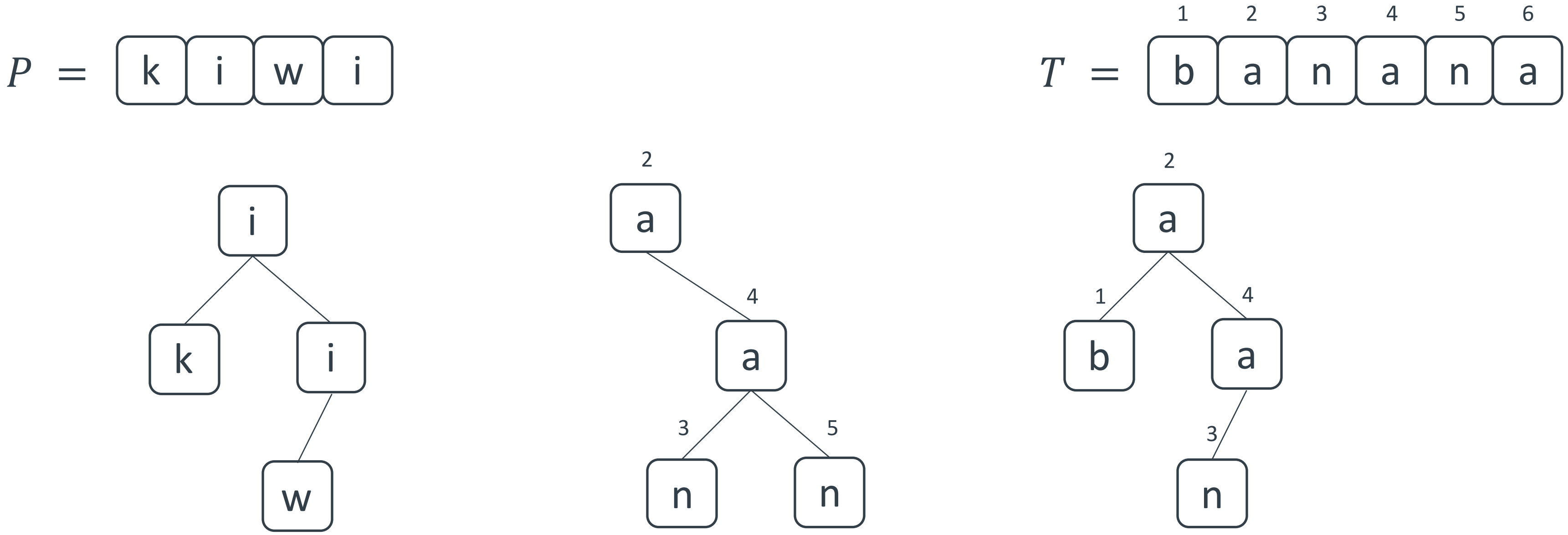
$\mathrm{ct}(P)$



$\mathrm{ct}(T[1..4])$

デカルト木照合

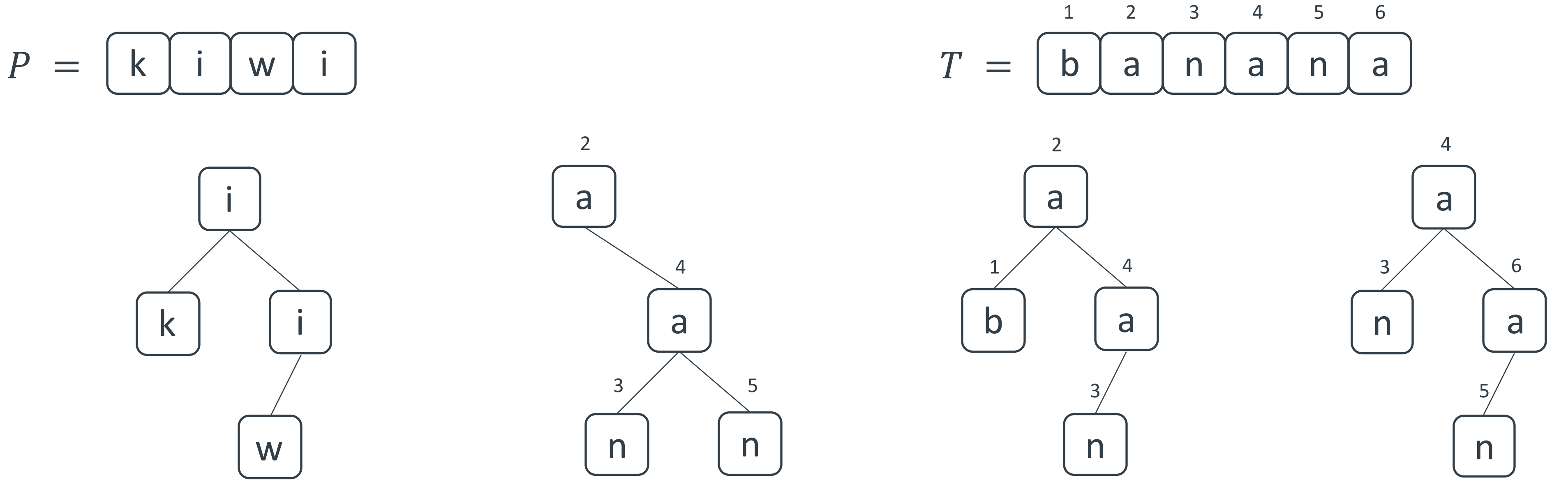
- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる



$\mathrm{ct}(P)$ $\mathrm{ct}(T[2..5])$ $\mathrm{ct}(T[1..4])$

デカルト木照合

- パターン P は出現するとは、 T の部分文字列が P と ct マッチする
- CTPM: T 中の P のすべての出現を数え上げる



ct(P) ct($T[2..5]$) ct($T[1..4]$) ct($T[3..6]$)

デカルト木照合

データ構造	必要なスペース 単位：ビット	建設のための 追加スペース	建設時間	CTPMの時間	参考
Suffix Tree	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$	$O(m \log \sigma)$	[Park+ '20]
Position Heap	$O(n \log n)$	$O(n \log n)$	$O(n \log \sigma)$	$O(m\sigma + m \log m + occ)$	[Nishimoto+ '21]
FM-Index (BWT)	$3n + o(n)$	$O(n \log n)$	$O(n \log n)$	$O(m)$	[Kim and Cho '21]

- σ アルファベットサイズ
- occ 出現の総数
- n テキストの合計長
- m パターンの長さ

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です

T =

1	2	3	4	5	6
b	a	n	a	n	a

親距離符号化

[Park+ '20]

- ∞ は、任意の整数よりも大きい特殊記号です

$T =$

1	2	3	4	5	6
b	a	n	a	n	a

$\langle T \rangle =$

1
∞

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です

$T =$

1	2	3	4	5	6
b	a	n	a	n	a

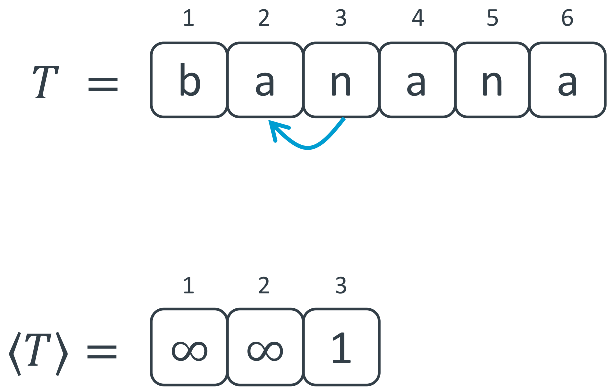
$\langle T \rangle =$

1	2
∞	∞

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



$T =$

1	2	3	4	5	6
b	a	n	a	n	a

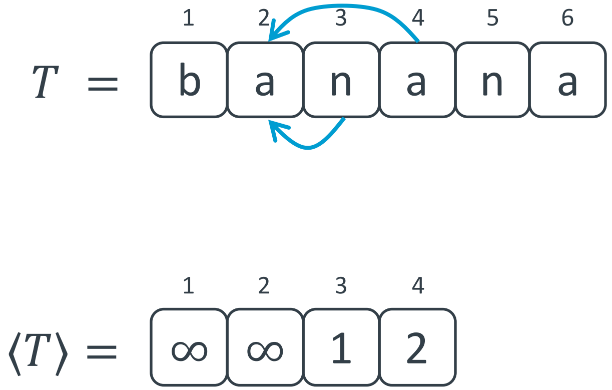
$\langle T \rangle =$

1	2	3
∞	∞	1

親距離符号化

[Park+ ‘20]

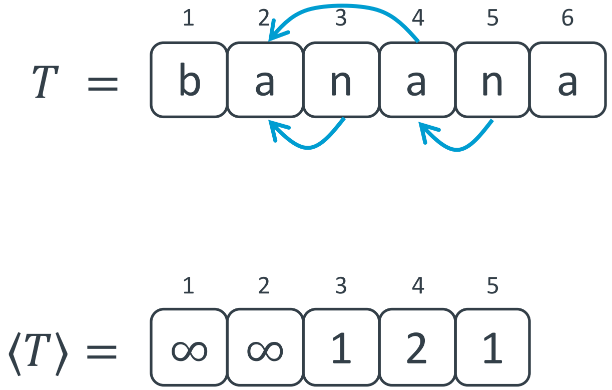
- ∞ は、任意の整数よりも大きい特殊記号です



親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



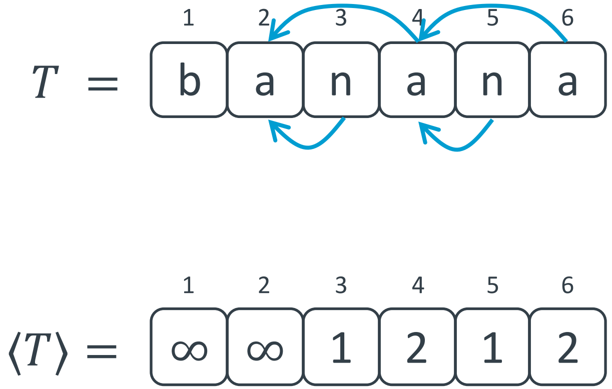
	1	2	3	4	5	6
T =	b	a	n	a	n	a

	1	2	3	4	5
$\langle T \rangle$ =	∞	∞	1	2	1

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



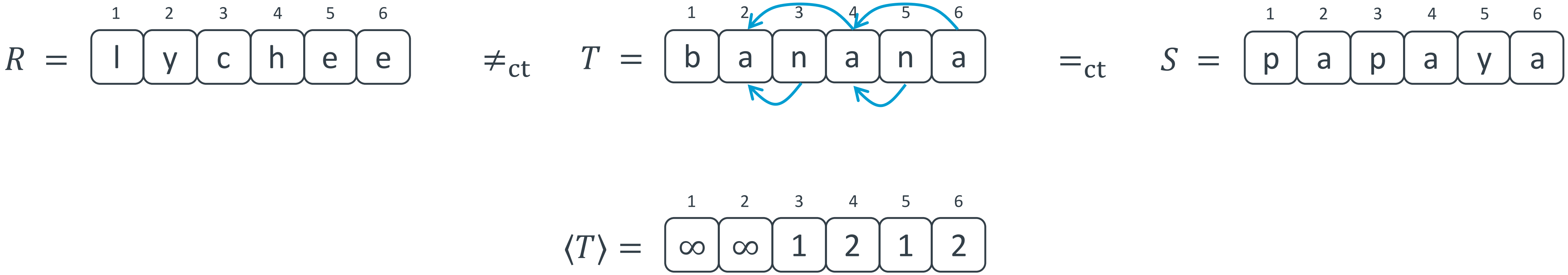
	1	2	3	4	5	6
$T =$	b	a	n	a	n	a

	1	2	3	4	5	6
$\langle T \rangle =$	∞	∞	1	2	1	2

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



	1	2	3	4	5	6
$R =$	l	y	c	h	e	e

$\neq_{\mathrm{ct}}$

	1	2	3	4	5	6
$T =$	b	a	n	a	n	a

$=_{\mathrm{ct}}$

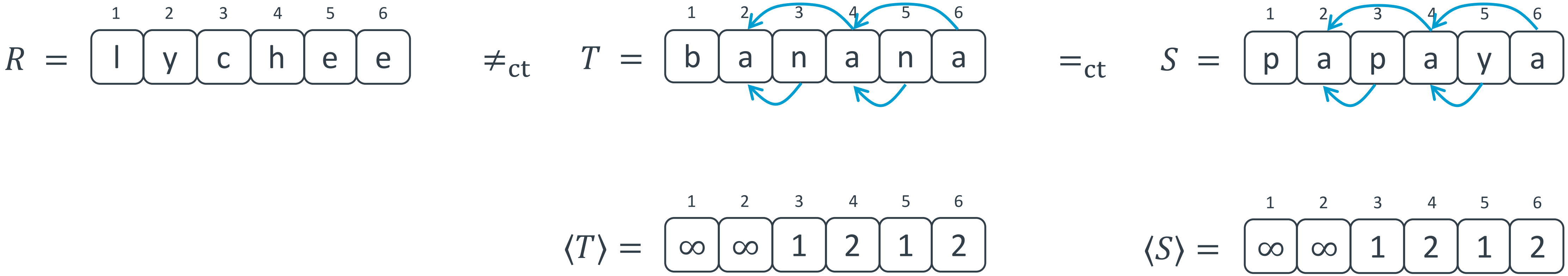
	1	2	3	4	5	6
$S =$	p	a	p	a	y	a

	1	2	3	4	5	6
$\langle T \rangle =$	∞	∞	1	2	1	2

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



$R = \boxed{l}\boxed{y}\boxed{c}\boxed{h}\boxed{e}\boxed{e} \neq_{\mathrm{ct}} T = \boxed{b}\boxed{a}\boxed{n}\boxed{a}\boxed{n}\boxed{a} =_{\mathrm{ct}} S = \boxed{p}\boxed{a}\boxed{p}\boxed{a}\boxed{y}\boxed{a}$

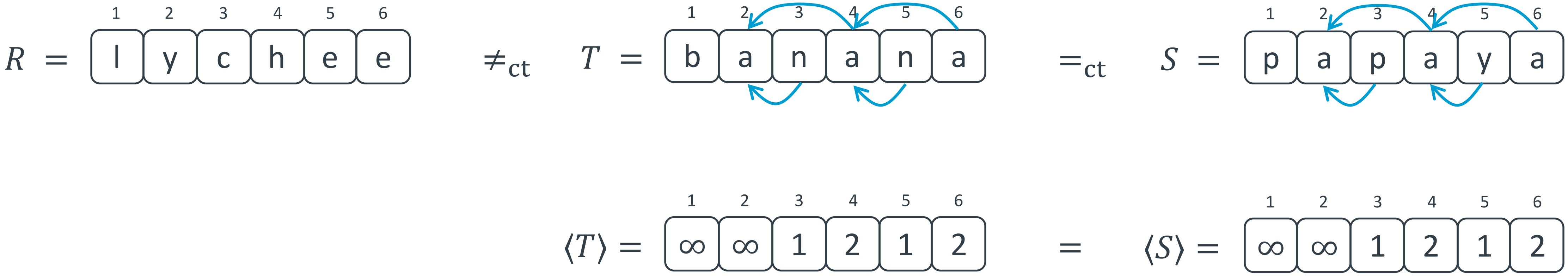
$\langle T \rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{2}$

$\langle S \rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{2}$

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



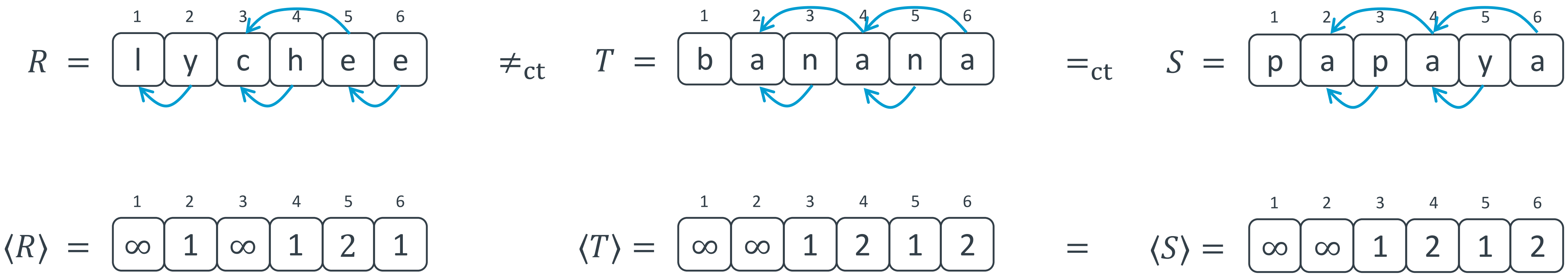
$$R = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \boxed{\text{l}} & \boxed{\text{y}} & \boxed{\text{c}} & \boxed{\text{h}} & \boxed{\text{e}} & \boxed{\text{e}} \end{array} \quad \neq_{\text{ct}} \quad T = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \boxed{\text{b}} & \boxed{\text{a}} & \boxed{\text{n}} & \boxed{\text{a}} & \boxed{\text{n}} & \boxed{\text{a}} \end{array} \quad =_{\text{ct}} \quad S = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \boxed{\text{p}} & \boxed{\text{a}} & \boxed{\text{p}} & \boxed{\text{a}} & \boxed{\text{y}} & \boxed{\text{a}} \end{array}$$

$$\langle T \rangle = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \boxed{\infty} & \boxed{\infty} & \boxed{1} & \boxed{2} & \boxed{1} & \boxed{2} \end{array} \quad = \quad \langle S \rangle = \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \\ \boxed{\infty} & \boxed{\infty} & \boxed{1} & \boxed{2} & \boxed{1} & \boxed{2} \end{array}$$

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



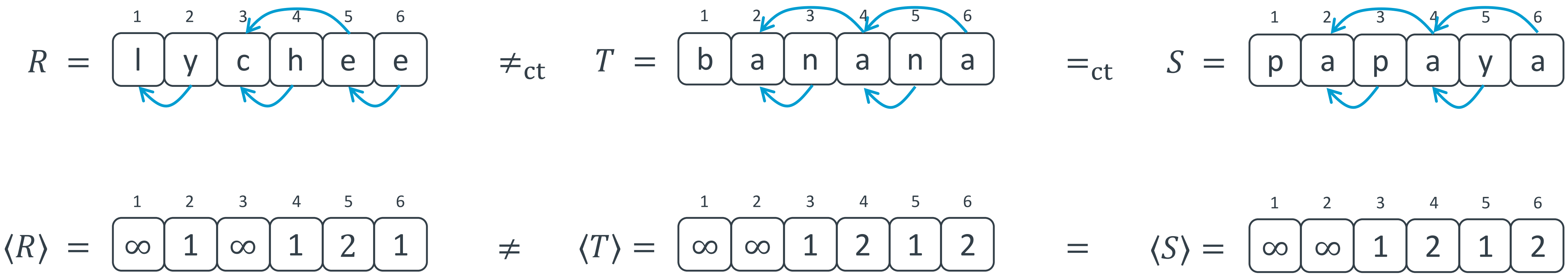
$$R = \boxed{\mathtt{l}}\boxed{\mathtt{y}}\boxed{\mathtt{c}}\boxed{\mathtt{h}}\boxed{\mathtt{e}}\boxed{\mathtt{e}} \quad \neq_{\mathrm{ct}} \quad T = \boxed{\mathtt{b}}\boxed{\mathtt{a}}\boxed{\mathtt{n}}\boxed{\mathtt{a}}\boxed{\mathtt{n}}\boxed{\mathtt{a}} \quad =_{\mathrm{ct}} \quad S = \boxed{\mathtt{p}}\boxed{\mathtt{a}}\boxed{\mathtt{p}}\boxed{\mathtt{a}}\boxed{\mathtt{y}}\boxed{\mathtt{a}}$$

$$\langle R\rangle = \boxed{\infty}\boxed{1}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1} \qquad \langle T\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{2} \quad = \quad \langle S\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{2}$$

親距離符号化

[Park+ ‘20]

- ∞ は、任意の整数よりも大きい特殊記号です



$R =$ l y c h e e $\neq_{\mathrm{ct}}$ $T =$ b a n a n a $=_{\mathrm{ct}}$ $S =$ p a p a y a

$\langle R \rangle = \infty\ 1\ \infty\ 1\ 2\ 1 \neq \langle T \rangle = \infty\ \infty\ 1\ 2\ 1\ 2 = \langle S \rangle = \infty\ \infty\ 1\ 2\ 1\ 2$

円形的な照合

- T を円形的なテキストと見なす
- P は T の境目でも出現可能

円形的な照合

- T を円形的なテキストと見なす
- P は T の境目でも出現可能

$P =$

k	i	w	i

$T =$

1	2	3	4	5	6
b	a	n	a	n	a

円形的な照合

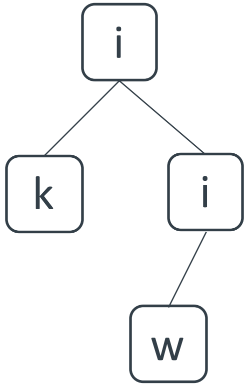
- T を円形的なテキストと見なす
- P は T の境目でも出現可能

P =

k	i	w	i

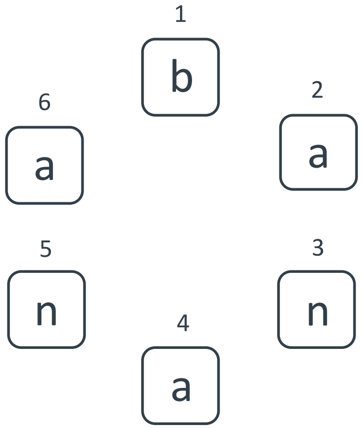
T =

1	2	3	4	5	6
b	a	n	a	n	a



$\mathrm{ct}(P)$

T =



円形的な照合

- T を円形的なテキストと見なす
- P は T の境目でも出現可能

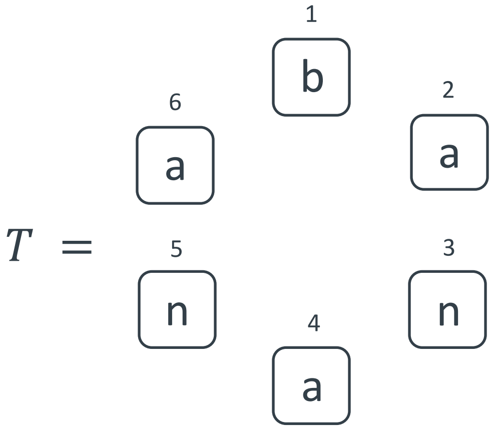
P =

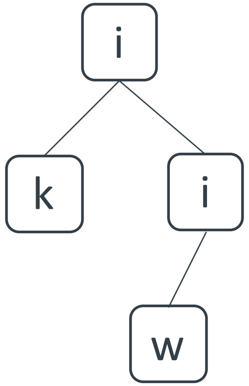
k	i	w	i

T =

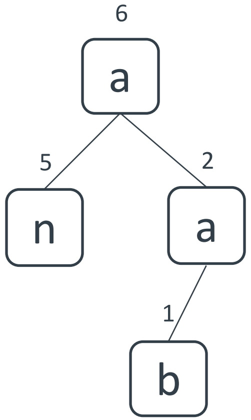
1	2	3	4	5	6
b	a	n	a	n	a

T =





$\mathrm{ct}(P)$



$\mathrm{ct}(T[5..6] \cdot T[1..2])$

問題の説明

- 回転文字列：文字列 T の任意のサイクリックシフト $T[i..n]T[1..i-1]$
- $T_1, \ldots, T_d$のすべての回転文字列と $T = T_1 \cdot \ldots \cdot T_d$中の開始位置を同一視する
- $X^\omega := X \cdot X \cdot X \cdot \ldots$ と定義し、 P でパターンを表現する
- 円形 CTPM: $\mathrm{C}(i)^\omega[..|P|] =_{\mathrm{ct}} P$ が成り立つような $T_1, \ldots, T_d$のすべての回転文字列を数え上げる

問題の説明

- 回転文字列：文字列 T の任意のサイクリックシフト $T[i..n]T[1..i-1]$
- $T_1, \ldots, T_d$のすべての回転文字列と $T = T_1 \cdot \ldots \cdot T_d$中の開始位置を同一視する

	1	2	3		4	5	6	7		8	9	10	11	12
$T =$	f	i	g	·	k	i	w	i	·	a	p	p	l	e

- $X^{\omega} := X \cdot X \cdot X \cdot \ldots$ と定義し、 P でパターンを表現する
- 円形 CTPM: $\mathrm{C}(i)^{\omega}[..|P|] =_{\mathrm{ct}} P$ が成り立つような $T_1, \ldots, T_d$のすべての回転文字列を数え上げる

問題の説明

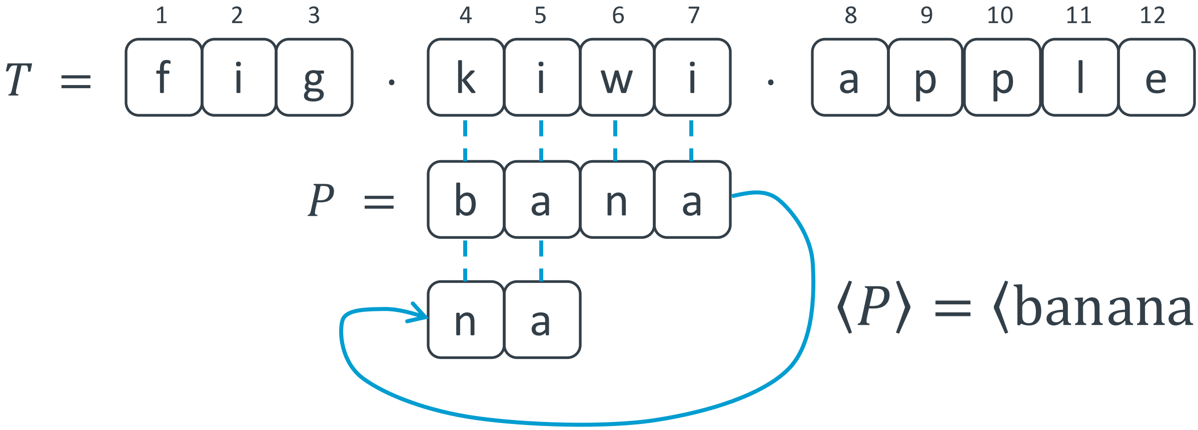
- 回転文字列：文字列 T の任意のサイクリックシフト $T[i..n]T[1..i-1]$
- $T_1, \ldots, T_d$のすべての回転文字列と $T = T_1 \cdot \ldots \cdot T_d$中の開始位置を同一視する

$$T = \overset{1}{\boxed{\mathrm{f}}}\overset{2}{\boxed{\mathrm{i}}}\overset{3}{\boxed{\mathrm{g}}} \cdot \overset{4}{\boxed{\mathrm{k}}}\overset{5}{\boxed{\mathrm{i}}}\overset{6}{\boxed{\mathrm{w}}}\overset{7}{\boxed{\mathrm{i}}} \cdot \overset{8}{\boxed{\mathrm{a}}}\overset{9}{\boxed{\mathrm{p}}}\overset{10}{\boxed{\mathrm{p}}}\overset{11}{\boxed{\mathrm{l}}}\overset{12}{\boxed{\mathrm{e}}} \Rightarrow \mathrm{C}(6) = \mathrm{wiki}$$

- $X^{\omega} \coloneqq X \cdot X \cdot X \cdot \ldots$ と定義し、 P でパターンを表現する
- 円形 CTPM: $\mathrm{C}(i)^{\omega}[..|P|] =_{\mathrm{ct}} P$ が成り立つような $T_1, \ldots, T_d$のすべての回転文字列を数え上げる

問題の説明

- 回転文字列：文字列 T の任意のサイクリックシフト $T[i..n]T[1..i-1]$
- $T_1, \ldots, T_d$のすべての回転文字列と $T = T_1 \cdot \ldots \cdot T_d$中の開始位置を同一視する



$\Rightarrow \mathrm{C}(6) = \mathrm{wiki}$

$\langle P \rangle = \langle \mathrm{banana} \rangle = \infty\infty 1212 = \langle \mathrm{kiwiki} \rangle = \langle \mathrm{C}(4)^{\omega}[..6] \rangle$

- $X^{\omega} \coloneqq X \cdot X \cdot X \cdot \ldots$ と定義し、 P でパターンを表現する
- 円形 CTPM: $\mathrm{C}(i)^{\omega}[..|P|] =_{\mathrm{ct}} P$ が成り立つような $T_1, \ldots, T_d$のすべての回転文字列を数え上げる

デカルト木照合

データ構造	必要なスペース 単位：ビット	建設のための 追加スペース	建設時間	CTPMの時間	参考
Suffix Tree	$O(n \log n)$	$O(n \log n)$	$O(n \log n)$	$O(m \log \sigma)$	[Park+ '20]
Position Heap	$O(n \log n)$	$O(n \log n)$	$O(n \log \sigma)$	$O(m\sigma + m \log m + occ)$	[Nishimoto+ '21]
FM-Index (BWT)	$3n + o(n)$	$O(n \log n)$	$O(n \log n)$	$O(m)$	[Kim and Cho '21]
FM-Index (eBWT)	$O(n \log \sigma)$	$O(n \log \sigma)$	$O\left(n \frac{\log \sigma \log n}{\log \log n}\right)$	$O\left(m \frac{\log \sigma \log n}{\log \log n}\right)$	本論
FM-Index (eBWT)	$3n + o(n)$	$O(n \log \sigma)$	$O\left(n \frac{\log \sigma \log n}{\log \log n}\right)$	$O(m)$	本論

CTPM

円形
CTPM

- σ アルファベットサイズ
- occ 出現の総数
- n テキストの合計長
- m パターンの長さ

回転文字列のソート

- パターンの出現が連続するように回転文字列をソートする
- [Kim and Cho ‘21]のように、符号化された回転文字列を辞書式にソートするのは駄目

	1	2	3		4	5	6	7		8	9	10	11	12
$T =$	f	i	g	·	k	i	w	i	·	a	p	p	l	e

位置で
タイブレーク

i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i) \rangle$
8	apple	∞1134
1	fig	∞12
5	iwik	∞121
7	ikiw	∞121
9	pplea	∞1∞∞∞∞
3	gfi	∞∞1
12	eappl	∞∞113
4	kiwi	∞∞12
6	wiki	∞∞12
2	igf	∞∞∞
11	leapp	∞∞∞11
10	pleap	∞∞∞∞1

回転文字列のソート

- パターンの出現が連続するように回転文字列をソートする
- [Kim and Cho '21]のように、符号化された回転文字列を辞書式にソートするのは駄目

	1	2	3		4	5	6	7		8	9	10	11	12
$T =$	f	i	g	·	k	i	w	i	·	a	p	p	l	e

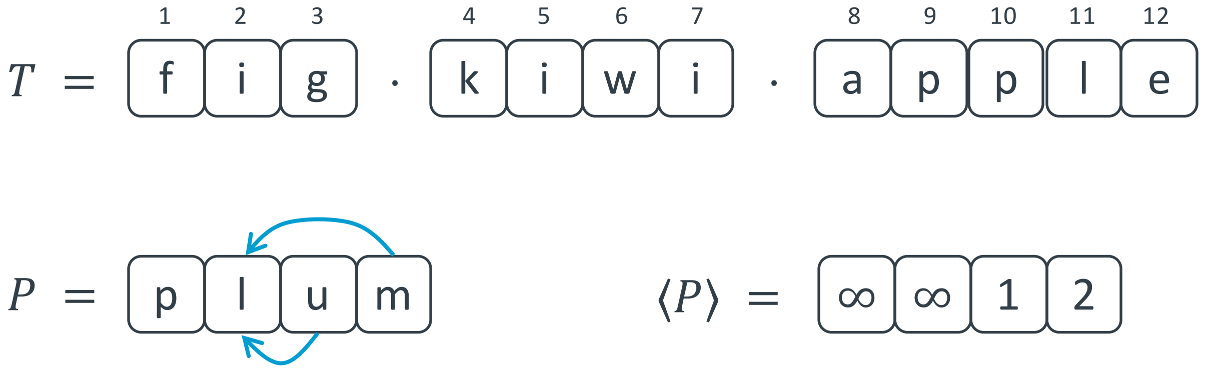
$P =$	p	l	u	m

位置で
タイブレーク

i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i) \rangle$
8	apple	∞1134
1	fig	∞12
5	iwik	∞121
7	ikiw	∞121
9	pplea	∞1∞∞∞
3	gfi	∞∞1
12	eappl	∞∞113
4	kiwi	∞∞12
6	wiki	∞∞12
2	igf	∞∞∞
11	leapp	∞∞∞11
10	pleap	∞∞∞∞1

回転文字列のソート

- パターンの出現が連続するように回転文字列をソートする
- [Kim and Cho '21]のように、符号化された回転文字列を辞書式にソートするのは駄目

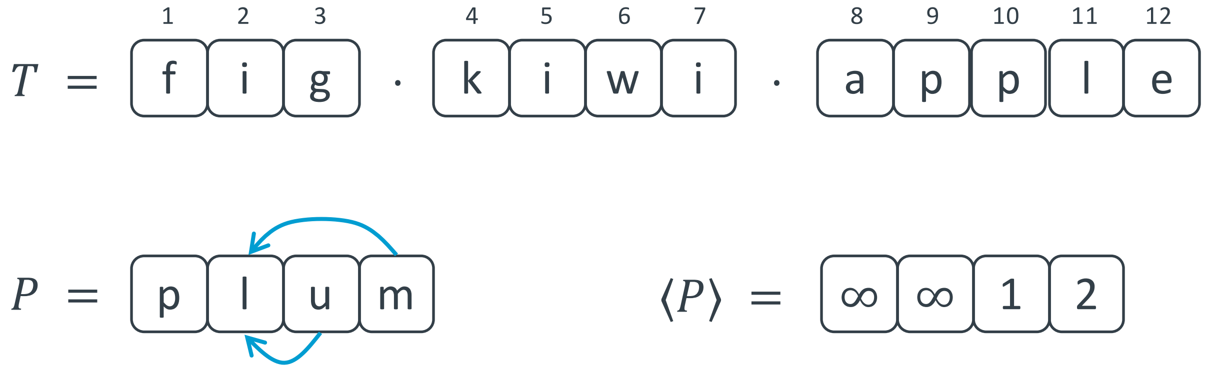


位置で
タイブレーク

i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i) \rangle$
8	apple	∞1134
1	fig	∞12
5	iwik	∞121
7	ikiw	∞121
9	pplea	∞1∞∞∞∞
3	gfi	∞∞1
12	eappl	∞∞113
4	kiwi	∞∞12
6	wiki	∞∞12
2	igf	∞∞∞
11	leapp	∞∞∞11
10	pleap	∞∞∞∞1

回転文字列のソート

- パターンの出現が連続するように回転文字列をソートする
- [Kim and Cho ‘21]のように、符号化された回転文字列を辞書式にソートするのは駄目



位置で
タイブレーク

i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i) \rangle$
8	apple	∞1134
1	fig	∞12
5	iwik	∞121
7	ikiw	∞121
9	pplea	∞1∞∞∞
3	gfig	∞∞12
12	eappl	∞∞113
4	kiwi	∞∞12
6	wiki	∞∞12
2	igf	∞∞∞
11	leapp	∞∞∞11
10	pleap	∞∞∞∞1

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T \rangle_r$ を定義

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義

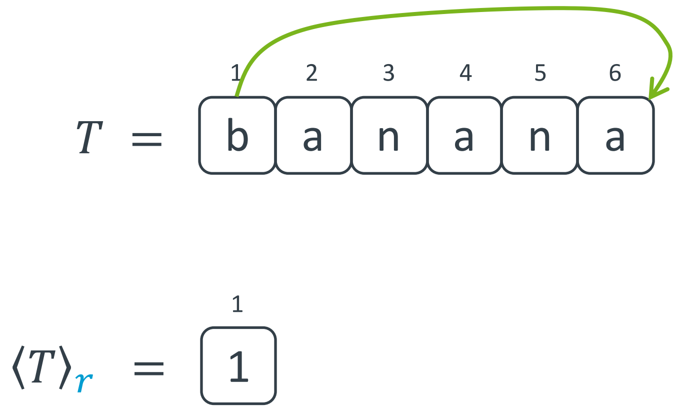
$T =$

1	2	3	4	5	6
b	a	n	a	n	a

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義



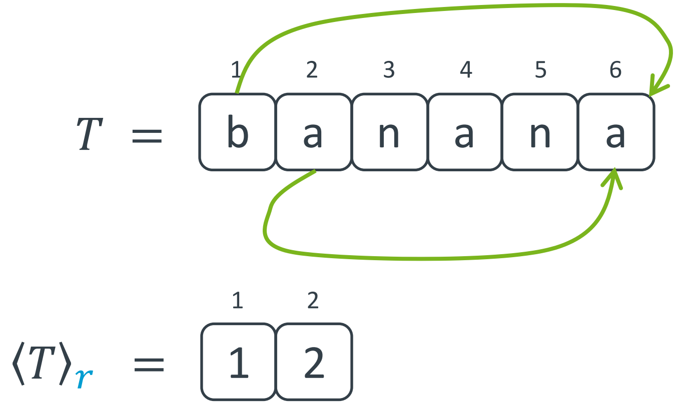
$T =$

$\langle T\rangle_r =$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

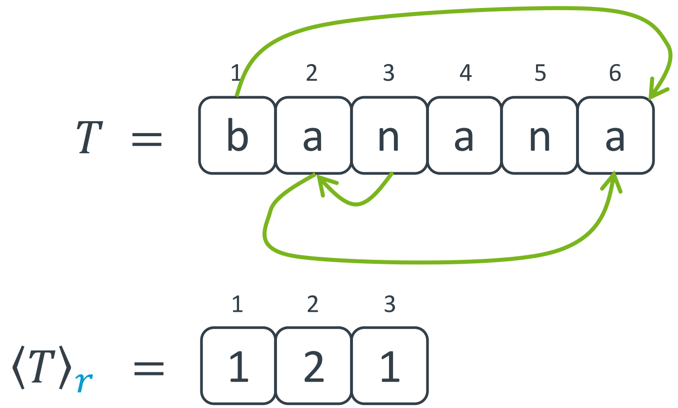
- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T \rangle_r$ を定義



文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

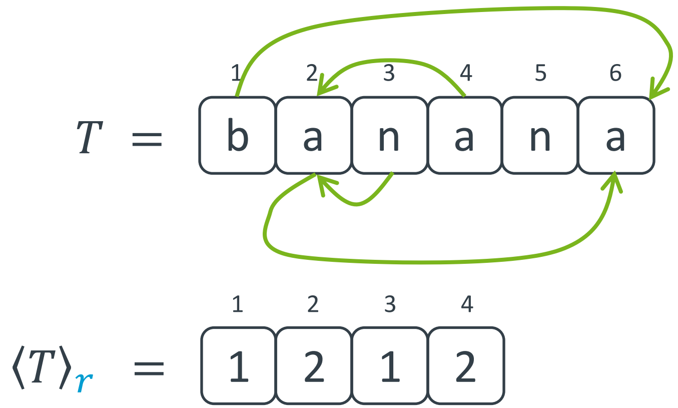
- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T \rangle_r$ を定義



文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

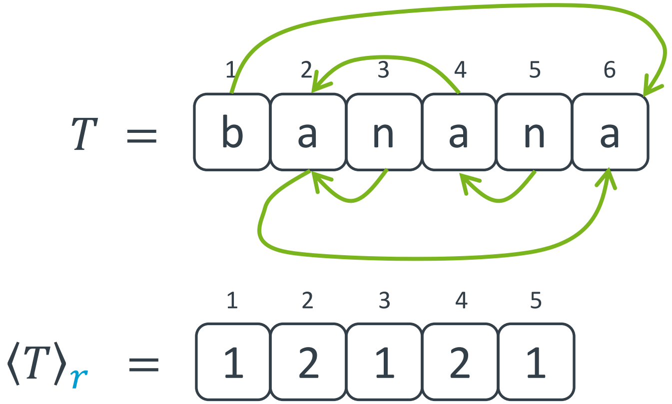
- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義



文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義



T =

1	2	3	4	5	6
b	a	n	a	n	a

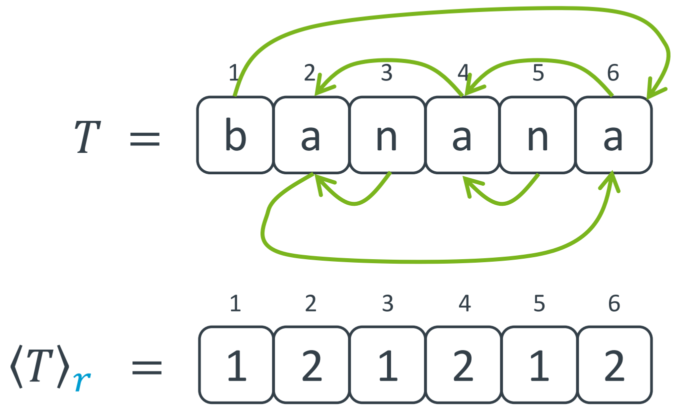
$\langle T\rangle_r$ =

1	2	3	4	5
1	2	1	2	1

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T \rangle_r$ を定義



$T =$

1	2	3	4	5	6
b	a	n	a	n	a

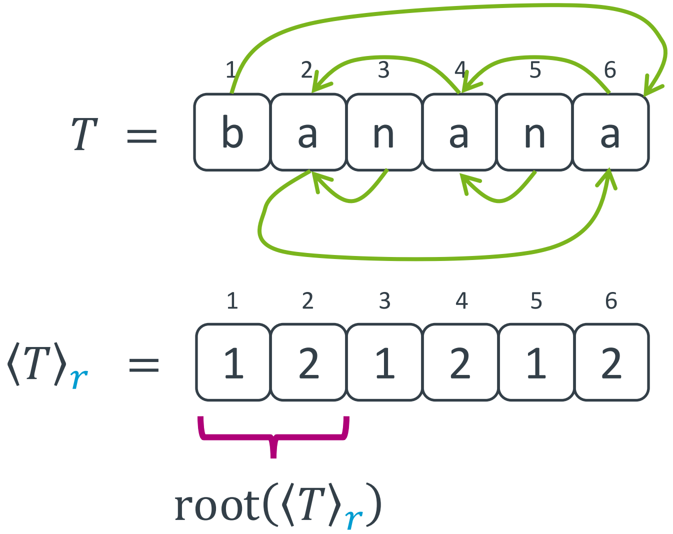
$\langle T \rangle_r =$

1	2	3	4	5	6
1	2	1	2	1	2

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

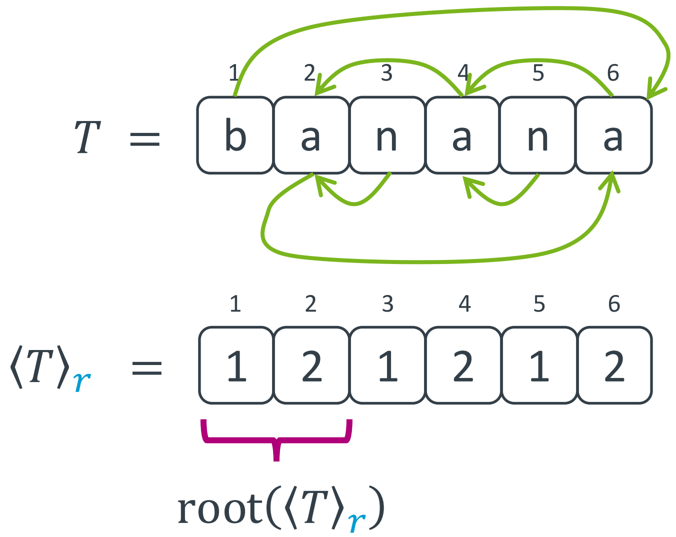
- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T \rangle_r$ を定義



文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義



T = b a n a n a

$\langle T\rangle_r$ = 1 2 1 2 1 2

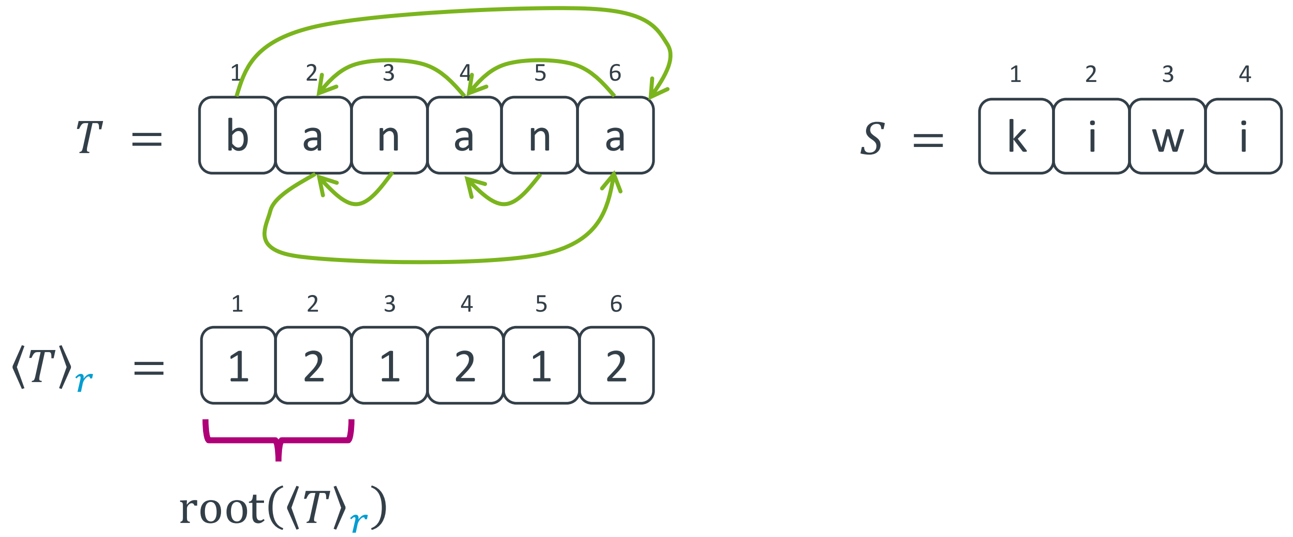
$\underbrace{1\ 2}_{\mathrm{root}(\langle T\rangle_r)}$

- ω 前順序：$X \preccurlyeq_\omega Y \Leftrightarrow \mathrm{root}(\langle X\rangle_r) = \mathrm{root}(\langle Y\rangle_r) \vee \exists i{:}\,\langle X^\omega[..i]\rangle < \langle Y^\omega[..i]\rangle$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義

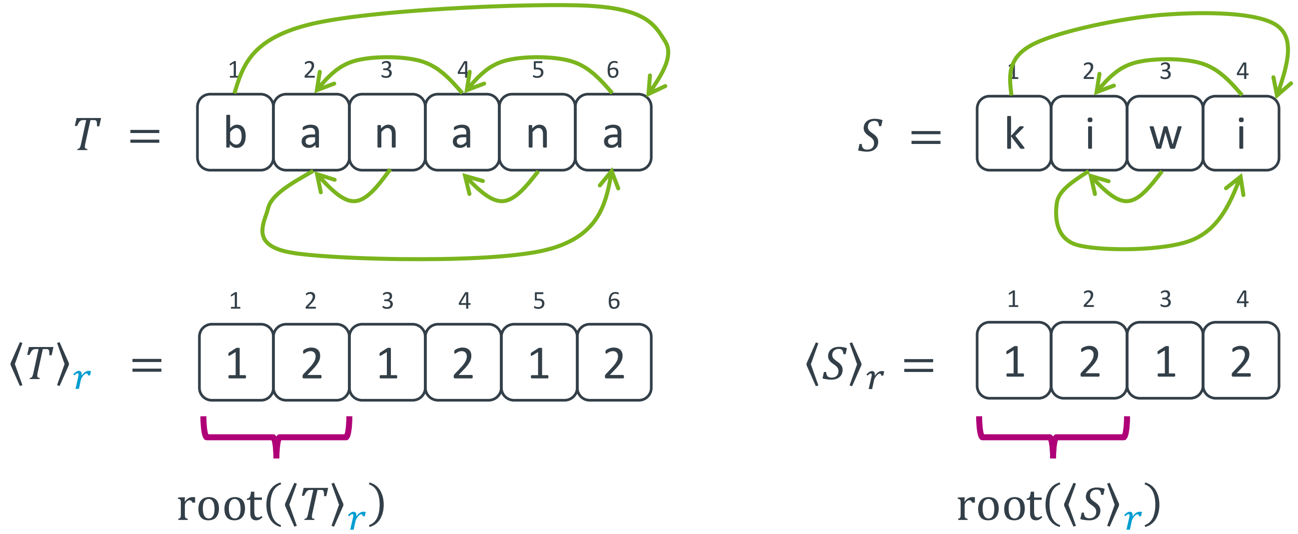


- ω 前順序：$X \preccurlyeq_{\omega} Y \Leftrightarrow \mathrm{root}(\langle X\rangle_r) = \mathrm{root}(\langle Y\rangle_r) \lor \exists i\colon \langle X^{\omega}[..i]\rangle < \langle Y^{\omega}[..i]\rangle$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義

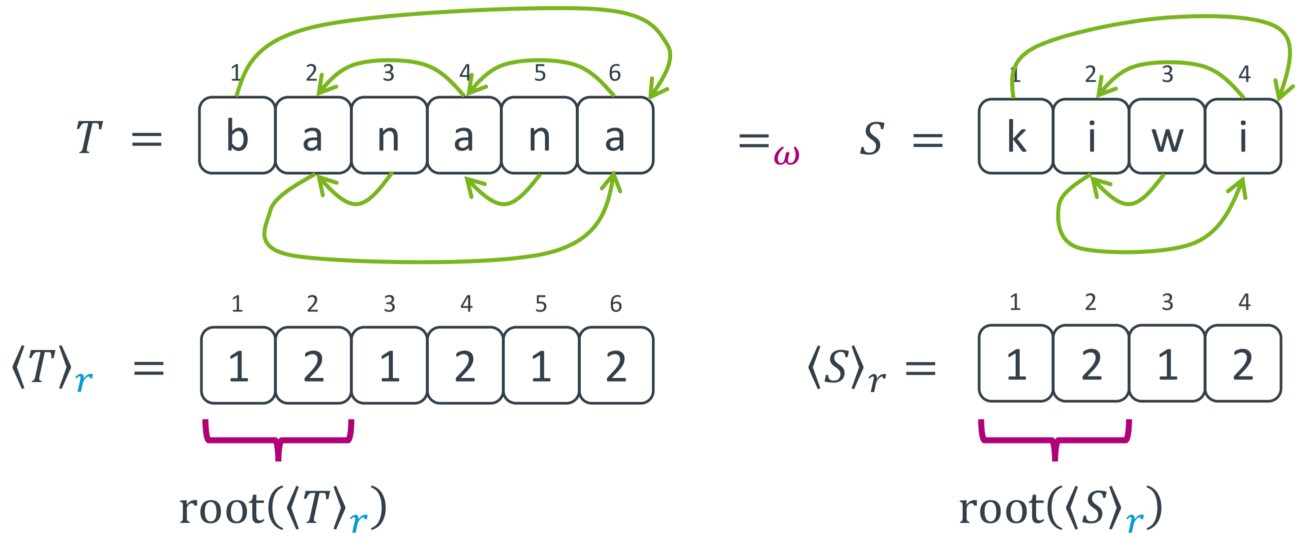


- ω 前順序：$X \preccurlyeq_{\omega} Y \Leftrightarrow \mathrm{root}(\langle X\rangle_r) = \mathrm{root}(\langle Y\rangle_r) \vee \exists i{:}\, \langle X^{\omega}[..i]\rangle < \langle Y^{\omega}[..i]\rangle$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義

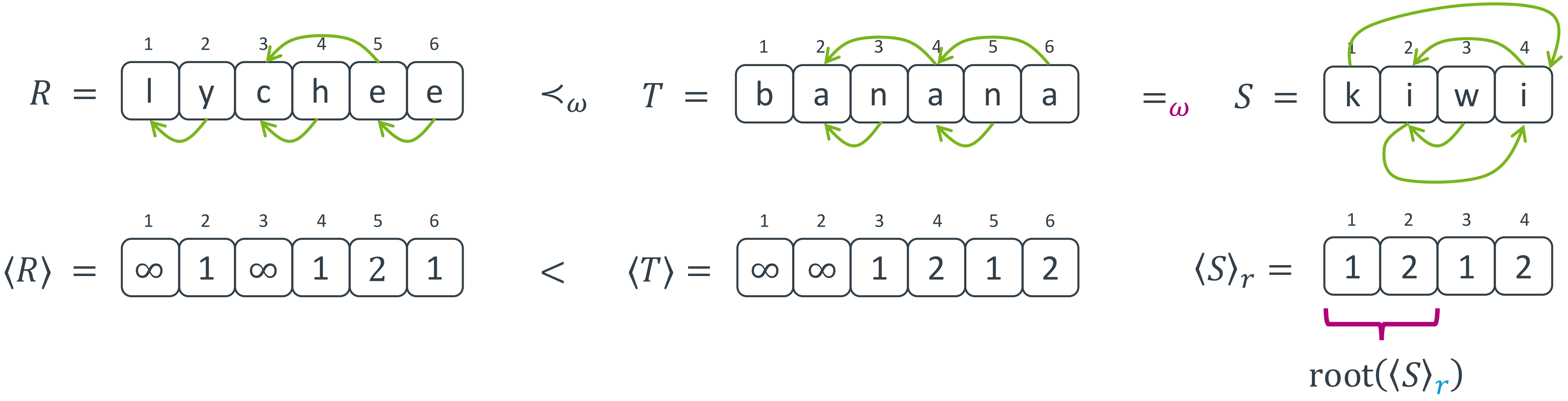


- ω 前順序：$X \preccurlyeq_\omega Y \Leftrightarrow \mathrm{root}(\langle X\rangle_r) = \mathrm{root}(\langle Y\rangle_r) \lor \exists i\colon \langle X^\omega[..i]\rangle < \langle Y^\omega[..i]\rangle$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- 「Mantaci+ '07」によるeBWTへの適応アプローチ
- 文字列 T の回転親距離符号化 $\langle T\rangle_r$ を定義



$R = \boxed{l}\boxed{y}\boxed{c}\boxed{h}\boxed{e}\boxed{e} \prec_\omega T = \boxed{b}\boxed{a}\boxed{n}\boxed{a}\boxed{n}\boxed{a} =_\omega S = \boxed{k}\boxed{i}\boxed{w}\boxed{i}$

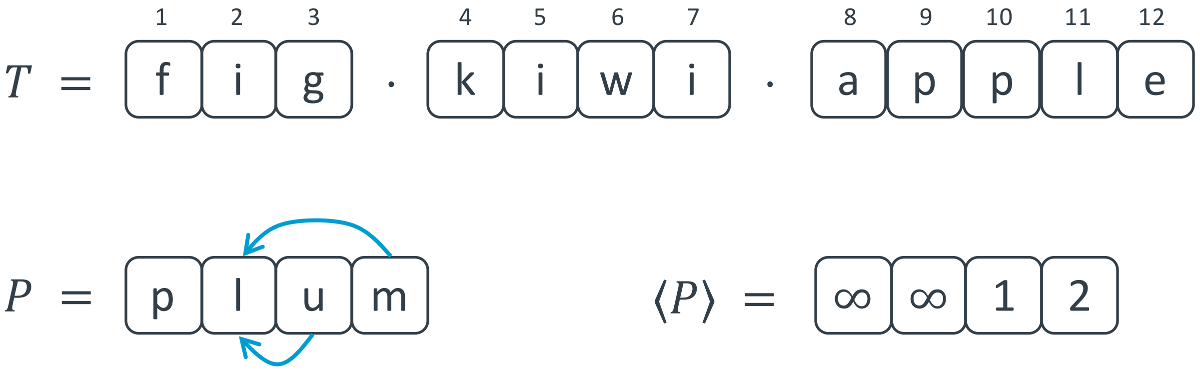
$\langle R\rangle = \boxed{\infty}\boxed{1}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1} < \langle T\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{2} \quad \langle S\rangle_r = \underbrace{\boxed{1}\boxed{2}}_{\mathrm{root}(\langle S\rangle_r)}\boxed{1}\boxed{2}$

- ω 前順序：$X \preccurlyeq_\omega Y \Leftrightarrow \mathrm{root}(\langle X\rangle_r) = \mathrm{root}(\langle Y\rangle_r) \vee \exists i{:}\, \langle X^\omega[..i]\rangle < \langle Y^\omega[..i]\rangle$

文字列Tのルート(root)とは繰り返される接頭辞

回転文字列のソート

- $X =_\omega Y \Leftrightarrow \langle X^\omega[..2z]\rangle = \langle Y^\omega[..2z]\rangle$, ただし $z = \max\{|X|, |Y|\}$によって ω 前順序の計算が可能
- 最左の列: conjugate array CA
- 赤い行: conjugate range $\mathrm{CR}(P)$



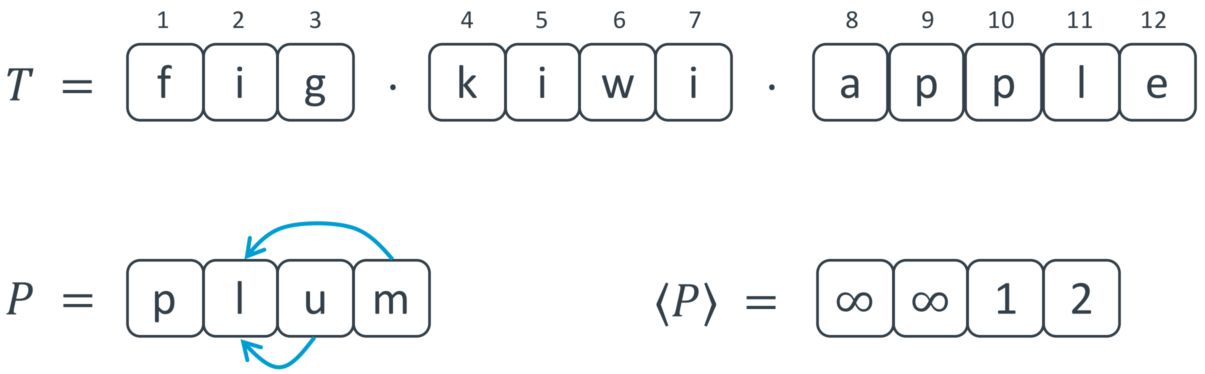
i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i)^\omega[..15]\rangle$
8	apple	∞11345113451134
5	iwik	∞12121212121212
7	ikiw	∞12121212121212
1	fig	∞12312312312312
9	pplea	∞1∞∞∞1134511345
12	eappl	∞∞1134511345113
4	kiwi	∞∞1212121212121
6	wiki	∞∞1212121212121
3	gfi	∞∞1231231231231
11	leapp	∞∞∞113451134511
2	igf	∞∞∞123123123123
10	pleap	∞∞∞∞11345113451

conjugate array : 回転文字列の配列（接尾辞配列の兄弟）

conjugate range ： パターンと一致する回天文字列の区間

回転文字列のソート

- $X =_\omega Y \Leftrightarrow \langle X^\omega[..2z]\rangle = \langle Y^\omega[..2z]\rangle$, ただし $z = \max\{|X|, |Y|\}$によって ω 前順序の計算が可能
- 最左の列: conjugate array CA
- 赤い行: conjugate range $\mathsf{CR}(P)$



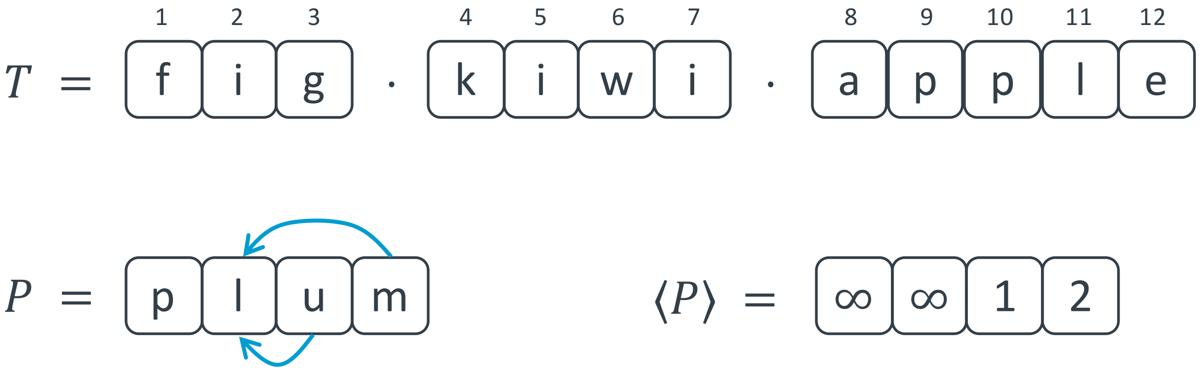
i	$\mathsf{C}(i)$	$\langle \mathsf{C}(i)^\omega[..15]\rangle$
8	apple	∞11345113451134
5	iwik	∞12121212121212
7	ikiw	∞12121212121212
1	fig	∞12312312312312
9	pplea	∞1∞∞∞1134511345
12	eappl	∞∞1134511345113
4	kiwi	∞∞1212121212121
6	wiki	∞∞1212121212121
3	gfi	∞∞1231231231231
11	leapp	∞∞∞113451134511
2	igf	∞∞∞123123123123
10	pleap	∞∞∞∞11345113451

conjugate array : 回転文字列の配列（接尾辞配列の兄弟）

conjugate range ： パターンと一致する回天文字列の区間

回転文字列のソート

- $X =_\omega Y \Leftrightarrow \langle X^\omega[..2z]\rangle = \langle Y^\omega[..2z]\rangle$, ただし $z = \max\{|X|, |Y|\}$によって ω 前順序の計算が可能
- 最左の列：conjugate array CA
- 赤い行：conjugate range $\mathrm{CR}(P)$



T = f i g · k i w i · a p p l e

P = p l u m

$\langle P\rangle$ = ∞ ∞ 1 2

位置で
タイブレーク

i	$\mathrm{C}(i)$	$\langle \mathrm{C}(i)^\omega[..15]\rangle$
8	apple	∞11345113451134
5	iwik	∞12121212121212
7	ikiw	∞12121212121212
1	fig	∞12312312312312
9	pplea	∞1∞∞∞1134511345
12	eappl	∞∞1134511345113
4	kiwi	∞∞1212121212121
6	wiki	∞∞1212121212121
3	gfi	∞∞1231231231231
11	leapp	∞∞∞113451134511
2	igf	∞∞∞123123123123
10	pleap	∞∞∞∞11345113451

conjugate array : 回転文字列の配列（接尾辞配列の兄弟）

conjugate range ：パターンと一致する回天文字列の区間

FL と LF 写像

- 回転文字列のランクは左右の回転文字列のランクに写像する（右：FL, 左:LF)
- タイブレークのため、テキスの符号化されたルートを循環させたい

FL と LF 写像

- 回転文字列のランクは左右の回転文字列のランクに写像する（右：FL, 左:LF)
- タイブレークのため、テキスの符号化されたルートを循環させたい

$$T_2 = T[4..7] = \boxed{\text{k}}\boxed{\text{i}}\boxed{\text{w}}\boxed{\text{i}}$$

$$\langle T_2 \rangle_r = \underbrace{\boxed{1}\boxed{2}}\underbrace{\boxed{1}\boxed{2}}$$

$$\text{root}(\langle T_2 \rangle_r)$$

FL と LF 写像

- 回転文字列のランクは左右の回転文字列のランクに写像する（右：FL, 左:LF)
- タイブレークのため、テキスの符号化されたルートを循環させたい

$$T_2 = T[4..7] = \boxed{\text{k}}\boxed{\text{i}}\boxed{\text{w}}\boxed{\text{i}}$$

$$\langle T_2 \rangle_r = \underbrace{\boxed{1}\boxed{2}}\underbrace{\boxed{1}\boxed{2}}$$

$\mathrm{root}(\langle T_2 \rangle_r)$

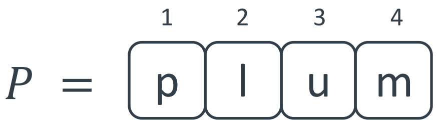
i	CA[i]	C(CA[i])	FL[i]	LF[i]
1	8	apple	5	6
2	5	iwik	7	7
3	7	ikiw	8	8
4	1	fig	11	9
5	9	pplea	12	1
6	12	eappl	1	10
7	4	kiwi	2	2
8	6	wiki	3	3
9	3	gfi	4	11
10	11	leapp	6	12
11	2	igf	9	4
12	10	pleap	10	5

FL と LF 写像

- 両方の写像を省メモリで表現したい
- [Kim and Cho '21]の整数表現を適応

FL と LF 写像

- 両方の写像を省メモリで表現したい
- [Kim and Cho '21]の整数表現を適用



$P =$

1	2	3	4
p	l	u	m

FL と LF 写像

- 両方の写像を省メモリで表現したい
- [Kim and Cho '21]の整数表現を適応

$\langle P[5..]\rangle =$

$\langle P[4..]\rangle = \boxed{\infty}$

$\langle P[3..]\rangle = \boxed{\infty}\boxed{\infty}$

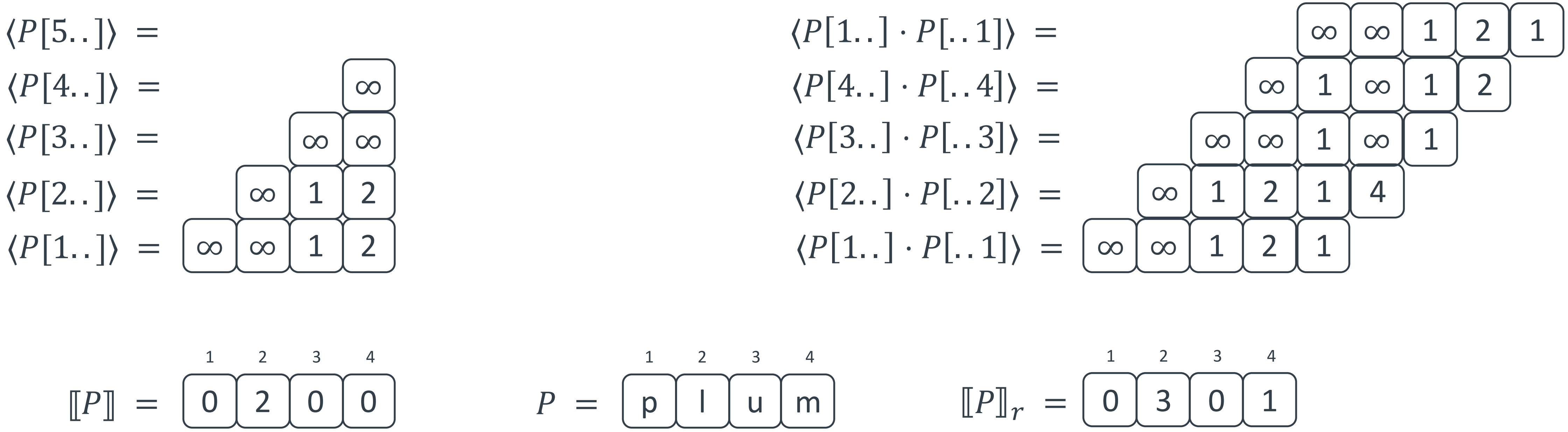
$\langle P[2..]\rangle = \boxed{\infty}\boxed{1}\boxed{2}$

$\langle P[1..]\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}$

$$[\![P]\!] = \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \boxed{0} & \boxed{2} & \boxed{0} & \boxed{0} \end{array} \qquad P = \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \boxed{\mathrm{p}} & \boxed{\mathrm{l}} & \boxed{\mathrm{u}} & \boxed{\mathrm{m}} \end{array}$$

FL と LF 写像

- 両方の写像を省メモリで表現したい
- [Kim and Cho '21]の整数表現を適応



$\langle P[5..]\rangle =$

$\langle P[4..]\rangle = \boxed{\infty}$

$\langle P[3..]\rangle = \boxed{\infty}\boxed{\infty}$

$\langle P[2..]\rangle = \boxed{\infty}\boxed{1}\boxed{2}$

$\langle P[1..]\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}$

$\langle P[1..]\cdot P[..1]\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}$

$\langle P[4..]\cdot P[..4]\rangle = \boxed{\infty}\boxed{1}\boxed{\infty}\boxed{1}\boxed{2}$

$\langle P[3..]\cdot P[..3]\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{\infty}\boxed{1}$

$\langle P[2..]\cdot P[..2]\rangle = \boxed{\infty}\boxed{1}\boxed{2}\boxed{1}\boxed{4}$

$\langle P[1..]\cdot P[..1]\rangle = \boxed{\infty}\boxed{\infty}\boxed{1}\boxed{2}\boxed{1}$

$\llbracket P \rrbracket =$

1	2	3	4
0	2	0	0

$P =$

1	2	3	4
p	l	u	m

$\llbracket P \rrbracket_r =$

1	2	3	4
0	3	0	1

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$\llbracket \mathrm{C}(\mathrm{CA}[i]) \rrbracket_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathbf{CA}[i]$	$\mathbf{C}(\mathbf{CA}[i])$	$\mathbf{FL}[i]$	$\mathbf{F}[i]$	$[\![\mathbf{C}(\mathbf{CA}[i])]\!]_r$	$\mathbf{L}[i]$	$\mathbf{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

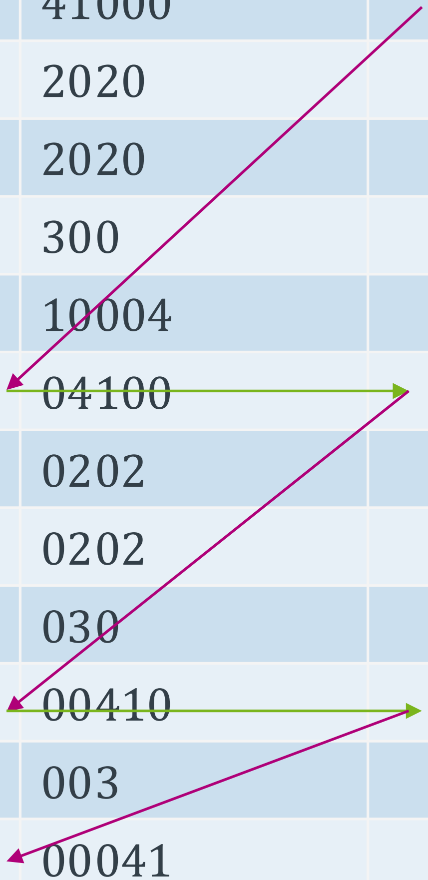
FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

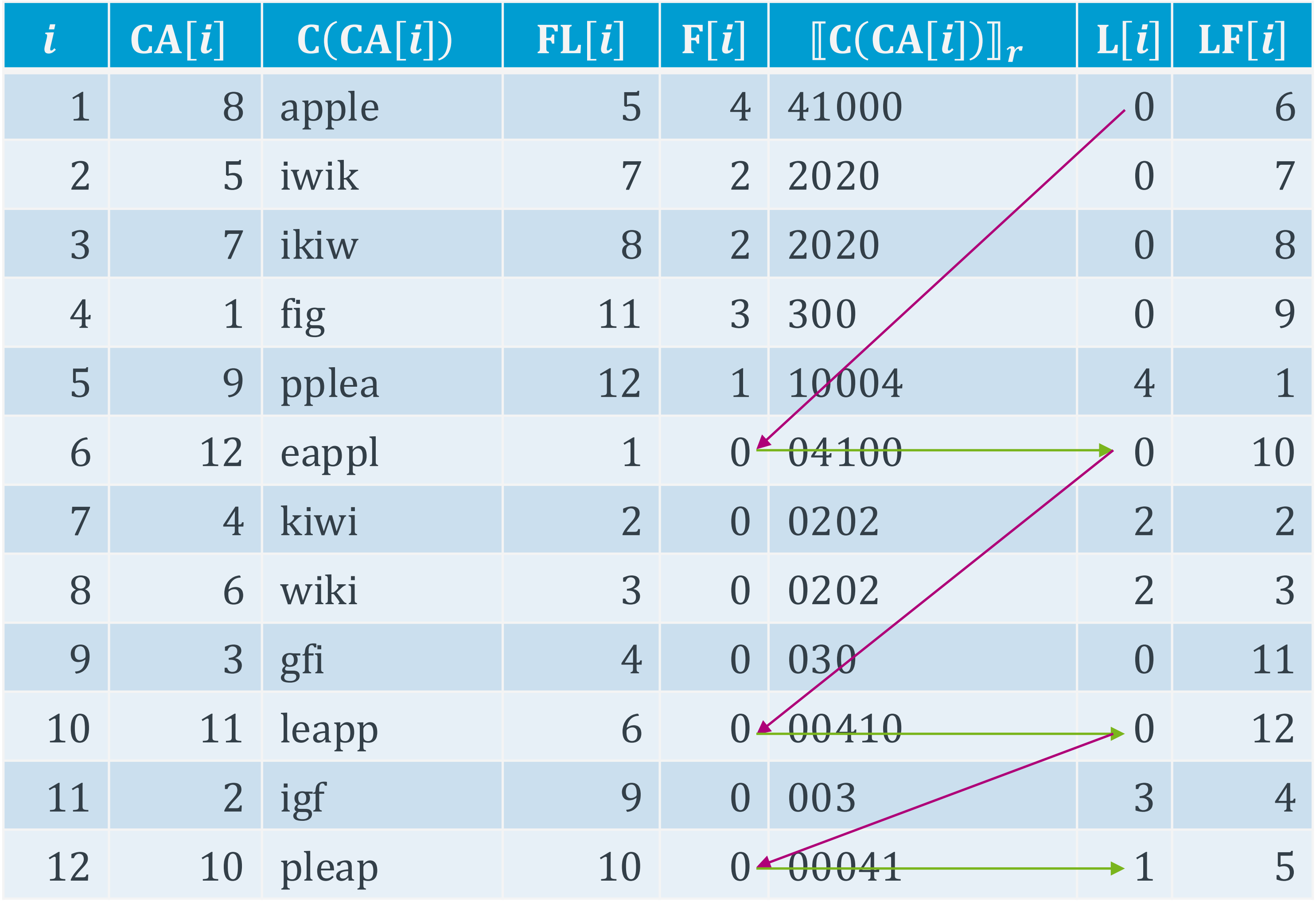
i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像



i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

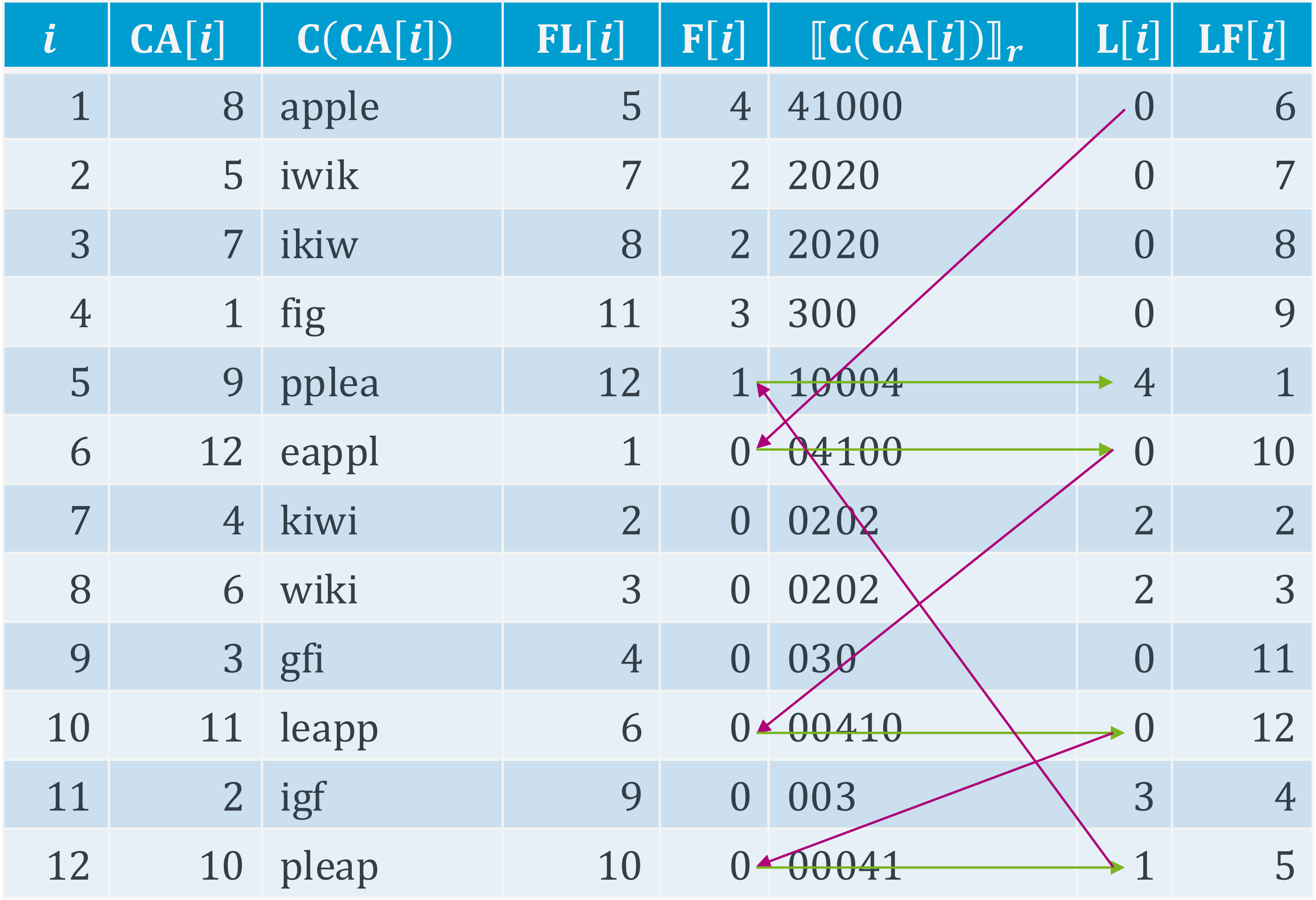


i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

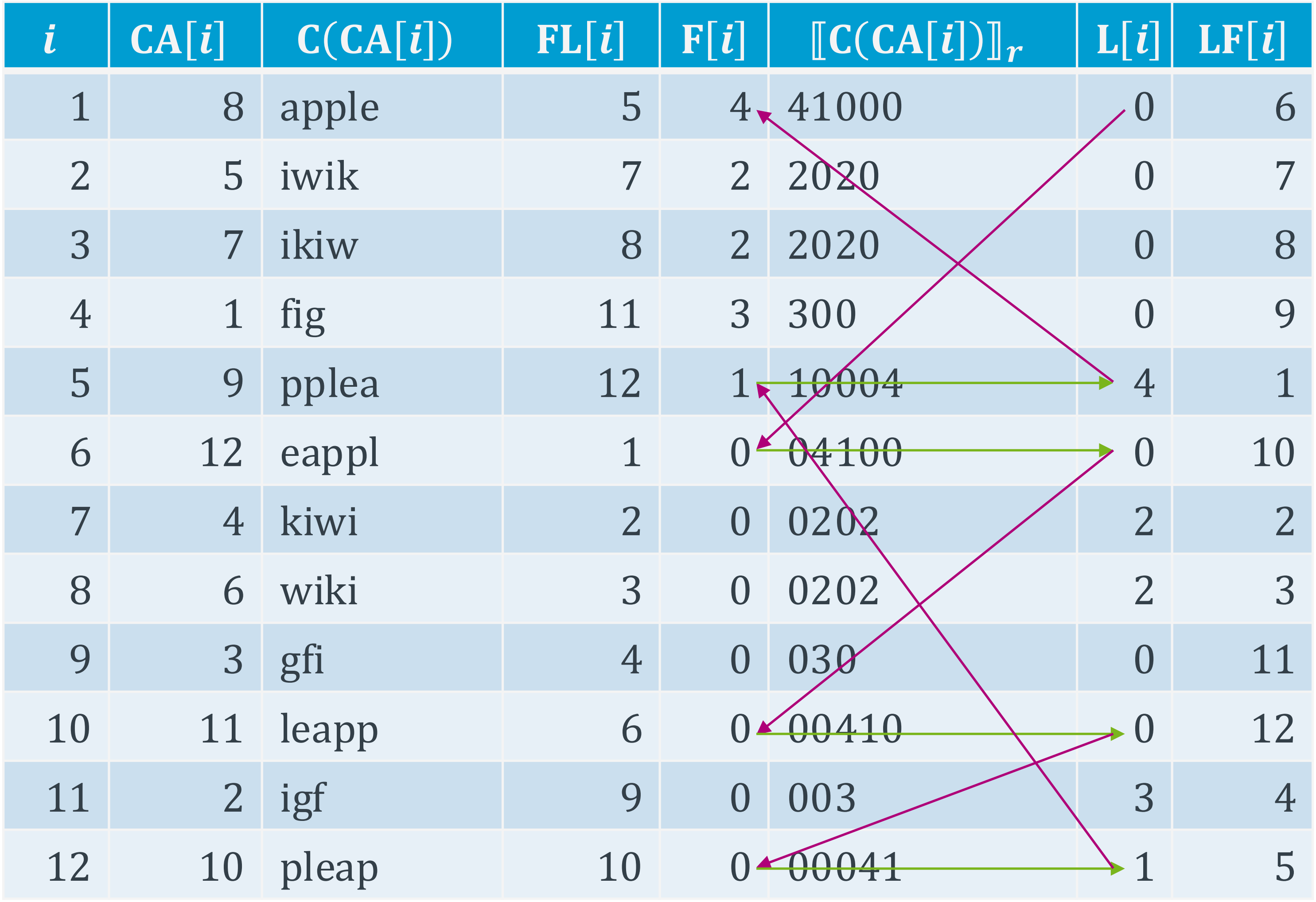
i	CA[i]	C(CA[i])	FL[i]	F[i]	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	L[i]	LF[i]
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像



i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像



i	CA[i]	C(CA[i])	FL[i]	F[i]	$\llbracket \mathrm{C}(\mathrm{CA}[i]) \rrbracket_r$	L[i]	LF[i]
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$\llbracket \mathrm{C}(\mathrm{CA}[i]) \rrbracket_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

i	$\mathrm{CA}[i]$	$\mathrm{C}(\mathrm{CA}[i])$	$\mathrm{FL}[i]$	$\mathrm{F}[i]$	$[\![\mathrm{C}(\mathrm{CA}[i])]\!]_r$	$\mathrm{L}[i]$	$\mathrm{LF}[i]$
1	8	apple	5	4	41000	0	6
2	5	iwik	7	2	2020	0	7
3	7	ikiw	8	2	2020	0	8
4	1	fig	11	3	300	0	9
5	9	pplea	12	1	10004	4	1
6	12	eappl	1	0	04100	0	10
7	4	kiwi	2	0	0202	2	2
8	6	wiki	3	0	0202	2	3
9	3	gfi	4	0	030	0	11
10	11	leapp	6	0	00410	0	12
11	2	igf	9	0	003	3	4
12	10	pleap	10	0	00041	1	5

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

i	CA[i]	F[i]	L[i]
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

P =

1	2	3	4
p	l	u	m

$[\![P]\!]$ =

1	2	3	4
0	2	0	0

i	CA[i]	F[i]	L[i]
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = \llbracket P \rrbracket[k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq \llbracket P \rrbracket[k]$

	1	2	3	4
P =	p	l	u	m

	1	2	3	4
$\llbracket P \rrbracket$ =	0	2	0	0

k	e_k	$\llbracket P \rrbracket[k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	
2	1	2	
1	2	0	

i	$\mathrm{CA}[i]$	$\mathrm{F}[i]$	$\mathrm{L}[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = \llbracket P \rrbracket[k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq \llbracket P \rrbracket[k]$

P =

1	2	3	4
p	l	u	m

$\llbracket P \rrbracket$ =

1	2	3	4
0	2	0	0

k	e_k	$\llbracket P \rrbracket[k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	
1	2	0	

i	CA[i]	F[i]	L[i]
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

	1	2	3	4
P =	p	l	u	m

	1	2	3	4
$[\![P]\!]$ =	0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	
1	2	0	

i	CA[i]	F[i]	L[i]
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

P =

1	2	3	4
p	l	u	m

$[\![P]\!]$ =

1	2	3	4
0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	$[6..12]$
1	2	0	

i	$\mathrm{CA}[i]$	$\mathrm{F}[i]$	$\mathrm{L}[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

P =

1	2	3	4
p	l	u	m

$[\![P]\!]$ =

1	2	3	4
0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	$[6..12]$
1	2	0	

i	CA$[i]$	F$[i]$	L$[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

	1	2	3	4
P =	p	l	u	m

	1	2	3	4
$[\![P]\!]$ =	0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	$[6..12]$
1	2	0	$[2..4]$

i	CA$[i]$	F$[i]$	L$[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

P =

1	2	3	4
p	l	u	m

$[\![P]\!]$ =

1	2	3	4
0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	$[6..12]$
1	2	0	$[2..4]$

i	$\mathrm{CA}[i]$	$\mathrm{F}[i]$	$\mathrm{L}[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

FL と LF 写像

- e_k: $\langle P[k..]\rangle$ 中に ∞ の出現の個数
- $j \in \mathrm{CR}(P[k+1..])$
- if $e_k > 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] = [\![P]\!][k]$
- if $e_k = 1$ then $LF[j] \in \mathrm{CR}(P[k..]) \Leftrightarrow L[j] \geq [\![P]\!][k]$

P =

1	2	3	4
p	l	u	m

$[\![P]\!]$ =

1	2	3	4
0	2	0	0

k	e_k	$[\![P]\!][k]$	$CR(P[k+1..])$
4	1	0	$[1..12]$
3	2	0	$[1..12]$
2	1	2	$[6..12]$
1	2	0	$[2..4]$

i	CA$[i]$	F$[i]$	L$[i]$
1	8	4	0
2	5	2	0
3	7	2	0
4	1	3	0
5	9	1	4
6	12	0	0
7	4	0	2
8	6	0	2
9	3	0	0
10	11	0	0
11	2	0	3
12	10	0	1

割愛した内容

- conjugate rangesを効率的に更新　（テキストの長さ n とは無関係）
- 構築した索引 (cBWT) は、動的な実装と静的な実装がある
  - 静的な実装は、[Kim and Cho ‘21]による索引の直接的な適応
  - 動的な実装は、[Hashimoto+ ‘22, Iseri+ ‘24]によって変更されたLCP配列を使用
- 構築アルゴリズムの方針は：
  - 単一テキスト索引の構築
  - 既存の索引を他のテキストで拡張
- [Hashimoto+ ‘22, Iseri+ ‘24]による技術を応用した構築

まとめと今後の取り組み

- cBWT 索引とは
  - CTPMを複数のテキストと(選択的に)循環テキストに答える
  - 動的な実装: $O(n \log \sigma)$ビット領域と $O(mt)$ 時間で CTPMを答える
  - 静的な実装: $3n + o(n)$ビット領域と$O(m)$時間で CTPMを答える
  - どちらの実装も、$O(n \log \sigma)$ビット領域と $O(nt)$ 時間で構築可能
- 今後の課題
  - 領域を縮める
  - 実験など

$$\boxed{t = \frac{\log \sigma \log n}{\log \log n}}$$