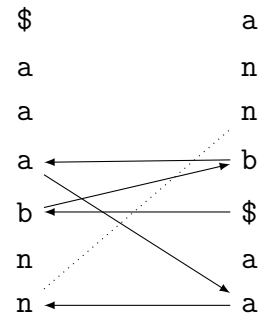


BWT (Burrows-Wheeler Transform)

BWT は文字列の変換方法である。作り方を banana の例文を使って、以下に説明する。

作り方	例																																																	
まずは、\$ を最後に追加する。 <ul style="list-style-type: none">\$ は他の文字より小さい。\$ は他の場所に現れない。	banana → banana\$ ただし \$ < a < b < n																																																	
全部の循環文字列を並べる。	<table><tr><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td></tr><tr><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td></tr><tr><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td></tr><tr><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td></tr><tr><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td></tr><tr><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td></tr><tr><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td></tr></table>	b	a	n	a	n	a	\$	a	n	a	n	a	\$	b	n	a	n	a	\$	b	a	a	n	a	\$	b	a	n	n	a	\$	b	a	n	a	a	\$	b	a	n	a	n	\$	b	a	n	a	n	a
b	a	n	a	n	a	\$																																												
a	n	a	n	a	\$	b																																												
n	a	n	a	\$	b	a																																												
a	n	a	\$	b	a	n																																												
n	a	\$	b	a	n	a																																												
a	\$	b	a	n	a	n																																												
\$	b	a	n	a	n	a																																												
循環文字列を辞書式順序でソートする。	<table><tr><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td></tr><tr><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td></tr><tr><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td></tr><tr><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td></tr><tr><td>b</td><td>a</td><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td></tr><tr><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td><td>n</td><td>a</td></tr><tr><td>n</td><td>a</td><td>n</td><td>a</td><td>\$</td><td>b</td><td>a</td></tr></table>	\$	b	a	n	a	n	a	a	\$	b	a	n	a	n	a	n	a	\$	b	a	n	a	n	a	n	a	\$	b	b	a	n	a	n	a	\$	n	a	\$	b	a	n	a	n	a	n	a	\$	b	a
\$	b	a	n	a	n	a																																												
a	\$	b	a	n	a	n																																												
a	n	a	\$	b	a	n																																												
a	n	a	n	a	\$	b																																												
b	a	n	a	n	a	\$																																												
n	a	\$	b	a	n	a																																												
n	a	n	a	\$	b	a																																												
最後の縦列を上から下まで読むと、BWT になっている。	banana\$ の BWT は annb\$aa																																																	

- 復元のため、BWT を縦書きにする (右)。
 - BWT の文字をソートすると、左の縦書き文字列を求める。(例：\$aaabnn)
- 復元したい時、右の \$ から始めて、以下の命令を適応する：
- 左の文字を見て、
 - 左の文字は x 番目の c なら、右の x 番目の c に移動する、
 - その作業を繰り返す。



雑学：Burrows さんと Wheeler さんは BWT という変換を圧縮のために 1994 年に発明した。現在、テキスト索引に対して、とても有名な方法となっている。

(クップル ドミニク, 令和 2 年 7 月)