

Burrows Wheeler Transformation

Um Daten komprimierungsfreundlicher zu gestalten möchten wir sie so transformieren, dass sie viele gleiche Buchstaben aufeinander folgen. Texte wie *KEINBEINREIN* beinhalten viele Wiederholungen von der Silbe *EIN* aber keine Buchstabendopplungen welche die kompression deutlich erleichtern. Eine Lösung dazu bietet die **Burrows–Wheeler–Transformation**.

Transformation

Zyklische Tabelle

Tragen Sie alle zyklischen Verschiebungen des Strings in eine Tabelle ein.

1	2	3	4	5	6	7	8	9	10	11	12
<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>
<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>
<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>
<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>
<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>
<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>
<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>
<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>
<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>

Sortieren

Sortieren Sie die Tabelle alphabetisch.

1	2	3	4	5	6	7	8	9	10	11	12
<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>
<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>
<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>
<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>
<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>
<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>
<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>
<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>
<i>N</i>	<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>
<i>R</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>K</i>	<i>E</i>	<i>I</i>	<i>N</i>	<i>B</i>	<i>E</i>	<i>I</i>	<i>N</i>

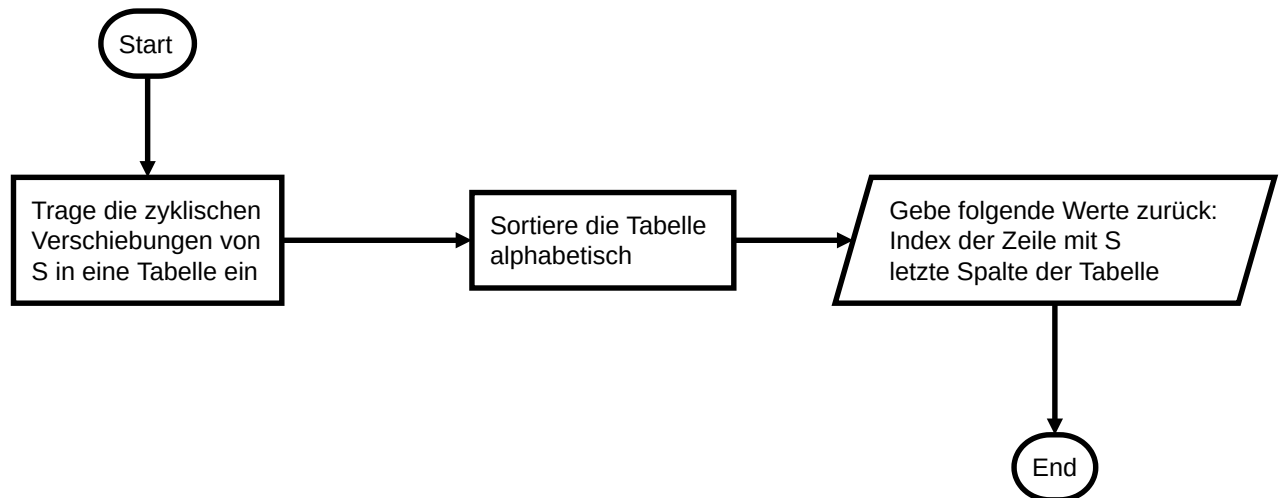
Rückgabe

Geben Sie den Index der Zeile in welcher der ursprüngliche Text steht und die letzte Spalte der Tabelle zurück. $s = \text{NKRBEENIIIN}$ und $i = 7$

Psydocode

```
def BWT(s):
    l_shift = lambda s, i: s[i:]+s[:i]
    table = sorted([l_shift(s, i) for i in range(len(s))])
    return ''.join(row[-1] for row in table), table.index(s)
```

```
BWT('KEINBEINREIN') -> ('NKRBEENIIIN', 7)
```



Inverse

Um die Transformation umzukehren wollen wir den String *NKRBEENIIIN* in die letzte Spalte der Tabelle eintragen. Danach sortieren wir die Tabelle und tragen den String *NKRBEENIIIN* in die vorletzte Spalte. Diesen Vorgang wiederholen wir bis die Tabelle gefüllt ist. Danach geben wir die Zeile der Tabelle mit dem Index i zurück.

1	2	3	4	5	6	7	8	9	10	11	12
										<i>N</i>	<i>B</i>
										<i>K</i>	<i>E</i>
										<i>R</i>	<i>E</i>
										<i>B</i>	<i>E</i>
										<i>E</i>	<i>I</i>
										<i>E</i>	<i>I</i>
										<i>E</i>	<i>I</i>
										<i>N</i>	<i>K</i>
										<i>I</i>	<i>N</i>
										<i>I</i>	<i>N</i>
										<i>I</i>	<i>N</i>
										<i>N</i>	<i>R</i>

1	2	3	4	5	6	7	8	9	10	11	12
										<i>B</i>	<i>E</i>
										<i>E</i>	<i>I</i>
										<i>E</i>	<i>I</i>
										<i>E</i>	<i>I</i>
										<i>I</i>	<i>N</i>
										<i>I</i>	<i>N</i>
										<i>I</i>	<i>N</i>
										<i>K</i>	<i>E</i>
										<i>N</i>	<i>B</i>
										<i>N</i>	<i>K</i>
										<i>N</i>	<i>R</i>
										<i>R</i>	<i>E</i>

Psydocode

```
def BWT_i(s, i):
    table = sorted(s)
    for _ in range(len(s)-1):
        table = [char + row for char, row in zip(s, table)]
        table = sorted(table)
    return table[i]
```

```
BWT_i('NKRBEENIIIN', 7) -> 'KEINBEINREIN'
```