گزارش پروژه پایان ترم

پیاده سازی الگوریتم BWT و backward-search

پویا ترابی

زمستان ۹۸

backward_search(pattern) تابع

الگوریتم جستجوی معکوس را با داشتن ماتریس های C و OCC اجرا می کند تا اولین و آخرین رخداد pattern را در آرایه پسوندی پیدا کند.

bwt(s) تابع

برای رشته ورودی مقادیر زیر را طبق الگوریتم محاسبه می کند:

- تبدیل BWT رشته ورودی
 - ماتریس C
 - ماتریس OCC
- آرایه پسوندی رشته ورودی

نتایج اجرا روی چند داده:

ورودی ۱:

Seq = GATGCGAGAGATG

Pattern to search for = GAT

خروجي هاي ١:

BWT: ggggggtcaa\$taa

C matrix:

{'\$': 0, 'a': 1, 'c': 5, 'g': 6, 't': 12}

OCC matrix:

{'\$': [0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1], 'a': [0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 2, 3, 4], 'c': [0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1], 'g': [1, 2, 3, 4, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6], 't': [0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2]}

number of pattern matches:

2

pattern matches:

gatg\$

gatgcgagagatg\$

ورودی۲:

Seq = GATGACACGTGTGCGGTCAATGGATTCGTCAGAAAAATCGTCAAAGATG

Pattern to search for = GAT

خروجي هاي ٢:

BWT: ggacaaacgccaaggagttatgttatata\$gttccgcttcgggataagagga

C matrix:

{'\$': 0, 'a': 1, 'c': 17, 'g': 25, 't': 40}

OCC matrix:

pattern matches:
gatg\$
gatgacacgtgtgtgcggtcaatggattcgtcagaaaaatcgtcaaagatg\$
gattcgtcagaaaaatcgtcaaagatg\$