# گزارش پروژه پایان ترم

# پیاده سازی الگوریتم BWT و backward-search

# پویا ترابی

# زمستان 98

# تابع backward\_search(pattern)

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.

# تابع bwt(s)

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:

ورودی1:

Seq = GATGCGAGAGATG

Pattern to search for = GAT

خروجی های1:

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, 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, 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, 0, 1, 1, 1, 1, 1, 2, 2, 2]}

number of pattern matches:

2

pattern matches:

gatg$

gatgcgagagatg$

ورودی2:

Seq = GATGACACGTGTGTGCGGTCAATGGATTCGTCAGAAAAATCGTCAAAGATG

Pattern to search for = GAT

خروجی های2:

BWT: ggacaaacgccaaggagttatgttatata$gttccgcttcgggataagagga

C matrix:

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

OCC matrix:

{'$': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1], 'a': [0, 0, 1, 1, 2, 3, 4, 4, 4, 4, 4, 5, 6, 6, 6, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 10, 10, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 11, 12, 12, 13, 14, 14, 15, 15, 15, 16], 'c': [0, 0, 0, 1, 1, 1, 1, 2, 2, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 6, 6, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8], 'g': [1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 5, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9, 9, 10, 11, 12, 12, 12, 12, 12, 13, 13, 14, 15, 15], 't': [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 2, 2, 3, 3, 4, 5, 5, 6,

6, 7, 7, 7, 7, 8, 9, 9, 9, 9, 9, 10, 11, 11, 11, 11, 11, 11, 12, 12, 12, 12, 12, 12, 12, 12]}

number of pattern matches:

3

pattern matches:

gatg$

gatgacacgtgtgtgcggtcaatggattcgtcagaaaaatcgtcaaagatg$

gattcgtcagaaaaatcgtcaaagatg$