

به نام خدا

امیررضا رجبی ۹۸۳۱۱۲۶

تمرین سوم

سوال اول :

Dynamic programming روش exhaustive اما FASTA و BLAST روشهایی heuristic

هستند که trade off ی بین سرعت و sensitivity دارند .

در روش داینامیک پروگرامینگ ما حافظه و زمان و هزینه زیادی صرف میکنیم که تمام جواب ها را چک کنیم و لزوما بهترین جواب را برگردانیم .

زمان های اجرا به ترتیب DP>FASTA>BLAST است .

حساسیت ها به ترتیب DP>FASTA>BLAST است.

در نهایت هر دو روش FASTA و BLAST از DP استفاده می شود ولی در روش BLAST نیاز به درخت جستجو است که حافظه را زیاد میکند .

برای پرسمان های با طول کوتاه FASTA روشی مناسب است اما فقط یک رشته بر میگرداند بر خلاف دو روش دیگر که داریم .

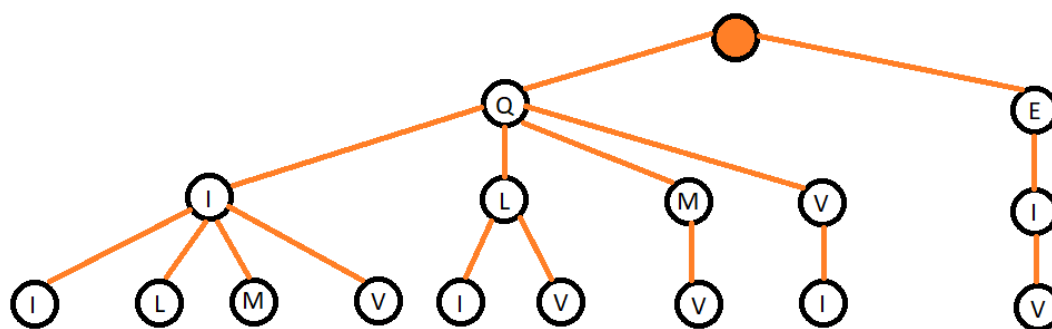
سوال دوم :

(الف)

Q I I	Q I L	Q I M	Q I V	Q L I	Q L V	Q M V	Q V I	Q V V	E I V
+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V	+ Q I V
5+4+3	5+4+1	5+4+1	5+4+4	5+2+3	5+2+4	5+1+4	5+3+3	5+3+4	2+4+4

SEQ	QII	QIL	QIM	QIV	QLI	QLV	QMV	QVI	QVV	EIV
Score	12	10	10	13	10	11	10	11	12	10

(ب)



سوال ۳ : انف)

Query: CACGTTGACAT

hash table →

index : 1 2 3 4 5 6 7 8 9 10 11

C A G T

1 2 4 5

3 8 7 6

9 10 11

a: ATGACATTGAA

target table: 1 2 3 4 5 6 7 8 9 10 11 12

A T G A C A T T C G A A

1 3 1 -2 -4 -4 -2 -3 -8 -6 -9 -10

~~TGACAT~~

7 4 4 4 -2 2 -1 -2 -6 -3 -3 -4

9 9 6 4 4 4 5 0 -1 -2

1 → TGACAT

سوال ۳: الف

b: C G A T T C G G A C A

target table b:

1	2	3	4	5	6	7	8	9	10	11
C	G	A	T	T	C	G	G	A	C	A
0	2	-1	1	0	-5	-3	-4	-7	-9	-9
2	5	5	2	1	-3	0	-1	-1	-7	-3
8	7	7	6	3			1	-1	-1	

→ ~~ACGACGACGACG~~ GACA

TGACAT : پس جواب می شود

سوال ۳: ب. target: T A C G T C G T

query: G T T A C C A C G

G T T T A A C C C A A C C G

index 1 2 3 4 5 6 7 8

G T T T A A C C C A C G

hash table 1 2 3 4 5 6 8

7

target table : T A C 1 2 3 4 5 6 7

T A A C C G G T T C C G G T

T A C جواب

است A C G

2 2 5 -3 2 -6

سوال چہارم :

سوال ۴ :

	A	B	C	D	
A	—	5	4	11	20
B	5	—	11.5	12	28.5
C	4	11.5	—	10	25.5
D	11	12	10	—	33
	20	28.5	25.5	33	

D بہ عنوان مرکز انتخاب کی گئی و دیگر ترتیب یا B و A و C متنازعہ

seq D: TCACTGGAC---

seq B: ---TTGCACATC

seq D: ----TCACTGGAC---

seq B: -----TTGCACATC

seq A: ACGCTAAC-----

seq D: -----TCACTGGAC---

seq B: -----TTGCACATC

seq A: ----ACGCTAAC-----

seq C: TCGGTAGATC-----

D TCACTGGAC---

B ---TTGCACATC

D ----TCACTGGAC

A ACGCTAAC----

D -----TCACTGGAC

C TCGGTAGATC-----

سوال پنجم :

clustalW همترازی را به صورت **global** انجام میدهد پس برای دنباله‌های با طول متفاوت مناسب نیست و همینطور نتیجه نهایی به ترتیب اضافه کردن **seq** ها وابسته است پس خطایی اگر داشته باشد از همترازی‌های پیشین منتشر میشود .

Tcofee علاوه بر همترازی **global** همترازی **local** برای تمام جفت ها انجام میدهد که **global** را با **clustal** و **local** را با **Lalign** انجام میدهد به دلیل **local** می تواند برای طول متفاوت مناسب باشد دقت بیشتری نسبت به **clustalW** دارد اما کندتر است .

Star روش هیورستیک است که سرعت بالا دارد ولی روش ما **triangle inequality** را ارضا کند و وابسته به ترتیب نیست و قانونی که دارد گپ همیشه گپ میماند

در **clsutal** ما با ساخت درخت **neighbor joining** با استفاده از همترازی گلوبال از درخت کمک میگیریم تا الاین کنیم