تمرین ۷ درس طراحی الگوریتم

سید صالح اعتمادی
امیر خاکپور
مهسا سادات رضوی
دانشگاه علم و صنعت ۹۸ـ۹۷

لطفا به نکات زیر توجه کنید:

- مهلت ارسال این تمرین شنبه ۳۱ فروردین ساعت ۱۱:۵۹ ب.ظ است.
- این تمرین شامل سوال های برنامه نویسی است، بنابراین توجه کنید که حتماً موارد خواسته شده را رعایت کنید. .
- نام شاخه، پوشه و پول ریکوست همگی دقیقا "$A7$" باشد.
- در صورتی که به اطلاعات بیشتری نیاز دارید می توانید با آیدی تلگرام sargdsra@ یا mhsarzvi@ در ارتباط باشید.
- اگر در حل تمرین شماره ی ۷ مشکلی داشتید،لطفا به sargdsra@ یا mhsarzvi@ مراجعه کنید.

موفق باشید.

توضیحات کلی تمرین

تمرین این هفته ی شما،۳ سوال دارد که باید به همه ی این سوال ها پاسخ دهید. برای حل این سری از تمرین ها مراحل زیر را انجام دهید:

۱. ابتدا مانند تمرین های قبل، یک پروژه به نام $A7$ بسازید.

۲. کلاس هر سوال را به پروژه ی خود اضافه کنید و در قسمت مربوطه کد خود را بنویسید. هر کلاس شامل دو متد اصلی است:

- متد اول: تابع solve است که شما باید الگوریتم خود را برای حل سوال در این متد پیاده سازی کنید.
- متد دوم: تابع process است که مانند تمرین های قبلی در TestCommon پیاده سازی شده است.

بنابراین با خیال راحت سوال را حل کنید و نگران تابع process نباشید! زیرا تمامی پیاده سازی ها برای شما انجام شده است و نیازی نیست که شما کدی برای آن بزنید.

۳. اگر برای حل سوالی نیاز به تابع های کمکی دارید؛ می توانید در کلاس مربوط به همان سوال تابع تان را اضافه کنید.

اکنون که پیاده سازی شما به پایان رسیده است، نوبت به تست برنامه می رسد. مراحل زیر را انجام دهید.

۱. یک UnitTest برای پروژه ی خود بسازید.

۲. فولدر TestData که در ضمیمه همین فایل قرار دارد را به پروژه ی تست خود اضافه کنید.

۳. فایل GradedTests.cs را به پروژه ی تستی که ساخته اید اضافه کنید. توجه کنید که بر خلاف تمرین‌های قبل برای هر سوال یک تست در نظر گرفته شده تا Timeout هر تست جداگانه محاسبه شود و در صورت حل نکردن هر تست بتوانید برای آن تست جداگانه Assert.Inconclusive بنویسید.

۱ پیدا کردن تمام تکرارهای یک الگو در یک رشته

در این سوال به دنبال پیدا کردن تعداد تکرارهای یک الگو بعنوان زیر رشته در رشته ی ورودی هستیم. توجه کنید که این زیررشته ها می توانند با یکدیگر هم پوشانی داشته باشند. برای مثال الگوی ATA سه بار در رشته ی CGATATATCCATAG تکرار شده است.

در این سوال باید الگوریتمی بنویسید که در یک رشته ورودی تعداد تکرارهای الگوهای داده شده را پیدا کند. در خط اول فایل ورودی رشته متنی که باید الگو را در آن جستجو کنیم وجود دارد و در خط دوم الگویی که به دنبال آن هستیم و باید آن را در رشته متنی جستجو کنیم وجود دارد.

در خروجی باید در هر خط ایندکس هایی از متن (با فرض شروع ایندکس گذاری از صفر) که زیر رشته از آنجا شروع شده را برگردانید. در صورتی که الگو در رشته متنی وجود نداشت ۱ـ را بعنوان خروجی برگردانید.

نمونه ۱

ورودی:

```
GT
TACG
```

خروجی:

```
-1
```

در این مثال طول الگو از رشته بیشتر است پس الگو در رشته وجود ندارد.

نمونه ۲

ورودی:

```
ATATA
ATA
```

خروجی:

```
0
2
```

نمونه ۳

ورودی:

```
GATATATGCATATACTT
ATAT
```

خروجی:

```
1
3
9
```

۲ تشکیل Suffix Array برای یک رشته ی طولانی

هدف این سوال بدست آوردن Suffix Array است اما این بار رشته مورد نظر طول بزرگی دارد. بنابراین الگوریتمی با پیچیدگی زمانی درجه دو نمی تواند برای این سوال مناسب باشد و نیازمندید الگوریتمی با پیچیدگی زمانی تقریبا خطی برای این سوال پیاده سازی کنید.
در فایل ورودی یک رشته وجود دارد که با حروف T G C A ساخته شده است و با نماد $ پایان می یابد.
خروجی Suffix Array برای رشته است که شامل لیستی از اعداد صحیح نشان دهنده ی ایندکس شروع پسوندهای مرتب شده ی رشته هستند.

نمونه ۱

ورودی:

```
AAA$
```

خروجی:

```
3 2 1 0
```

Sorted suffixes:

```
3   $
2   A$
1   AA$
0   AAA$
```

نمونه ۲

ورودی:

```
GAC$
```

خروجی:

```
3 1 2 0
```

Sorted suffixes:

```
3   $
1   AC$
2   C$
0   GAC$
```

نمونه ۳

ورودی:

```
GAGAGAGA$
```

خروجی:

```
8 7 5 3 1 6 4 2 0
```

Sorted suffixes:

```
8   $
7   A$
5   AGA$
3   AGAGA$
1   AGAGAGA$
6   GA$
4   GAGA$
2   GAGAGA$
0   GAGAGAGA$
```

نمونه ۴

ورودی:

```
AACGATAGCGGTAGA$
```

خروجی:

```
15 14 0 1 12 6 4 2 8 13 3 7 9 10 11 5
```

Sorted suffixes:

```
15   $
14   A$
0    AACGATAGCGGTAGA$
1    ACGATAGCGGTAGA$
12   AGA$
6    AGCGGTAGA$
4    ATAGCGGTAGA$
2    CGATAGCGGTAGA$
8    CGGTAGA$
13   GA$
3    GATAGCGGTAGA$
7    GCGGTAGA$
9    GGTAGA$
10   GTAGA$
11   TAGA$
5    TAGCGGTAGA$
```

۳ تطبیق الگوها با استفاده از Suffix Array

در این سوال باید الگوریتم تطبیق چندگانه الگوها را با استفاده از Suffix Array پیاده سازی کنید.

در این سوال باید تمامی تکرارهای الگوهای داده شده در رشته ورودی را بدست آورید. خط اول از فایل ورودی رشته مورد نظر است که با حروف G C T A ساخته شده است و طول آن بین ۱ تا ۱۰۰۰۰۰ کاراکتر است. در خط بعدی یک عدد صحیح وجود دارد که نشان دهنده تعداد الگوها برای یافتن در رشته است. در هر یک از n خط بعدی الگوهای مورد نظر برای پیدا کردن در رشته آمده است.

در فایل خروجی در هر خط تمامی ایندکس هایی از متن که هر یک از الگوها در آنجا به عنوان زیررشته وجود دارند را برگردانید. اگر در یک ایندکس بیش از یک الگو یافت شد تنها یک بار آن را در خروجی چاپ کنید و اگر جوابی نیافتید ۱ـ را یعنوان خروجی برگردانید.

نمونه ۱

ورودی:

```
AAA
1
A
```

خروجی:

```
0
1
2
```

نمونه ۲

ورودی:

```
ATA
2
C
G
```

خروجی:

```
-1
```

نمونه ۳

ورودی:

```
ATATATA
3
ATA
C
TATAT
```

خروجی:

```
0
2
1
4
```