

تمرین ۵ درس ساختمان داده

مهسا سادات رضوی

امیر خاکپور

سید صالح اعتمادی

دانشگاه علم و صنعت ۹۸-۹۷

لطفاً به نکات زیر توجه کنید:

- مهلت ارسال این تمرین شنبه ۱۷ 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 دقیقاً "A۵" 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 ۵ مشکلی داشتید، لطفاً به @sargdsra یا @mhsarzvi مراجعه کنید.

موفق باشید.

توضیحات کلی تمرین

تمرین این هفته ی شما، ۵ 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 A5 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 بنویسید.

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 مانند تمرین های قبل، لازم نیست که برای هر سوال TestMethod بنویسید. تمامی آنچه که برای تست هر سوالتان نیاز دارید از قبل در این فایل برای شما پیاده سازی شده است.

دقت کنید که **TestCommon** تغییر یافته است. بنابراین شما باید نسخه ی جدید آن را با دستور **git Pull** دریافت کنید . با استفاده از **TestCommon** بروز شده میتوانید در **Constructor** هر سوال از اجرای تعدادی تست کیس دلخواه جلوگیری کنید.

1. `ExcludeTestCaseRangeInclusive`
2. `ExcludeTestCases`

چنانچه کد خود را با حذف تعدادی تست کیس یا تغییر **Timeout** بفرستید مقدار متناسب و قابل توجهی نمره کاسته میشود.

۱ ساختن درخت (Trie) از روی تعدادی الگو

یک درخت برای مجموعه‌ای از الگوها بصورت زیر تعریف می‌شود:

- یک درخت تنها یک ریشه دارد.
- هر یال با یک حرف از حروف الفبا برچسب گذاری می‌شود.
- یال‌هایی که از یک گره خارج می‌شوند دارای برچسب‌های متمایزاند.
- الگوها از طریق پیمایش یک مسیر از ریشه درخت به سمت برگ‌ها بدست می‌آیند.
- هر مسیر از ریشه درخت تا برگ‌ها نشان دهنده‌ی یک الگو است.

در این سوال شما باید الگوریتمی بنویسید که یک درخت از رشته الگوهای داده شده در ورودی را تشکیل دهد.

خط اول فایل ورودی یک عدد صحیح بین ۱ تا ۱۰۰ است که این عدد نشان دهنده تعداد الگوهای داده شده برای ساخت درخت است. در هر یک از خطوط بعدی الگوها برای ساخت درخت آورده شده‌اند. طول این رشته الگوها بین ۱ تا ۱۰۰ کاراکتر از بین مجموعه کاراکترهای C G T A است. در این سوال فرض بر آن است که هیچ یک از الگوها زیررشته‌ای از الگوهای دیگر نیستند.

در فایل خروجی ساختار درخت در قالب یک لیست با فرمت مشخص نشان داده شده است. در درختی که با n گره می‌سازید، گره ریشه را با برچسب صفر شماره گذاری کنید و باقی گره‌ها را متناسب با ترتیب رشته‌ها در ورودی با اعداد ۱ تا $n-1$ شماره گذاری کنید. هر خط از فایل خروجی نشان دهنده یالی از درخت ساخته شده است. هر یال را با یک سه تایی به فرم $u \rightarrow v : c$ نشان می‌دهیم. دو عدد اول یعنی u و v نشان دهنده شماره گره مبدا و مقصد یال است و کاراکتر سوم یعنی c نشان دهنده برچسب یال مورد نظر است

دقت کنید که ترتیب نشان دادن یال‌ها در خروجی، متناسب با ترتیب رشته‌های الگو در ورودی است.

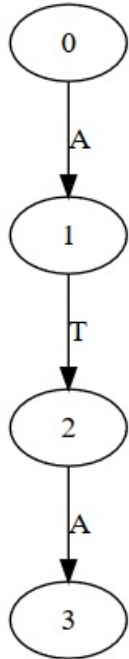
نمونه ۱

ورودی:

1 ATA

خروجی:

0->1:A
1->2:T
2->3:A

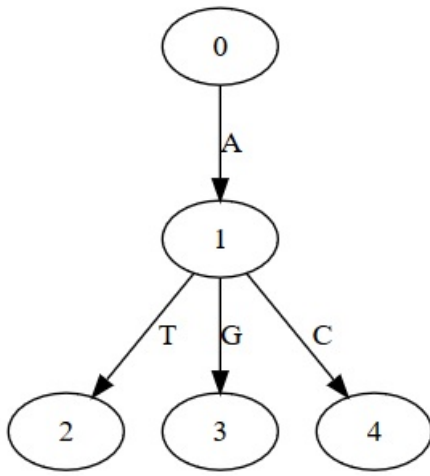


نمونه ۲
ورودی:

3
AT
AG
AC

خروجی:

0->1:A
1->2:T
1->3:G
1->4:C

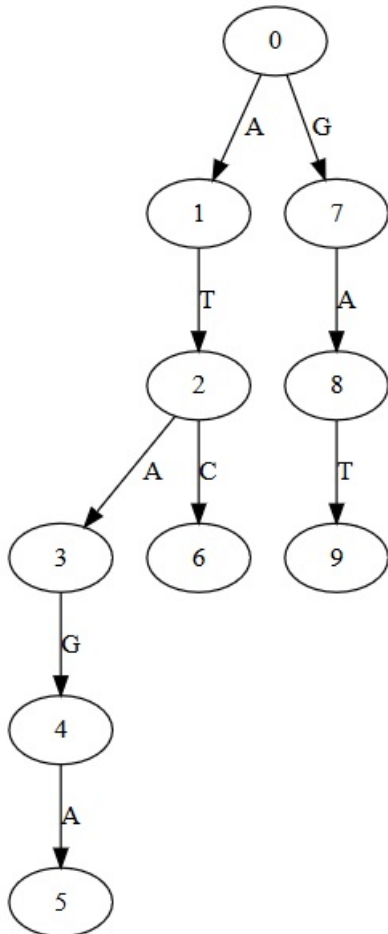


نمونه ۳
ورودی:

3
ATAGA
ATC
GAT

خروجی:

0->1:A
1->2:T
2->3:A
3->4:G
4->5:A
2->6:C
0->7:G
7->8:A
8->9:T



۲ تطبیق الگوهای چندگانه

در این سوال باید الگوریتمی بنویسید که محل تعدادی الگو به عنوان زیر رشته را در یک متن برگرداند.

در خط اول از فایل ورودی یک رشته از متن مورد نظر با طولی بین ۱ تا ۱۰۰۰۰ کاراکتر متشکل از C G T A داده شده است. در خط دوم یک عدد صحیح بین ۱ تا ۵۰۰۰ که نشان دهنده تعداد الگوها است و در هر یک از خطوط بعدی رشته الگوهای داده شده برای یافتن در متن داده شده است. در این سوال فرض بر آن است که هیچ یک از الگوها زیررشته‌ای از الگوهای دیگر نیست.

در فایل خروجی تمامی index هایی از متن (بصورت صعودی) که در آن یک رشته از الگوها، بعنوان زیر رشته یافت می شود وجود دارد (index گذاری روی متن از صفر شروع می شود). در صورتی که پاسخی نیافت مقدار ۱ - را برگرداند.

نمونه ۱

ورودی:

AAA

1

AA

خروجی:

0 1

نمونه ۲

ورودی:

AA

1

T

خروجی:

-1

نمونه ۳

ورودی:

AATCGGGTTCAATCGGGT
2
ATCG
GGGT

خروجی:

1 4 11 15

۳ تطبیق الگوهای چندگانه‌ی تعمیم یافته

در این سوال راه حل سوال قبل تعمیم داده می شود بطوریکه شما باید الگوریتمی بنویسید که حالتی را که یک الگوی ورودی زیررشته الگوی دیگری باشد را نیز شامل شود. در این حالت رشته‌هایی از الگوها وجود دارند که در طی مسیری از ریشه درخت تا یک گره داخلی (و نه لزوما برگ) ساخته می‌شوند.

در خط اول فایل ورودی یک رشته از متن مورد نظر با طولی بین ۱ تا ۱۰۰۰۰ کاراکتر متشکل از C G T A داده شده است. در خط دوم یک عدد صحیح بین ۱ تا ۵۰۰۰ که نشان دهنده تعداد الگوها است و در هر یک از خطوط بعدی رشته الگوهای داده شده برای یافتن در متن داده شده است.

در این سوال فرض بر آن است که الگوها می توانند زیررشته‌ای از الگوهای دیگر باشد. در فایل خروجی تمامی index هایی از متن (بصورت صعودی) که در آن یک رشته از الگوها بعنوان زیر رشته یافت می شود وجود دارد (index گذاری روی متن از صفر شروع می‌شود). در صورتی که پاسخی نیافت مقدار ۱- را برگرداند.

اگر بیش از یک الگو در i امین index یافت شود، تنها یکبار i را در خروجی نشان دهید.

نمونه ۱

ورودی:

AAA

1

AA

خروجی:

0 1

نمونه ۲

ورودی:

ACATA

3

AT

A

AG

خروجی:

0 2 4

۴ ساختن suffix tree برای یک رشته

درخت‌های پسوندی یک ساختار داده قدرتمند است که در زمینه‌هایی مثل pattern matching و compression کاربرد دارد. هدف از این سوال پیاده‌سازی این داده ساختار است. در این سوال باید الگوریتمی بنویسید که درخت پسوندی از یک رشته ورودی تشکیل دهد. در خط اول فایل ورودی یک رشته که با نماد \$ پایان می‌یابد وجود دارد. این رشته با طولی بین ۱ تا ۵۰۰۰ از مجموعه کاراکترهای C G T A ساخته شده است. در هر یک از خط‌های فایل خروجی یک رشته که نشان دهنده‌ی برجسب یالی از درخت است وجود دارد. ترتیب نشان دادن یال‌ها در خروجی باید مطابق با پیمایش درخت بصورت اول عمق (dfs) باشد.

نمونه ۱

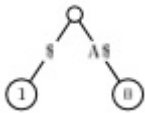
ورودی:

A\$

خروجی:

A\$

\$



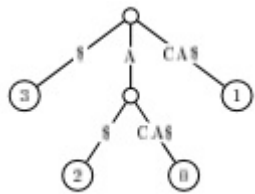
نمونه ۲

ورودی:

ACA\$

خروجی:

A
CA\$
\$
CA\$
\$



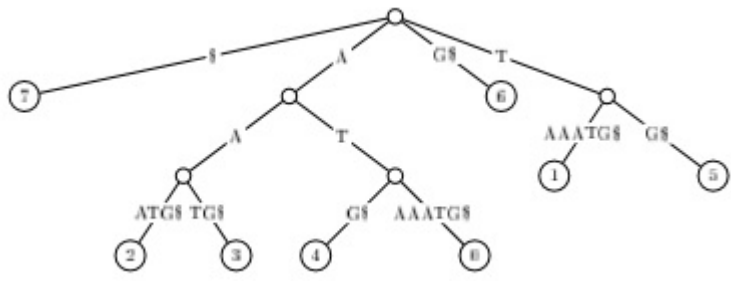
نمونه ۳

ورودی:

ATAAATG\$

خروجی:

A
 T
 AAATG\$
 G\$
 A
 ATG\$
 TG\$
 T
 AAATG\$
 G\$
 G\$
 \$



۵ پیدا کردن کوچکترین زیر رشته غیر مشترک دو رشته

در این سوال دو رشته به شما داده می‌شود و هدف پیدا کردن کوچکترین زیر رشته غیر مشترک بین آن دو رشته است.

شما باید الگوریتمی بنویسید که کوچکترین زیررشته از رشته اول، که در رشته دوم وجود ندارد را پیدا کند.

در دو خط اول فایل ورودی ۲ رشته ورودی وجود دارد. رشته‌های ورودی با طولی بین ۱ تا ۲۰۰۰ متشکل از C G T A ساخته شده‌اند. طول این رشته‌ها با هم برابراند.

در فایل خروجی کوچکترین زیر رشته از رشته اول که در رشته دوم وجود ندارد قرار دارد. (ممکن است چند جواب متفاوت برای دو رشته وجود داشته باشد که در اینصورت باید زیر رشته‌ای که در index کوچکتری وجود دارد را برگردانید.)

نمونه ۱

ورودی:

A
T

خروجی:

A

نمونه ۲

ورودی:

AAAAAAAAAAAAAAAAAAAAA
TTTTTTTTTTTTTTTTTTTTT

خروجی:

A

نمونه ۳

ورودی:

CCAAGCTGCTAGAGG
CATGCTGGGCTGGCT

خروجی:

CC

نمونه ۴

ورودی:

ATGCGATGACCTGACTGA
CTCAACGTATTGGCCAGA

خروجی:

ATG