



تمرین کامپیوتری سوم

سیستم‌های عامل - بهار ۱۳۹۹

طراحی چندریسه‌ای

دانشکده مهندسی برق و کامپیوتر

مسئول تمرین:

مهلت تحویل:

استاد:

محمد مریدی

ساعت ۲۳:۵۵ روز ۶ اردی‌بهشت ماه ۱۳۹۹

مهدی کارگهی

مقدمه

هدف از این تمرین آشنایی شما با مفاهیم اولیه طراحی چندریسه‌ای^۱ یک مسئله است. در این تمرین شما به تحلیل داده‌هایی که از وبگاه^۲ Goodreads استخراج شده است می‌پردازید. پیشنهاد می‌شود در ابتدا به بررسی مجموعه‌های داده^۳‌ای که در کنار این پرونده^۴ بارگذاری شده است پرداخته و سپس شرح تمرین را مطالعه فرمایید.

goodreads

شرح تمرین

در این تمرین شما به تحلیل داده‌هایی که از وبگاه Goodreads استخراج شده است می‌پردازید. در ابتدا برنامه شما اقدام به خواندن و تجزیه^۵ مجموعه داده‌های داده شده می‌کند و آن‌ها را در حافظه^۶ خود ذخیره می‌کند. پس از استخراج داده‌ها، برنامه اقدام به بدست آوردن محبوب‌ترین کتاب در ژانری مشخص، بر اساس معیاری که توضیح داده خواهد شد می‌کند. در این تمرین شما به دو روش این

^۱ Muti-Threaded Design

^۲ Website

^۳ Datasets

^۴ File

^۵ Parse

^۶ Memory

مسئله را پیاده‌سازی می‌کنید. همچنین در قالب گزارش کاری که در کنار این پرونده بارگذاری شده است به مقایسه این روش‌ها و بررسی پیاده‌سازی‌های انجام شده می‌پردازید.

دقت شود که مجموعه‌های داده به منظور یادگیری شما تهیه گردیده است و مسئولیتی در خصوص محتوای کتاب‌های داده شده از سوی مسئولان تمرین قبول نخواهد شد.

معیار محبوبیت کتاب

برنامه نوشته شده باید پس از استخراج داده‌ها، اقدام به چاپ کردن اطلاعات محبوب‌ترین کتاب در ژانری که به برنامه داده شده است کند. معیار محبوبیت در این تمرین به صورت زیر تعریف می‌شود:

$$\text{معیار محبوبیت کتاب} = \frac{1}{10} \times (\text{author_average_rating} + \frac{\sum_{\text{review} \in \text{reviews_of_book}} \text{review.rating} \times \text{review.number_of_likes}}{\text{total_book_reviews_likes}})$$

- *author_average_rating*
 - میانگین امتیاز نویسنده کتاب.
- *review.rating*
 - امتیازی که کاربر برای کتاب ثبت کرده است.
- *review.number_of_likes*
 - تعداد دفعاتی که نظر کاربر مورد پسند سایرین قرار گرفته است.
- *total_book_reviews_likes*
 - تعداد کل دفعاتی که نظراتی که برای کتاب مورد بررسی ثبت شده است مورد پسند سایرین قرار گرفته است.

پیاده‌سازی سری⁷

در این بخش از تمرین شما به پیاده‌سازی سری برنامه خواسته شده می‌پردازید. سعی کنید در این بخش از تمرین بهترین پیاده‌سازی که می‌توانید را از لحاظ زمان اجرا انجام دهید. پس از انجام این بخش از تمرین به کامل کردن بخش مربوطه در گزارش کار اقدام کنید.

⁷ Serial

پیاده‌سازی چندریسه‌ای

در این بخش از تمرین به موازی‌سازی اعمال صورت گرفته در توابعی که در بخش قبل به عنوان Hotspot⁸ از آن‌ها یاد کردید می‌پردازید.

- تجزیه کردن پرونده‌ها و ذخیره‌سازی آن‌ها در حافظه از اعمال زمان‌گیر در خیلی از برنامه‌ها است که احتمالاً از توابع مربوط به آن‌ها (در کنار سایر توابع) به عنوان Hotspot های برنامه یاد کرده‌اید.

- برای موازی‌سازی این بخش می‌توانید مجموعه‌های داده‌ای که در اختیارتان قرار گرفته است را به پرونده‌هایی کوچکتر تقسیم کرده و اعمال مربوطه را توسط چندین ریسه انجام دهید.

- برای مثال می‌توانید اطلاعات مربوط به کتاب‌ها را که در پرونده books.csv آمده است را به چهار پرونده کوچکتر شکسته (books_[0-3].csv) و استخراج هر پرونده را توسط یک ریسه انجام دهید.

- **دقت کنید** که مجاز به تغییر ساختار مجموعه‌های داده، از قبیل تغییر ستون‌های مربوطه نیستید و تنها می‌توانید نمونه‌های موجود در هر پرونده را بین پرونده‌های کوچکتر تقسیم کنید.

- **توجه کنید** که اگر موازی‌سازی را بدون تقسیم مجموعه‌های داده به پرونده‌های کوچکتر انجام دهید به شما **نمره امتیازی** تعلق خواهد گرفت. بطور مثال برای استخراج اطلاعات کتاب‌ها تنها از یک پرونده books.csv استفاده کنید و این پرونده به پرونده‌های کوچکتر شکسته نشود.

دقت کنید که این بخش از تمرین باید به صورت **چندریسه‌ای** پیاده‌سازی گردد و سایر پیاده‌سازی‌ها قابل قبول نیست. پس از انجام این بخش از تمرین به کامل کردن بخش مربوطه در **گزارش کار** اقدام کنید.

⁸ توابعی که در برنامه‌تان بیشترین زمان اجرا را به خود اختصاص می‌دهند.

ورودی و خروجی برنامه

برای اجرای برنامه ژانری که قصد جستجو در آن وجود دارد از طریق آرگومان‌های خط فرمان⁹ به برنامه شما داده می‌شود. نمونه اجرای برنامه در زیر آمده است:

نمونه اجرا
./Goodreads.out Historical

قالب و نمونه خروجی این اجرای برنامه در زیر آمده است:

قالب خروجی
id: <id> Title: <title> Genres: <book_genre_1>, <book_genre_2> Number of Pages: <number_of_pages> Author: <author_name> Average Rating: <average_rating>

برای نمایش میانگین امتیازهای کتاب، مقدار آن را گرد شده تا دو رقم اعشار نمایش دهید.

نمونه خروجی
id: 138685 Title: Dissolution Genres: Historical, Mystery Number of Pages: 456 Author: C.J. Sansom Average Rating: 0.75

نکات تکمیلی

- تمام خروجی‌های برنامه را در جریان خروجی استاندارد¹⁰ چاپ کنید.
- در صورتی که در محاسبه معیار محبوبیت کتاب تقسیم بر صفر رخ داد، مقدار حاصل تقسیم را برابر با صفر در نظر بگیرید.

⁹ Command Line

¹⁰ Standard Output Stream

- همچنین تضمین می‌شود که ورودی‌هایی که به برنامه شما داده می‌شود صحیح هستند و نیازی به بررسی صحت ورودی توسط برنامه شما نیست.
- **دقت شود** که تجزیه پرونده‌های CSV بخشی از تمرین است و استفاده از کتابخانه‌های مربوطه جهت تجزیه پرونده‌ها قابل قبول نیست.

نحوه‌ی تحویل

- برای انجام پروژه یک مخزن¹¹ خصوصی در سایت **Gitlab** ایجاد کنید و حساب UT_OS_TA را با دسترسی Maintainer به مخزن خود اضافه کنید (برای عبور از تحریم می‌توانید از سرویس رایگان **شکن** استفاده کنید).
- طراحی درست، کارایی¹² برنامه و شکستن برنامه به بخش‌های کوچکتر تأثیر زیادی در نمره‌ی تمرین دارد.
- به منظور سنجش توانایی شما در شکستن برنامه بخش‌های کوچکتر از Commit هایی که در مخزن مربوطه در Gitlab انجام داده‌اید بهره گرفته خواهد شد. بنابراین سعی کنید Commit های معناداری داشته باشید و از Commit هایی که تغییرات زیادی در پرونده‌های پروژه صورت گرفته است **خودداری فرمایید** چرا که موجب کسر نمره شما خواهد شد.
- **دقت کنید** که مخزن شما حتماً باید دارای **دو شاخه**¹³ مجزا باشد که در یک شاخه پیاده‌سازی سری و در شاخه دیگر پیاده‌سازی موازی آورده شده است.
- برنامه شما باید در سیستم عامل لینوکس و با مترجم g++ با استاندارد c++11 ترجمه و در زمان معقول برای ورودی‌های آزمون اجرا شود.
- **دقت کنید** که پروژه شما باید دارای Makefile باشد. همچنین در Makefile خود مشخص کنید که از استاندارد c++11 استفاده می‌کنید.
- نکته‌هایی که در جلسه توجیهی تمرین گفته می‌شود و یا در فروم‌های مربوطه مطرح می‌شود بخشی از تمرین هستند؛ بنابراین به آن‌ها توجه داشته باشید.
- برای تحویل پروژه گزارش کار خود را در قالب PDF بارگذاری کرده و آدرس مخزن را به همراه **دو شناسه** که معرف آخرین Commit های شما در شاخه‌های سری و موازی است را در CECM قرار دهید.
- هدف این تمرین یادگیری شماسست. لطفاً تمرین را خودتان انجام دهید. در صورت کشف تقلب مطابق قوانین درس با آن برخورد خواهد شد.

¹¹ Repository

¹² Performance

¹³ Branch