



## تمرین کامپیوتری سوم



سیستم‌های عامل - بهار ۱۳۹۹

گزارش کار

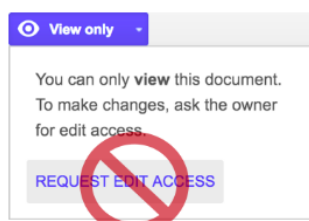
دانشکده مهندسی برق و کامپیوتر

نام و نام خانوادگی:

تاریخ:

استاد:

مهدی کارگهی

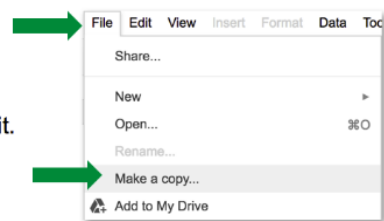


### How to use this template:

This is a view-only file and cannot be edited.

**Create your own copy** of this template to edit.

In the menu, click **File > Make a copy...**



### فهرست مطالب

2	مقدمه
3	پیاده‌سازی سری
3	سوال اوّل
3	سوال دوم
4	سوال سوم
4	جدول اوّل
4	پیاده‌سازی چندریسه‌ای
4	سوال چهارم
5	سوال پنجم
5	جدول دوم

## مقدمه



در این مسئله به تحلیل داده‌هایی که از وبگاه Goodreads استخراج شده است پرداخته شده است. در ابتدا برنامه اقدام به خواندن و تجزیه<sup>1</sup> مجموعه داده‌های داده شده می‌کند و آن‌ها را در حافظه<sup>2</sup> خود ذخیره می‌کند. پس از استخراج داده‌ها، برنامه اقدام به بدست آوردن محبوب‌ترین کتاب در ژانری مشخص، بر اساس معیاری که توضیحات تمرین آمده است می‌کند. این تمرین به دو روش این مسئله پیاده‌سازی شده است که در ادامه گزارش، نتایج حاصل آمده است.



---

<sup>1</sup>Parse

<sup>2</sup>Memory

## پیاده‌سازی سری

### سوال اول

چرا برای پیاده‌سازی یک برنامه بصورت چندریسه‌ای، بهتر است ابتدا این برنامه بصورت سری پیاده‌سازی شود؟

جواب

تا hotspot های برنامه مشخص شوند و آن قسمت هایی از برنامه که امکان موازی سازی آنها وجود دارد شفاف شوند.

### سوال دوم

در پیاده‌سازی سری، چه نکاتی را برای افزایش کارایی<sup>3</sup> برنامه در نظر گرفتید؟ (ساختمان‌های داده<sup>4</sup> مورد استفاده، Option های کامپایلر و...)

جواب

خواندن خط به خط فایل ها بجای کاراکتر به کاراکتر برای کمینه کردن I/O overhead

Allocate کردن instance ها روی استک به جای هیپ

مرتب سازی review ها بر اساس book\_id و جستجوی باینری review ها برای حساب کردن محبوبیت هر کتاب به جای هر بار جستجوی کامل review ها برای هر کتاب.

پاس دادن آرگومان های توابع با reference به منظور حذف overhead کپی کردن آرگومان ها.

بهینه سازی حداکثری با 03 flag-.

---

<sup>3</sup> Performance

<sup>4</sup> Data Structures

## سوال سوم

با بررسی زمان اجرای بخش‌های مختلف برنامه،<sup>5</sup> Hotspot های برنامه را مشخص کنید.

جواب

خواندن از فایل‌ها

جستجوها

## جدول اول

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را بازای ورودی نمونه‌ای که در شرح تمرین آمده است، در جدول زیر بیاورید.

میانگین	اجرای ششم	اجرای پنجم	اجرای چهارم	اجرای سوم	اجرای دوم	اجرای اول
0.503	0.498	0.501	0.528	0.491	0.518	0.485

## پیاده‌سازی چندریسه‌ای

### سوال چهارم

اگر هنگام موازی‌سازی برنامه به زمان اجرای بیشتری نسبت به حالت سری برخورد کنید، چه رویکردهایی را برای کاهش زمان اجرا و استفاده حداکثری از موازی‌سازی پیش می‌گیرید؟

جواب

بجای استفاده از یک container مشترک و استفاده از mutex برای جلوگیری از concurrency، به تعداد رشته‌ها container ایجاد و در نهایت بعد از اتمام کار همه thread ها، container هارا concatenate میکنیم.

---

<sup>5</sup>توابعی که در برنامه‌تان بیشترین زمان اجرا را به خود اختصاص می‌دهند.

برای بهره‌وری هرچه بیشتر از چند رشته‌ای تعداد رشته‌ها را برابر با تعداد هسته‌های منطقی معمول سیستم‌ها (4 رشته) در نظر می‌گیریم.

سوال پنجم

در هنگام پیاده‌سازی این بخش، به چه چالش‌هایی برخورد کردید و بیان کنید که به چه صورت آن‌ها را رفع کردید.

جواب

جلوگیری از rival race:

که در جواب سوال قبل چالش اصلی و رفع آن توضیح داده شده است.

جدول دوم

زمان‌های اجرای ۶ اجرای متوالی از برنامه و میانگین آن‌ها را برای ورودی نمونه‌ای که در شرح تمرین آمده است، در جدول زیر بیاورید.

میانگین	اجرای ششم	اجرای پنجم	اجرای چهارم	اجرای سوم	اجرای دوم	اجرای اول
0.365	0.369	0.356	0.362	0.373	0.363	0.368

میزان تسریع برنامه نسبت به حالت سری را نیز در زیر بیاورید.

میزان تسریع	میانگین زمان اجرای موازی	میانگین زمان اجرای سری
29%	0.365	0.503