



## تمرین کامپیوتری سوم



سیستم‌های عامل - پاییز ۱۳۹۹

گزارش کار

دانشکده مهندسی برق و کامپیوتر

نام و نام خانوادگی: نازنین یوسفیان

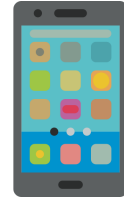
تاریخ: 24/8/99

استاد:

دکتر مهدی کارگهی

|   |                       |
|---|-----------------------|
| 2 | مقدمه                 |
| 3 | پیاده‌سازی سری        |
| 3 | سوال اول              |
| 3 | سوال دوم              |
| 3 | جدول اول              |
| 3 | پیاده‌سازی چندریسه‌ای |
| 3 | سوال سوم              |
| 4 | سوال چهارم            |
| 4 | سوال پنجم             |
| 4 | جدول دوم              |

## مقدمه



در این تمرین به تحلیل داده‌هایی که از مشخصات و قیمت فروش گوشی‌های موبایل جمع‌آوری شده‌است پرداخته شده است. در ابتدا برنامه اقدام به خواندن و تجزیه مجموعه داده<sup>1</sup>ی ارائه شده می‌کند و آنها را در حافظه خود ذخیره می‌کند. پس از استخراج داده‌ها و ویژگی‌های آنها، برنامه اقدام به نرمال‌سازی<sup>2</sup> داده‌ها و در نهایت اقدام به تعیین طبقه قیمتی گوشی‌ها می‌کند. این تمرین به دو روش این مسئله پیاده‌سازی شده است که در ادامه گزارش، نتایج حاصل آمده است.



---

<sup>1</sup> Dataset

<sup>2</sup> Data Normalization

## پیاده‌سازی سری

### سوال اوّل

چرا برای پیاده‌سازی یک برنامه بصورت چندریشه‌ای، بهتر است ابتدا این برنامه بصورت سری پیاده‌سازی شود؟  
چون task ها مشخص می شوند و می توان تصمیم گرفت که کدام یک از آن ها را می توان به صورت موازی پیاده سازی کرد.  
همچنین قسمت هایی که زمان زیادی برای اجرا می برند نیز مشخص می شود و می توان اقدام به موازی سازی آن ها کرد.

### سوال دوم

با بررسی زمان اجرای بخش‌های مختلف برنامه، Hotspot<sup>3</sup> های برنامه را مشخص کنید.  
تابع read\_data() که برای خواندن داده ها از فایل استفاده می شود و هنگام خواندن مینیمم و ماکزیمم هر ستون را نیز محاسبه می کند.  
تابع normalize() که برای نرمالیزه کردن داده ها استفاده می شود و در آخر تابع calculate() که تعداد جواب های به دست آمده درست را محاسبه می کند.

### جدول اوّل

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 اوّل |
|---------|-----------|------------|-------------|-----------|-----------|------------|
| 0.059   | 0.064     | 0.058      | 0.055       | 0.057     | 0.065     | 0.059      |

<sup>3</sup> توابعی که در برنامه‌تان بیشترین زمان اجراها را به خود اختصاص می‌دهند.

## پیاده‌سازی چندریسه‌ای

### سوال سوم

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؟

افزایش زمان اجرا ممکن است به این دلیل باشد که از lock استفاده کردیم و thread ها نیاز دارند با داده ای که قفل شده است کار کنند ولی باید منتظر ریسه ای که آن را قفل کرده است بمانند تا کارش تمام شود و قفل را آزاد کند. برای رفع این مشکل باید در جاهای محدودی از قفل استفاده کنیم و استفاده از آن را کمتر کنیم.

هم چنین اگر تعداد ریسه هایی که می‌سازیم زیاد باشند، زمانی که باری context switch استفاده می شود زیادتر می شود و باعث می شود بیشتر طول بکشد. ساخت خود ریسه ها نیز زمانی را به خود اختصاص می دهد. اگر تعداد ریسه ها به اندازه تعداد هسته ها باشد سریع تر می شود.

### سوال چهارم

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

در ابتدا 4 ریسه می ساختم و هر کدام از آنها داده مربوط به خود را از فایل مشخص شده می خواند و مقدار مینیمم و ماکزیمم کلی را قفل کرده و آپدیت می کرد. سپس عملیات نرمالایز را انجام میداد. این باعث می شد که مثلاً یک ریسه در حال انجام عملیات نرمال سازی باشد در صورتی که یک ریسه دیگر مقدار مینیمم و ماکزیمم را آپدیت نکرده باشد و این مقدار به درستی ثبت نشده باشد.

برای حل این مشکل ابتدا 4 ریسه ساختم که عملیات خواندن از فایل و به دست آوردن مینیمم و ماکزیمم را انجام دهند و هنگامی که این کار تمام شد (بعد از join) چهار ریسه دیگر ساختم که بقیه الگوریتم را اجرا کنند.

## سوال پنجم

با توجه به تجربه‌ای که در پیاده‌سازی این تمرین بدست آوردید، به نظر شما در چه مواقعی از قفل<sup>4</sup> در یک طراحی چندریسه‌ای ضروری است؟ تاثیر استفاده از قفل‌ها را بر روی کارایی<sup>5</sup> سامانه بیان کنید.

در هنگامی که ریشه‌ها از یک داده مشترک استفاده می‌کنند، برای تغییر آن داده باید از قفل استفاده شود. ممکن است که دو ریشه سعی کنند همزمان مقدار یک داده را تغییر دهند که مورد نظر ما نیست و این اتفاق نباید همزمان بیفتد.

استفاده از قفل باعث پایین آمدن کارایی سامانه می‌شود.

## جدول دوم

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 |
|---------|-----------|------------|-------------|-----------|-----------|-----------|
| 0.028   | 0.032     | 0.023      | 0.025       | 0.032     | 0.033     | 0.023     |

میزان تسریع (  $\frac{Serial\ Time}{Parallel\ Time}$  ) برنامه نسبت به حالت سری را در زیر بیاورید.

| میزان تسریع | میانگین زمان اجرای موازی | میانگین زمان اجرای سری |
|-------------|--------------------------|------------------------|
| 2.107       | 0.028                    | 0.059                  |

<sup>4</sup> Lock

<sup>5</sup> Performance