

پروژه ی درس ساختمان داده

استاد : دکتر حسین رحمانی

طراح : سایین اعلا

مریم جعفری

یزدان ماستری فراهانی

اهداف پروژه :

1. طراحی و انتخاب داده ساختار مناسب برای ذخیره سازی و بازیابی اطلاعات
2. ذخیره‌ی درست اطلاعات در RAM و عدم مغایرت اطلاعات ذخیره شده با ورودی
3. تلاش در جهت ارائه الگوریتم بهینه
4. پیاده‌سازی توابع لازم با کمینه‌ی پیچیدگی زمانی و حافظه‌ای
5. آشنایی با نحوه‌ی کار تیمی و تقسیم کار
6. آشنایی بیشتر با مباحث Search Engine و Text Mining

مقدمه^۳

همانطور که تا به حال در درس ساختمان‌های داده، با ساختارهای مناسب جهت ذخیره و بازیابی اطلاعات به همراه تحلیل پیچیدگی زمانی و حافظه‌ای آن‌ها آشنا شده‌اید، در این پروژه از شما انتظار می‌رود با توجه به مطالبی که تا کنون آموخته‌اید، مناسب‌ترین داده ساختار را با کمینه‌ی پیچیدگی زمانی و حافظه‌ای، طراحی و پیاده‌سازی کنید.

برخی از نکات مربوط به پروژه به شرح زیر است :

- پروژه در قالب تیم‌های سه نفره باید انجام گیرد.
- در پیاده سازی پروژه می‌توان از هر کتابخانه و ابزار موجود استفاده کرد. (هنگام ارائه در مورد کتابخانه و داده ساختارهای استفاده شده و دلیل استفاده از آن سوال می‌شود و باید قادر به توضیح مفاهیم باشید)
- پروژه بخش مهمی از نمره را شامل می‌شود. پس انجام آن را به روز های پایانی موکول نکنید.
- برای ارزیابی عادلانه دانشجویان، تنها زبان برنامه نویسی **C#** برای پیاده‌سازی پروژه قابل استفاده خواهد بود.

فاز های مختلف پروژه

1. فاز پیاده‌سازی پروژه و آپلود در کوئرا : پس از طراحی و پیاده‌سازی پروژه فایل‌های خود را در سایت کوئرا آپلود کنید. آخرین مهلت ارسال فایل های پروژه ۳۰ دی ساعت ۱۲ می‌باشد.
2. ارائه : پس از ارسال پروژه ملزم به ارائه‌ی پروژه می‌باشید. در ارائه از ساختار کد و نحوه‌ی پیاده‌سازی آن سوال می‌شود و آمادگی پاسخ‌گویی به سوالات را داشته باشید. ارائه‌ی پروژه با هماهنگی منتور هر گروه در تاریخ های ۳و ۴ بهمن انجام خواهد شد.

توضیحات پروژه

تمام کتاب‌ها و اسناد کتابخانه‌ای آقای X به هم ریخته شده و پاره شده است. می‌دانیم که هر برگه شامل ۹۹ خط می‌باشد که هر خط ۱۰ کلمه دارد. آقای X از شما برای مرتب کردن کتاب‌هایش کمک کنید.

یک راهنما نیز برای مرتب کردن کتابخانه در اختیار شما قرار داده می‌شود. (فایل Genres) این فایل شامل ۱۸ ژانر می‌باشد که هر کدام دارای ۱۱۰ کلمه می‌باشد آقای X می‌گوید که هر برگه فقط شامل ۹ ژانر می‌باشد.

قسمت اول (۱۰ نمره)

در ابتدا آقای X برای پیدا کردن برگه‌های کتاب‌هایش نیاز به کمک دارد و از شما می‌خواهد که تعداد برگه‌هایی را پیدا کنید که :

کلمه‌ی 'temperature__68c' در آن حداقل ۱۰ بار تکرار شده است. (جواب = ۶)

کلمه‌ی 'SquareNumber_9' در آن حداقل 17 بار تکرار شده است. (جواب = ۴)

کلمه '12:45' در آن حداقل 3 بار تکرار شده است. (جواب = ۲)

تعداد درخواست‌های او به قدری زیاد است که باید برنامه‌ای برای آن پیاده کنیم. تابعی بنویسید که با گرفتن کلمه و حداقل تعداد تکرار آن، تعداد برگه‌ها را به ما برگرداند.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۱۰ ثانیه

- تعداد برگه‌ها : ۱۰۰/۰۰۰

قسمت دوم (۱۰ نمره)

ژانر اصلی : ژانری می باشد که در یک برگه ژانر میانه آن برگه باشد یا به عبارت دیگر ژانر پاراگراف وسط آن برگه باشد.

آقای X از شما می خواهد که برای هر ژانر تعداد برگه هایی که آن ژانر، ژانر اصلی می باشد را بیابید برای این منظور تابعی بنویسید که با گرفتن ژانر به عنوان ورودی تعداد برگه های مربوط به آن ژانر را برگرداند.

به طور مثال در برگه اول ژانر اصلی ژانر Colors است

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۱۰ ثانیه

- تعداد برگه ها : ۱۰۰/۰۰۰

قسمت سوم (۱۰ نمره)

مجاورت ساده : دو برگه با هم مجاورت دارند اگر N کلمه‌ی مشترک در یک ژانر داشته باشند.

حال آقای X به شما یک برگه می‌دهد و از شما می‌خواهد برگه‌هایی که در ژانر G با آن مجاورت دارند را بیابید و به ازای هر برگه مقدار N را به شما می‌دهد. تابعی بنویسید که با گرفتن برگه و ژانر و N ، تعداد و شماره برگه‌هایی که مجاورت ساده با برگه‌ی ورودی دارند را به ما برگرداند.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۲۰ ثانیه

- تعداد برگه‌ها : ۱۰/۰۰۰

برگه‌ها : $(Number * 10_000, (Number + 1) * 10_000)$

$Number =$ شماره یکان گروه

قسمت چهارم (۱۰ نمره)

مجاورت : دو برگه با هم مجاورت دارند اگر در N کلمه ی مشترک در یک ژانر داشته باشند.

حسنى به شما يك برگه مى‌دهد و از شما مى‌خواهد برگه‌هایی که با آن مجاورت دارند را بیابید و به ازای هر برگه مقدار N را به شما مى‌دهد. تابعی بنویسید که با گرفتن برگه و N تعداد و شماره برگه‌های مجاور برگه‌ی ورودی را به ما برگرداند.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۲۰ ثانیه

- تعداد برگه ها : ۵۰۰۰

برگه ها : $((\text{Number} - 1) * 5_000, \text{Number} * 5_000)$

$\text{Number} =$ شماره گروه

قسمت پنجم (۱۰ نمره)

مجاورت تام : دو برگه با هم مجاورت تام دارند اگر در M ژانر N کلمه‌ی مشترک داشته باشند. (اگر در متن X کلمه W ده بار آمده باشد و در متن Y کلمه‌ی W بیست بار آمده باشد کلمه‌ی W ده بار مشترک است)

آقای X به شما یک برگه می‌دهد و از شما می‌خواهد برگه‌هایی که با آن مجاورت دارند را بیابید و به ازای هر برگه مقدار N را به شما می‌دهد. تابعی بنویسید که با گرفتن برگه و N و M تعداد و شماره برگه‌هایی که مجاورت با برگه‌ی ورودی دارند را به ما برگرداند.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۲۰ ثانیه

- تعداد برگه‌ها : ۱۰۰۰

برگه‌ها : $(Number * 1000, (Number + 1) * 1000)$

$Number =$ شماره گروه

قسمت ششم (۱۵ نمره)

کتاب : برگه‌هایی که در یک ژانر مشخص با هم مجاورت داشته باشند.

آقای X از شما می‌خواهد که در همه‌ی ژانرها تعداد کتاب‌ها را به او برگردانید. برای این منظور تابعی بنویسید که با گرفتن ژانر تعداد کتاب‌های ممکن برای آن ژانر و برگه‌های هر کتاب را برگرداند.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۲۰ ثانیه

- تعداد برگه‌ها : ۵۰۰

برگه‌ها : $(\text{Number} * 500, (\text{Number} + 1) * 500)$

Number = شماره گروه

قسمت هفتم (۱۵ نمره)

آقای X از شما می خواهد که در یک ژانر کتاب هایی را برگردانید که کلمه ی W در آن حداقل N بار آمده باشد .
تابعی بنویسید که با گرفتن ژانر و کلمه و N تعداد کتاب های موجود و برگه های هر کتاب را بیابد.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۲۰ ثانیه

- تعداد برگه ها : ۵۰۰

برگه ها : $(Number * 500, (Number + 1) * 500)$

Number = شماره گروه

قسمت هشتم (۲۰ نمره)

کتاب کامل : کتابی هست که در یک ژانر برگه هایش دو به دو باهم مجاورت داشته باشند.
آقای X از شما می خواهد بگویید بزرگترین کتاب کامل چند برگه دارد و شامل چه برگه هایی می باشد.

- حداکثر حافظه : ۱۰۲۴ مگابایت

- مدت زمان اجرا : ۵۰ ثانیه

- تعداد برگه ها : ۱۰۰

برگه ها : $(\text{Number} * 100, (\text{Number} + 1) * 100)$

$\text{Number} =$ شماره گروه

راهنمایی

1. ۱۰۰ میلیون تراکنش (به جز buffer و WriteLine) در زبان برنامه‌نویسی #c یک ثانیه زمان می‌برد.
2. هر integer چهار بایت می‌باشد.

ویژگی های محیط کاربری (Console)

رابط کاربری طراحی شده باید ویژگی های زیر را داشته باشد :

در ابتدا یک ورودی از شماره ی ۱ تا ۸ از ما بگیرد . سپس ورودی های تابع آن قسمت از پروژه را از ما بگیرد و خروجی نهایی و مدت زمان اجرای آن دستور و میزان حافظه ی استفاده شده را چاپ کند این عمل تا وقتی که عدد ۰ را وارد کنیم تکرار شود .

برای مثال عدد ۱ را وارد می کنیم سپس از ما کلمه و حداقل تعداد تکرار آن را بگیرد و تابع قسمت ۱ پروژه را اجرا کند وبعد از آن تعداد برگه های واجد شرایط (خروجی تابع) و مدت زمان اجرای برنامه و حافظه ی استفاده شده را برگرداند.

نحوه ی ارسال

کد نهایی پروژه را به صورت یک فایل به فرمت DS_Project_GroupNumber.cs در کوئرا آپلود کنید تابع مربوط به هر قسمت را به صورت Part_i نام گذاری کنید. علاوه بر آن یک فایل PDF که شامل توضیح مختصر از کار کردن توابع و روند اجرای پروژه است را با نام DS_Project_GroupNumber.pdf ارسال کنید.

ارسال فایل پروژه توسط یکی از اعضای گروه کافی است.

با آرزوی موفقیت و سلامتی