



دانشگاه گیلان

دکتر سید ابولقاسم میرروشندل

طراح: محمد دهقانپور

پروژه درس ساختمان های داده (DS992) درخت هافمن و min heap

داستان پروژه:

ممد که در یک شرکت نرم افزاری مشغول به کار است و در حال توسعه یک نرم افزار برای انتقال و ثبت اسناد الکترونیک است و بعد از به پایان رسیدن مراحل توسعه نرم افزار، نسخه آزمایشی از این نرم افزار در اختیار افراد قرار گرفته تا مورد بررسی قرار بگیرد، در این مرحله متاسفانه متوجه شده ایم که با توجه به بالا بودن حجم اسناد و محدودیت در ارسال و ثبت اسناد باید تغییری در ساختار ذخیره سازی اسناد کنیم و با توجه به این که او در درس ساختمان داده با موضوع درخت هافمن برای فشردن سازی رابطه خوبی نداشت از شما میخواهد که به او کمک کنید ...

هدف پروژه:

آشنایی با درخت هافمن و نحوه پیاده سازی آن با استفاده از min heap

صورت پروژه:

الگوریتم درخت هافمن برای فشردن سازی استفاده می شود که برای این منظور نیاز است تعداد حروف استفاده شده در متن و تعداد تکرار هر کدام از آنها را استخراج کنیم و در مراحل بعدی باید دو حرف با کمترین تعداد تکرار گرفته شود. در اینجا قرار است با استفاده از min heap یک صف اولیت بسازیم تا درخت تشکیل شود.

کلاس هیپ باید شامل توابع اضافه کردن و حذف کردن باشد

پروژه شامل دو قسمت میشود:

- 1.Encode: در این مرحله یک فایل به نام `main.txt` در اختیار شما قرار میگیرد که شامل متنی است که باید فشرده سازی روی آن اعمال شود و برنامه شما باید درخت هافمن را برای این متن تشکیل داده و به عنوان خروجی کلید فشرده سازی را نمایش دهد و همچنین متن فشرده را در فایلی به نام `encode.txt` ثبت کند.
و همچنین باید تعداد بایت مورد نیاز با استفاده از درخت هافمن و بدون استفاده از آن را در خروجی نمایش دهید (هر حرف فایل اصلی ۸ بیت و هر واحد فایل فشرده ۱ بیت در نظر بگیرید).
- 2.decode: در این مرحله کلیدها و متن فشرده در اختیار شما قرار میگیرد و شما باید متن اصلی را در فایلی به نام `decode.txt` ذخیره کنید.

نمونه برای قسمت اول:

Input:Hello World!

Output: 100010111001101010111101010100001110110011

نمونه برای قسمت دوم:

Input:100010111001101010111101010100001110110011 H:1#e:2# :1#l:3#o:2#W:1#r:1#d:1#!:1

Output:Hello World!

قسمت امتیازی:

- 1.قابلیت فشرده سازی داده های دیگر مانند تصویر و ...
- 2.استفاده از یک روش غیر بهینه برای به دست آورد کمترین تکرار و مقایسه آن با روش استفاده از `min heap` در تشکیل درخت هافمن و نمایش مقایسه آنها در نمودار.(هزینه اجرا را در دو روش برای چندین مثال به دست آورده و در نمودار نمایش دهید).

نکات تکمیلی:

- توجه کنید هر کارکتر در متن یک حرف است.
- توجه کنید هزینه اجرایی با استفاده از min heap نباید بیشتر از $O(n \log n)$ باشد.
- پیاده سازی به صورت تک نفره است و هیچ محدودیتی برای زبان برنامه نویسی وجود ندارد.
- همه ساختمان داده های مورد نیاز باید پیاده سازی شوند. استفاده از ساختمان داده های آماده مجاز نیست.
- توجه کنید کد های شما باید خوانا و دارای کامنت گذاری مناسب باشد.
- بحث و بررسی میان دانشجویان آزاد است اما هر دانشجو موظف است پروژه را به تنهایی انجام دهد و در هنگام تحویل، به تمام جزئیات کد کاملاً مسلط باشد. با موارد **تقلب** و **کپی کردن**، طبق تشخیص دوستان حل تمرین، برخورد جدی خواهد شد.
- زمان بندی و چگونگی تحویل پروژه، متعاقباً اعلام خواهد شد.