

پیشنهاد دهنده کیبورد

ساختمان داده‌ها - دکتر آرش شفیعی

پروژه پایانی - شماره ۱

دانشگاه اصفهان

پاییز ۱۴۰۳

—

گام های پروژه

گام اول:

پروژه جدیدی بسازید

گام دوم:

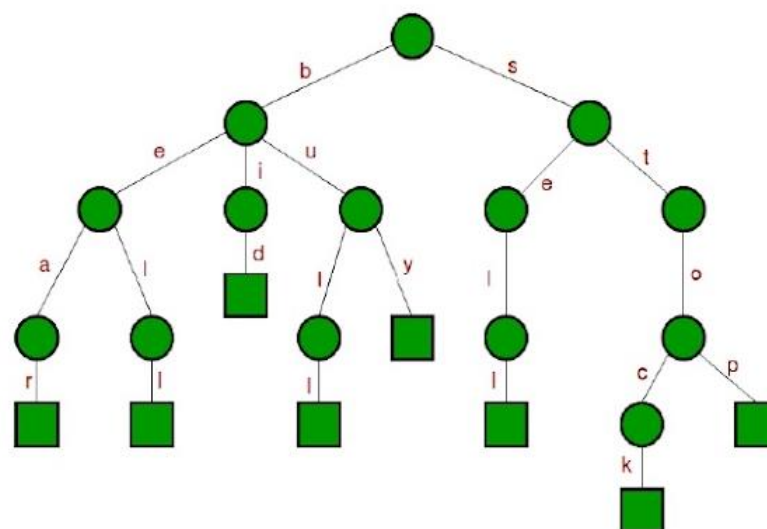
ما فایل های اسنادی داریم که حاوی کلمات انگلیسی هستند.

<https://star-academy.github.io/codestar-documents/assets/files/the-20-newsgroups-b28960092a8cf8e833bba736d4f3d433.zip>

اسناد داده شده را بخوانید و به نحوی ویرایش کنید که فاقد هر گونه علائم نگارشی بوده و کلمات آن با اسپیس از هم جدا شده باشد. (کاراکتر اسپیس جداکننده تمامی کلمات است).

گام سوم:

درخت زیر را مشاهده کنید.



در واقع در این درخت، حروف تشکیل دهنده کلمات متن داده شده را روی یال ها به گونه ای پخش می‌کنیم که با پیمایش از ریشه درخت به سمت برگ ها، با رسیدن به هر برگ، مجموعه حروف پیموده شده یکی از کلماتی است که در متن وجود دارد. با استفاده از این روش برای پیدا کردن یک کلمه با طول m در یک متن با طول n به جای اینکه پیچیدگی زمانی از مرتبه $O(n)$ باشد، مرتبه زمانی ما برابر $O(m)$ خواهد بود.

همچنین در هر برگ لیست نام اسنادی که کلمه موجود در برگ، در آن‌ها بوده است نیز قرار دارد تا هنگام پیمایش درخت برای سرچ کلمه نام اسناد را دسترسی داشته باشیم.

برنامه شما باید برای تمامی متن‌های داده شده یک درخت واحد مشابه درخت بالا تولید کرده که عملیات‌های زیر را پشتیبانی کند:

- بررسی وجود یا عدم وجود یک کلمه در متن‌های داده شده
- سرچ کلمه و تعیین نام اسناد حاوی آن کلمه (یعنی کلمه مورد نظر را در کنسول یا سرچ باکس وارد میکند و شماره تمامی اسنادی که آن کلمه در آن‌ها وجود دارد در خروجی داده شود).
- پشتیبانی از عبارات شرطی:
برای مثال داکيومنت هایی که را پیدا کنیم که حتماً شامل عبارات `get` و `help` و همچنین حداقل یکی از عبارات `illness` و `disease` باشند و شامل عبارت `cough` نباشند.

`get help +illness +disease -cough`

- توجه کنید برای این پروژه حق استفاده از `hashMap` و مانند آن را ندارید.
- توجه کنید عملیات سرچ شما باید با استفاده از یک درخت ساخته شده انجام شود.
- دقت کنید ساختمان داده درخت را باید خودتان پیاده‌سازی کنید و حق استفاده از درخت‌های آماده را ندارید. همچنین درخت شما باید عملیات افزودن، حذف کردن، بروزرسانی جستجوی نودها را داشته باشد.

گام چهارم :

در نهایت یک رابط کاربری (کنسولی یا گرافیکی) طراحی کرده که یک رشته ورودی از کلماتی که قرار است سرچ کند دریافت کرده و پاسخ های مورد نظر را بدهد. (پاسخ های مورد نظر در واقع لیست نام اسناد شامل کلمات می باشد).

ویژگی های امتیازی:

- در صورت نبود یک کلمه در تمام متون، کلمات مشابه با یک اختلاف (تغییر در حروف، کم و زیاد شدن تعداد حروف) را نشان داده و سپس سرچ کند.

نکات تکمیلی :

- این پروژه بصورت تک نفری باید پیاده سازی شود.
- بستر پیاده سازی پروژه روی گیت هاب می باشد.
- سعی کنید هریک از بخش ها را در یک کامیت جداگانه انجام دهید.
- رعایت اصول کدنویسی تمیز بخش بسیار زیادی از نمره را به خود اختصاص می دهد و در صورتی که کد کاملاً به شکل غیر اصولی پیاده سازی شده باشد، تحویل گرفته نمی شود.
- استفاده از هر زبان، فریمورک و رابط های گرافیکی کاملاً آزاد است.