



۱. (۳۰٪) [عبارت منظم] برای هر یک از قسمت‌های زیر کد عبارت منظم را بنویسید و با استفاده از ویرایشگر notepad++ به ترتیب بر روی پیکره زیر/ اعمال نمایید (کل فایل‌های پیکره را با هم ادغام کرده و در یک فایل بگذارید). علاوه بر بیان عبارات منظم هر بخش، پیکره را پس از این اصلاحات به عنوان پاسخ تمرین بفرستید.

الف) سطرهایی که به علامت‌های «نقطه، سؤال و تعجب» ختم نشده‌اند را به سطر بعدیشان بچسبانید.

ب) فضای خالی ابتدا و انتهای همه سطرها را حذف کنید.

ج) فضاهای خالی که بیش از یک کاراکتر هستند به یک فاصله عادی (space) تبدیل کنید.

د) اگر بیش از یک نیم‌فاصله در جایی استفاده شده، به یک نیم‌فاصله تبدیل کنید.

ه) اگر پیش از علائم نگارشی « ({ [: ، ؛ ؟ . ! فضای خالی هست حذف کنید و اگر پس از این علائم فضای خالی نیست یک فاصله اضافه کنید.

و) اگر پیش از علائم نگارشی «) }] (: ، ؛ ؟ . ! فضای خالی نیست یک فاصله اضافه کنید و اگر پس از این علائم فضای خالی هست حذف کنید.

ز) کاراکترهای غیر استاندارد برای فارسی (مثل ک و ی) را نرمال‌سازی کنید.

در پیکره‌ای که در مرحله قبل نرمال‌سازی کردید، با استفاده از کد عبارت منظم موارد زیر را جستجو کنید. کد عبارت منظم مورد استفاده را نیز گزارش دهید.

۱-۱ تعداد همه فعل‌های صرف شده از ریشه "گرفتن" در زمان‌های ماضی ساده، ماضی استمراری و ماضی نقلی را در پیکره به دست آورید.

۱-۲ فعل به اصطلاح مجهول در فارسی از ساختار «ریشه + (ته/ده) + فعل شدن» ساخته می‌شود. حداقل ۱۰ مورد (غیر تکراری) از این ساختار را در پیکره بیابید و گزارش دهید.

۲. (۲۰٪) [مبدل حالت متناهی] می‌خواهیم یک FST ساده طراحی کرده و از آن استفاده کنیم.

الف) FST را با استفاده از نمادهای خروجی {N, PREP, V, ADJ} طراحی کنید که همه جملات زیر



را بپذیرد. FST شما دنباله کلمات زیر را گرفته و دنباله گرامری آن را تولید می‌کند.

- امروز هوا سرد است.
- به دانشگاه می‌رویم.
- هوای تهران آلوده است.
- لباس زیبایی پوشید.
- باز می‌گردیم.
- من امروز کلاس دارم.

PREP	ADJ	V	N
از	آلوده	است	امروز
به	زیبا	باز می‌گردیم	تهران
در	سرد	پوشید	خانه
		دارم	دانشگاه
		می‌رویم	کلاس
		می‌مانم	لباس
			من
			هوا

ب) برای FST قسمت الف، هفت تایی مرتبط با تعریف FST، $(Q, \Sigma, \Delta, q_0, F, \delta(q,i), \sigma(q,w))$ را تعیین کنید.

ج) با پیمایش جملات زیر در FST طراحی شده قست الف، خروجی مرتبط و پذیرش یا عدم پذیرش آن‌ها را تعیین نمایید.

- از تهران باز می‌گردیم.
- کلاس زیبا در دانشگاه
- من امروز در خانه می‌مانم.



۳. (۵۰٪) [پیااده‌سازی: تصحیح غلط املائی با روش لونشتاین] یک تابع با نام MyLevenshtein بنویسید که دو رشته ورودی را گرفته و فاصله بین آنها را به عنوان خروجی برگرداند. بنابراین تعریف این تابع به صورت زیر خواهد بود

$$\text{Distance} = \text{MyLevenshtein}(\text{String1}, \text{String2})$$

با استفاده از تابع لونشتاین نوشته شده، یک برنامه دیگر را بنویسید که تابع MyLevenshtein را فراخوانی کرده و فاصله بین ۷ کلمه آزمون را که دارای غلط املائی است (موجود در فایل TestWords.txt)، با ۱۶ کلمه مرجع (موجود در فایل RefWords.txt) به صورت دو به دو محاسبه کنید. خروجی برنامه شما باید جدولی مشابه آنچه در زیر می‌بینید، باشد. در گزارش خود نیز جدول کامل شده را بیاورید. بدیهی است که برای هر کلمه آزمون، آن کلمه مرجعی که مقدار فاصله کمتری دارد، به عنوان کاندید جهت اصلاح کلمه آزمون انتخاب می‌شود. برای هر کلمه آزمون، کلمه مناسب جایگزین را گزارش کنید.

کلمه آزمون	همچنونان	کالیبراسونی	قریبه	حق تدریس	بین مللی	بلوچستان	کارگروهیان
کلمه مرجع							
فره							
همچنین							
.							
.							
.							
بلوچستان							