

تکلیف شماره ۲

فایل paragraphs.docx را دانلود کرده و مراحل زیر را پیاده سازی کنید:

- فایل را با استفاده از ابزار های معرفی شده تبدیل به فایل txt کنید.
- پیش پردازش های لازم (حذف فضا های خالی اضافه، تبدیل فاصله به نیم فاصله، ایجاد فاصله مناسب بعد از ویرگول یا نقطه، و...) انجام دهید.
- برای هر پاراگراف جدول مجزا ساخته و اطلاعات زیر را در جدول قرار دهید:
 - تعداد جملات
 - تعداد کل کلمات
 - تعداد فعل ها
 - تعداد اسم ها
 -
- ۵ جمله دلخواه انتخاب کرده و:
 - همه توکن ها را پیدا کنید.
 - ریشه فعل ها و ریشه کلمات را پیدا کنید.
 - لیست کلماتی که به درستی در خروجی مراحل قبل ظاهر شده اند را پیدا کنید.

توجه: برای پیاده سازی از کتابخانه هضم استفاده شود.

توجه ۲: مراحل انجام کار را بصورت مرحله به مرحله و همراه با کد و خروجی و توضیحات مربوطه بصورت یک گزارش کامل نوشته شود.

توجه ۳: فایل های زیپ شده شامل (فایل گزارش، فایل txt پس از انجام پیش پردازش ها، فایل کد) باشد.