



تمرین چهارم درس تبدیل متن به گفتار

رفع ابهام و هنجارسازی متن

استاد درس: دکتر محمد مهدی همایون پور

## تمرین چهارم درس تبدیل متن به گفتار

در تکلیف چهارم قصد داریم تا دانشجویان با نحوه رفع ابهام و هنجارسازی (نرمالسازی) متن آشنا شوند. در واقع در تکلیف قبل فرض شده بود که متن فارسی ورودی دارای هیچ ابهام و مورد غیر نرمالی نیست، اما در این تکلیف می‌خواهیم با نحوه برخورد با این موارد نیز آشنا شویم. در این تمرین هر دانشجو در ابتدا باید تمرین‌های خواسته شده را با توضیحات کامل جواب دهد. دانشجویان علاوه بر ارسال کدها و فایل اکسل، تصاویر و توضیحاتی در خصوص این برنامه و نیز گزارشی بر روی فعالیت‌های درخواست شده در این تکلیف تحویل می‌دهند.

۱) از یک روزنامه اینترنتی یا از یک وبلاگ مثال‌هایی از موارد زیر پیدا نموده و برای رفع ابهام آنها راه حل ارائه کنید:

- الف) ابهام در تشخیص انتهای جمله و پیشنهاد راه حل رفع ابهام برای آنها (۳ مورد)
- ب) ابهام در واحدسازی متن به دلایلی چون حذف فاصله ضروری یا گذاشتن فاصله اضافی (۱۰ مورد) و پیشنهاد راه حل رفع ابهام برای ۵ مورد از آنها.
- ج) یافتن کوتاه نوشت و سر نام (۱۰ مورد) با ذکر نوع یا نحوه ساخته شدن هریک از کوتاه نوشت‌ها یا سرنام‌ها.
- د) برنامه تکلیف ۳ را طوری کامل کنید که کوتاه‌نوشت و سرنام‌های نوشته شده در قسمت "ج" را به درستی به معادل واجنویسی آن تبدیل کند.

- برای این کار باید کوتاه‌نوشت و سرنام‌های مورد نظر را در دیکشنری اضافه کنید.
- به عنوان مثال در صورت مشاهده BBC آن را به صورت \bibisi\ واجنویسی کند و یا در صورت مشاهده "ک.پ." آنرا به صورت \kodeposti\ واجنویسی کند.
- برای رفع مشکل تایپ راست به چپ در زبان فارسی به خصوص در زمانی که در بین آن از کلمات انگلیسی استفاده می‌کنید، می‌توانید فایل متنی txt را در نرم افزار Word باز کنید و یا در نرم‌افزار Notepad از دکمه Ctrl+Shift استفاده کنید.

۲) یک متن از یک روزنامه اینترنتی شامل ۱۰ نوع مختلف از موارد غیر نرمال انتخاب و اقدامات زیر را انجام دهید:

- الف) موارد غیر نرمال متن انتخاب شده را تشخیص دهید و آنها را گروه‌بندی کنید.
- ب) برای هر گروه یک برچسب مناسب پیشنهاد دهید و هر مورد غیر نرمال را توسط آن برچسب در متن انتخاب شده مشخص کنید.
- ج) برای شناسایی خودکار هر یک از گروه‌های غیر نرمال موجود در متن انتخاب شده راه حل ارائه کنید.
- د) برای هنجارسازی هر یک از گروه‌های غیر نرمال موجود در متن انتخاب شده راه حل ارائه کنید.
- ه) برنامه تکلیف ۳ را طوری کامل کنید که بتواند اعداد را به شکلی که توضیح داده می‌شود، بخواند.
- در صورت مشاهده شده یک تاریخ شمسی به فرمت روز/ماه/سال آن را به صورت کامل بخواند. به عنوان مثال در صورت مشاهده عبارت "۱۳۵۷/۱۱/۲۲" آن را به صورت زیر بخواند:

\bist-o do-e yAzdah-e hezAr-o sisad-o panjah-o haft\

## تمرین چهارم درس تبدیل متن به گفتار

- اعداد سه رقمی را رقم به رقم بخواند. به عنوان مثال با مشاهده "۶۷۲"، واج نویسی آن به صورت \SeS haft do\ ارائه نماید. برای انجام این بخش باید همه موارد مورد نیاز را به دیکشنری اضافه کنید.
- (و)- برنامه تکلیف ۳ را طوری کامل کنید که کاراکترهای غیر نرمال " ( ) - + = < > ~ \ / % \$ # @ { } [ ] " را نیز پشتیبانی کرده و به معادل واجنویسی آن تبدیل کند. به عنوان مثال کاراکتر "(" را به صورت \parAntezbAz\ واجنویسی کند. برای انجام این بخش باید این کاراکترها و واجنویسی آنها را به دیکشنری اضافه کنید.

پیامبر اکرم (صلی الله علیه و آله) :

من کان فی حاجة اخیه کان الله فی حاجته

کسی که در پی حاجت مردم باشد، خداوند نیز در پی برآوردن حاجت اوست.