



دکتر داودآبادی

زمستان ۱۴۰۳

مبانی پردازش زبان و گفتار

تمرین سری اول

طراحان: امین عباسی، جواد فرجی

تاریخ انتشار: ۱۷ اسفند

تاریخ تحویل: ۲۵ اسفند



قوانین و مقررات انجام و تحویل تمارین

۱. پاسخ سوالات را به صورت کامل با توجه به نوع و شماره هر سوال در قالب گزارش PDF در یک فایل فشرده به شکل زیر قرار داده و تا زمان تعیین شده در LMS بارگذاری نمایید.

HW-[Number]-[Name]-[LastName]-[StudentID].zip

۲. لطفا منابع استفاده شده شامل هرگونه منبع (کد، مقاله، کتاب، صفحه وب و...) در حل هر سوال را ذکر نمایید. از نظر اخلاقی نیز استفاده از منبع بدون ارجاع، سرقت علمی به حساب می‌آید.

۳. استفاده از مدل‌های زبانی هوش مصنوعی صرفاً باید در جهت کمک به فهم دقیق‌تر سوالات باشد. لطفاً در نهایت تحلیل خود را به عنوان پاسخ بنویسید.



سوالات تئوری

۱ سوال اول

با توجه به جمله زیر به سوالات پاسخ دهید.

The quick brown fox jumps over the lazy dog. The dog barked, and the fox ran away. Meanwhile, another fox appeared near the forest, and walked in a level path led to the lake.

۱.۱

یک الگوی regex بنویسید تا همه موارد fox یا dog را در متن پیدا کند.

۲.۱

الگوی regex سوال قبل را طوری تغییر دهید تا کلماتی که با f یا d شروع می‌شوند و با x یا g پایان می‌یابند (مثلاً fox یا dog) را بیابد.

۳.۱

یک الگوی regex بنویسید تا همه کلماتی که حداقل دو حرف صدادار دارند (مثلاً over یا barked) را پیدا کند.

۴.۱

یک الگوی regex بنویسید تا کلماتی که با یک حرف شروع و با همان حرف پایان می‌یابند را بیابد.

۵.۱

یک الگوی regex بنویسید تا همه کلماتی که با حرف l شروع می‌شوند و در انتهای جمله قرار دارند را پیدا کند.

۲ سوال دوم

Stemming و Lemmatization را توضیح دهید، سپس تفاوت های آنها را با مثالی در زبان فارسی بیان کنید.

۳ سوال سوم

سه کلمه و سه بردار جانمایی داده شده است، به نظر شما کدام بردار مربوط به کلمه apple است؟ چرا؟

king [0.1, -0.9, 0.2] •

queen [0.7, 0.5, 0.3] •

apple [0.8, 0.6, -0.4] •

۴ سوال چهارم

با استفاده از ۵ جمله زیر:

<s> the quick brown fox jumps over the lazy dog </s>

<s> the lazy dog barks at the quick fox </s>

<s> a quick brown fox jumps over the lazy dog </s>

<s> the lazy dog sleeps all day </s>

<s> the quick fox jumps over the lazy dog </s>

احتمال جمله زیر را با توجه به bigram های آن محاسبه کنید، از add-one smoothing استفاده کنید.

<s> the lazy fox jumps </s>



سوالات عملی

۵ سوال اول

در نوتبوک Q5.ipynb با استفاده از نامپای، جانمایی word2vec را از پایه پیاده سازی کنید.

۶ سوال دوم

در نوتبوک Q6.ipynb با توجه به توضیحات مدلی را با استفاده از دادگان Text8 یک بار در حالت CBOW و بار دیگر در حالت Skip-gram آموزش دهید و نتایج را با هم مقایسه کنید.