



به نام پروردگار
پردازش زبان طبیعی تمرین اول
موعد تحویل: ۴ اسفند ۹۸



پورامینی (pouramini@gmail.com)

سوال اول

پردازش زبان طبیعی را تعریف کنید. تفاوت آن با زبانشناسی محاسباتی (computational linguistics) چیست؟ برخی از کاربردهای آن را به شکل مختصر توضیح دهید.

سوال دوم

برای هر یک از موارد زیر یک عبارت منظم بنویسید (در هر مورد دقت کنید که نویسه‌های غیر مجاز در آن نباشد). در قسمت الف (آدرس ایمیل) به طور مختصر توضیح دهید که هر یک از قسمت‌های عبارت منظم ارائه شده چه کاری انجام می‌دهند.

- آدرس ایمیل (طول پسوند دامنه را بین دو تا حداکثر ۶ حرف در نظر بگیرید)
- تاریخ با قالب YYYY-MM-dd (سال باید با ۱ یا ۲ شروع شود و همینطور روز و ماه باید در محدوده اعداد مربوطه باشد).

- آدرس URL.

○ مثال:

<https://en.wikipedia.org/Punctuation#References>

سوال سوم

قطعه بندی متن به جملات sentence segmentation را تعریف کنید، مشکلات آن را نام ببرید، و برخی روش‌های رایج بویژه استفاده از یک طبقه‌بند مانند طبقه‌بند naïve bayse را به این منظور توضیح دهید.

سوال چهارم

الگوریتم min-edit distance در حالتی که علاوه بر عملگرهای Insert/delete دو عملگر دیگر substitution و transposition داشته باشد را در نظر بگیرید. transposition عملگری است که دو حرف کنار یکدیگر را جابجا میکند.

کمترین فاصله بین دو کلمه prosperity و properties را با رد نظر گرفتن هزینه اعمال در دو حالت زیر بدست آورید:

الف) هزینه همه اعمال را ۱ در نظر بگیرید.

ب) هزینه insert/delete = 1 و هزینه substitution/transposition = 2

سوال پنجم

مقاله زیر را مطالعه کرده و روش ارائه شده را تشریح کنید.

[An Improved Error Model for Noisy Channel Spelling Correction](#)

pdf مقاله را می توانید از آدرس زیر دانلود کنید:

<https://www.aclweb.org/anthology/P00-1037/>

لطفا به قواعد حل تمرین که در CECM قرار داده شده است توجه کنید.