



هدف تمرین:

- آشنایی با Part Of Speech tagging و مفهوم Chunk در NLTK
- آشنایی با کدگذاری IOB در NLTK
- آشنایی با Named Entity Recognition در NLTK
- آشنایی با Named Entity Recognition در Spacy

مراحل:

۱. در این تمرین شما با دو متن زیر کار می‌کنید.

- Natural language processing is fun! This text is a sample text.
- Jensen Huang, the CEO of Nvidia, the nation's most valuable semiconductor company, with a stock price of \$645 a share and a market cap of \$400 billion, is out to create the metaverse, what Huang describes "a virtual world that is a digital twin of ours." Huang credits author Neal Stephenson's Snow Crash, filled with collectives of shared 3-D spaces and virtually enhanced physical spaces that are extensions of the Internet, for conjuring the metaverse. This is already playing out with the massively popular online games like Fortnite and Minecraft, where users create richly imagined virtual worlds.

۲. در ادامه می‌خواهیم با استفاده از NLTK، عمل POS را بر روی متن a انجام دهیم.

- بررسی کنید مفهوم Chunk و non-phrase chunking در POS چیست؟
- عمل POS tagging را روی متن a انجام دهید و tag هر token را در خروجی چاپ کنید.
- بررسی کنید چگونه میتوان با عبارات منظم، عمل chunking را روی نتیجه قسمت b اعمال کرد؟
- با استفاده از عبارت منظم
NP: {<DT>?<JJ>*<NN>}
عمل chunking را روی نتیجه قسمت b اعمال کنید.
- عبارت منظم بالا و نتیجه حاصل از قسمت d را تحلیل کنید.
- درخت حاصل از قسمت d را به صورت بصری نمایش دهید. خروجی باید از لحاظ بصری مانند تصویر زیر باشد.



g. دو مثال دیگر از chunking با دو عبارت منظم دیگر متفاوت با عبارت منظم گفته شده در قسمت d بزنید و برای آن‌ها دو جمله مثال بزنید که اعمال عبارت‌های منظم مثال زده شده بر روی آن جمله‌ها باعث ایجاد حداقل یک chunk بشود. خروجی POS tagging و همچنین درخت حاصل (مانند قسمت f) را در خروجی چاپ کنید.

۳. در ادامه می‌خواهیم عمل IOB encoding را بر روی متن b اعمال کنیم.

a. بررسی کنید نحوه کدگذاری IOB در NLTK به چه صورت است؟

b. با استفاده از عبارت منظم داده شده در قسمت ۲ بخش d و اعمال آن بر روی متن b، عمل IOB

encoding را با کمک NLTK بر روی متن b اعمال کرده و نتیجه را در خروجی چاپ کنید.

c. معنی O، B-NP و I-NP در خروجی قسمت b چیست؟

۴. حال می‌خواهیم با استفاده از مدل از پیش آموزش داده شده Stanford برای NER، عمل NER را روی متن b اعمال کنیم.

a. بررسی کنید چگونه باید از مدل از پیش آموزش داده شده (ارسال شده به پیوست صورت تمرین در

گروه) را با استفاده از NLTK استفاده کرد؟

b. با استفاده از این مدل، عمل NER را بر روی متن b اعمال کنید.

c. موجودیت‌های سازمانی و نام افراد را در خروجی چاپ کنید.

۵. یکی دیگر از کتابخانه‌های پرکاربرد و خوب برای کار در زمینه NLP کتابخانه‌ی Spacy است. این

کتابخانه را نصب کنید. برای نصب آن می‌توانید به آدرس <https://spacy.io/usage> بروید و دستورات

نصب را مشاهده کنید. در ادامه می‌خواهیم با استفاده از این کتابخانه عمل NER را بر روی متن b اعمال کنیم.

a. بررسی کنید چگونه می‌توان با استفاده از کتابخانه Spacy عمل NER را انجام داد؟

b. عمل NER را بر روی متن b اعمال کنید.

c. موجودیت‌های نام افراد، نام سازمان‌ها، پول و اعداد را در خروجی چاپ کنید.

۶. نتیجه حاصل از NER قسمت ۴ و قسمت ۵ را با هم مقایسه کنید.

- نتیجه نهایی را در قالب یک فایل Jupyter notebook و یک گزارش PDF در یک فایل فشرده zip که نام آن با فرمت NLP_Name_Family_Ex4 (به جای Name اسم و به جای Family نام خانوادگی خودتان را قرار دهید) است قرار دهید و به ایمیل abedi.a1997@gmail.com ارسال کنید.

موفق باشید