

## سوالات تئوری:

1.

الف) چالش های NER را توضیح دهید.

ب) تاثیر مفهوم متن بر میزان دقت سیستم های NER را توضیح دهید.

ج) چگونگی بهبود محدودیت های HMM توسط CRF ها را توضیح دهید.

د) یک خطا در برچسب گذاری هر یک از جملات زیر که با مجموعه Penn Treebank برچسب گذاری شده اند، پیدا کنید.

- I/PRP need/VBP a/DT flight/NN from/IN Atlanta/NN
- Does/VBZ this/DT flight/NN serve/VB dinner/NNS
- I/PRP have/VB a/DT friend/NN living/VBG in/IN Denver/NNP
- Can/VBP you/PRP list/VB the/DT nonstop/JJ afternoon/NN flights/NNS

ه) توضیح دهید که روش برچسب گذاری BIO برای named entity ها چگونه استفاده می شود. و تفاوت این روش را از برچسب گذاری IO و برچسب گذاری BIOES بررسی کنید؟

## سوالات عملی:

1. نوت بوک Q1 را تکمیل کنید و نتایج هر مرحله را تحلیل کنید.

2. نوت بوک Q2 را تکمیل کنید و نتایج هر مرحله را تحلیل کنید.

3.

الف) دیتاهای موجود در فولدر data را بررسی کرده و دلایلی که ممکن است این name entity ها مشکل ساز شوند را بیان کنید.

ب) نوت بوک Q3 را تکمیل کنید.