

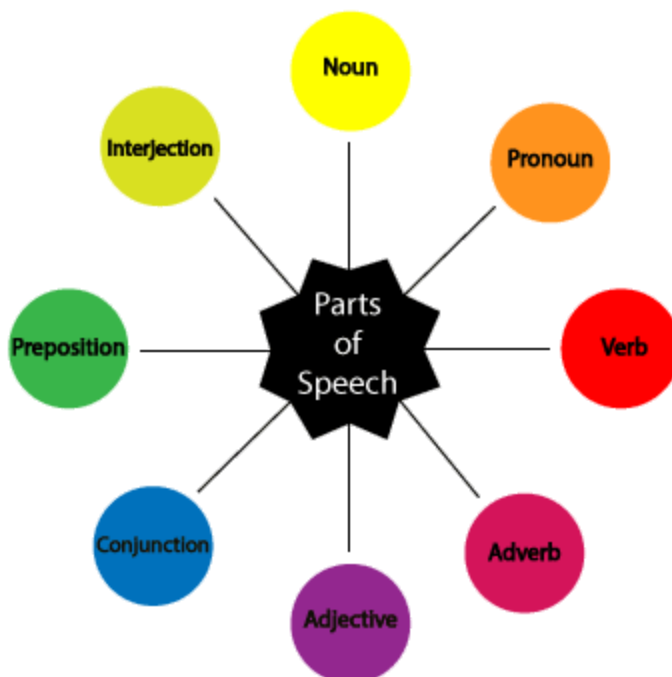
## تکلیف چهارم درس یادگیری عمیق

### شبکه های بازگشتی

این تکلیف شامل دو بخش است و در هر دو بخش به مسائل مربوط به NLP می پردازیم. در بخش اول چند مدل بازگشتی برای حل مسئله **Chunking** ارائه می دهید و نتایج را با هم مقایسه می کنید. در بخش دوم مدل هایی برای تولید نام های فارسی بر اساس شبکه های بازگشتی ارائه خواهید داد. کدها و گزارش مربوط به هر بخش را در یک **jupyter notebook** قرار دهید. داده های مورد نیاز برای شما فراهم شده است. به ایده ها و تحلیل و مقایسه های جالب نمره های تشویقی اضافه اختصاص داده می شود. برای هر بخش مدل هایی بر پایه LSTM و GRU و RNN ساده را پیاده سازی کنید و با هم مقایسه نمایید. برای انجام این تکلیف می توانید از **pytorch** یا **Tensorflow** استفاده نمایید.

### بخش اول: مساله Chunking

این مساله با مساله POS یا **Part of Speech** ارتباط نزدیکی دارد. در POS ورودی مدل یک جمله است و خروجی مدل رشته ای است که نقش هر کلمه در جمله را معین می کند. نقش های مختلف کلمات در زبان انگلیسی با شکل زیر نشان داده شده است.



تفاوت مساله **chunking** با POS در این است که در اینجا ممکن است به مجموعه ای از کلمات یک نقش نسبت داده شود. در این حالت برچسب کلمه اول این گروه با پیشوند **B** و بقیه با پیشوند **I** مشخص می شوند. به مثال زیر توجه کنید.

جمله ورودی مدل:

He reckons the current account deficit will narrow to only # 1.8 billion in September .

خروجی :

[NP He ] [VP reckons ] [NP the current account deficit ] [VP will narrow ] [PP to ] [NP only # 1.8 billion ] [PP in ] [NP September ] .

یعنی خروجی معادل **the current account deficit** عبارت خواهد بود از B\_NP I\_NP I\_NP I\_NP

برای اطلاعات بیشتر در مورد این مسئله و فرمت داده ها لینک زیر را مشاهده کنید:

<https://www.clips.uantwerpen.be/conll2000/chunking/>

مدل هایی بر اساس شبکه های عصبی بازگشتی برای انجام کار شرح داده شده ارائه و آموزش دهید. دیتاست های آموزش و آزمون لازم برای شما فراهم شده است و ضمیمه این فایل است.

بخش دوم: مسئله تولید نام جدید برای نوزادان

در این بخش فایلی در اختیار دارید که معروف ترین نام های فارسی برای نوزادان دختر و پسر را در بر دارد. مدل هایی بر اساس شبکه های عصبی بازگشتی ارائه و آموزش دهید که قادر باشند نام های جدیدی تولید کنند.