



## هدف تمرین:

- آشنایی با ابهام در گرامر
- آشنایی با ایجاد گرامر و اعمال آن بر روی جمله
- آشنایی با انواع parser در NLTK

## مراحل:

- ۱- در تمرین چهارم با مفاهیمی مثل Chunk و noun-phrase chunking و کاربرد گرامرها در POS tagging آشنا شدیم. در این تمرین می‌خواهیم با مفاهیمی چون ابهام در گرامر و probabilistic context free grammar و نحوه ایجاد درخت تجزیه آنها در NLTK آشنا شویم.
- ۲- ابتدا تابعی تعریف کنید که یک گرامر و token های مربوط یک جمله را ورودی بگیرد و با استفاده از ChartParser کتابخانه NLTK، درخت تجزیه جمله ورودی را ایجاد کند. درخت تجزیه از لحاظ ظاهری باید شبیه به درخت‌های ایجاد شده در تمرین چهارم باشد.
- ۳- در فایل‌های پیوست این تمرین، فایلی با نام Grammar1.cfg ارسال شده است. ابتدا بررسی کنید چگونه می‌توان این فایل گرامر، که در حقیقت یک فایل متنی است را در NLTK بارگذاری کنید. سپس با استفاده از این گرامر و با کمک تابعی که در بخش ۲ ایجاد کرده‌اید، درخت جمله "the man saw the dog with the telescope" را رسم کنید.
  - a. چند درخت برای این جمله ایجاد شده است؟
  - b. دلیل ایجاد هر درخت با توجه به گرامر داده شده چیست؟
  - c. تحلیل خود را به صورت کامل در فایل توضیحات بیان کنید.
- ۴- حال می‌خواهیم خود گرامر برای جملات موجود در اسلایدهای درس ایجاد کنیم و ابهام را در آنها بررسی کنیم. برای جملات زیر (که از اسلایدهای درس، قسمت مثال از ابهام برداشته شده است) گرامری ایجاد کنید، به صورتی که درخت ایجاد شده شبیه به درخت‌های موجود در اسلایدهای درس شود. نام فایل هر کدام از گرامرهای ایجاد شده را دقیقاً با نام فایل ذکر شده یکی در نظر بگیرید و فرمت آن‌ها را نیز cfg بگذارید. (راهنمایی: برای این کار می‌توانید از گرامر نمونه داده شده یعنی Grammar1.cfg و مقایسه آن با درخت‌های مربوط به این جمله در اسلاید ۴۴ درس parsing، ایده برداری کنید.)

نام فایلی که باید ایجاد کنید	اسلاید مربوطه	جمله
Grammar2.cfg	اسلاید ۴۵	he drove down the street in the car
Grammar3.cfg	اسلاید ۴۶	saw her duck with the telescope
Grammar4.cfg	اسلاید ۴۸	the fast car mechanic

۵- با استفاده از گرامر زیر، درخت‌های تجزیه جمله “i shot an elephant in my pajamas” را رسم کنید و ابهامات در آن را بررسی کرده و توضیح دهید.

```

S -> NP VP
PP -> P NP
NP -> Det N | Det N PP | 'i'
VP -> V NP | VP PP
Det -> 'an' | 'my'
N -> 'elephant' | 'pajamas'
V -> 'shot'
P -> 'in'

```

۶- درخت تجزیه جمله “the dog saw a man in the park” را با استفاده از گرامر زیر و RecursiveDescentParser رسم کنید.

```

S -> NP VP
VP -> V NP | V NP PP
PP -> P NP
V -> "saw"
NP -> Det N | Det N PP
Det -> "a" | "the"
N -> "man" | "dog" | "park"
P -> "in"

```

۷- درخت تجزیه جمله “Mary saw a dog” را با استفاده از گرامر زیر و ShiftReduceParser رسم کنید.

```

S -> NP VP
VP -> V NP
V -> "saw"
NP -> "Mary" | Det N
Det -> "a"
N -> "dog"

```

- ۸- تفاوت روند RecursiveDescentParser و ShiftReduceParser را به صورت مختصر بیان کنید.
- ۹- در ادامه می‌خواهیم نحوه کار با PCFG در NLTK را بررسی کنیم. برای این کار گرامر toy\_pcfg2 را از NLTK فراخوانی کنید. با استفاده از این گرامر درخت تجزیه جمله "Jack saw Bob with my cookie" را، با استفاده از parserهای InsideChartParser، RandomChartParser و UnsortedChartParser رسم نموده و تفاوت نتیجه حاصل بین این parserها را شرح دهید. (این سه parser را باید از پکیج parse، مائول pchart کتابخانه NLTK فراخوانی کنید).
- ۱۰- با توجه به اسلاید ۵۶ درس، گرامر مربوط به دو جدول موجود در این اسلاید را بنویسید. توجه کنید که باید احتمالات نیز در نظر گرفته شود. سپس با استفاده از ViterbiParser و گرامر ایجاد شده، درخت تجزیه "the dog sleeps" را رسم کنید و احتمال این جمله را با استفاده از NLTK محاسبه 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\_Ex7 (به جای Name اسم و به جای Family نام خانوادگی خودتان را قرار دهید) است قرار دهید و به ایمیل [abedi.a1997@gmail.com](mailto:abedi.a1997@gmail.com) ارسال کنید.

**موفق باشید**