Môn: Xử lý ngôn ngữ tự nhiên

GVHD: Ts. Nguyễn Tuấn Đăng

HVTH: Phạm Mạnh Trung Nguyên

MSSV: 2100000168

Lóp: 21MTH1A.BD

# **BÀI LÀM**

#### 1. Văn phạm CFG

### a. Nam đang đọc sách ở thư viện.

$S \rightarrow NP VP$	PROPERN → [Nam]
NP → PROPERN	ADV → [đang]
VP → ADVP V NP PP	V → [đọc]
ADVP → ADV	N → [sách]
NP → DET NOM	P → [ở]
NOM → N	N → [thư_viện]
PP → P NP	
NP → DET NOM	
$NOM \rightarrow N$	

### b. Nam thích đọc sách.

$S \rightarrow NP VP$	PROPERN → [Nam]
NP → PROPERN	$V \rightarrow [thích]$
$VP \rightarrow V VP$	V → [đọc]
$VP \rightarrow V NP$	N → [sách]
NP → DET NOM	
NOM → N	

## c. Nam có nhiều cuốn sách rất hay.

$S \rightarrow NP VP$	PROPERN → [Nam]
NP → PROPERN	V → [có]
$VP \rightarrow V NP$	Q → [nhiều]
NP → DET NOM	N → [cuốn]
DET → Q	$N \rightarrow [sách]$
NOM → N N ADJP	ADV → [rất]
ADJP → ADVP ADJ	ADJ → [hay]
ADVP → ADV	

## d. Nó mới mua hai cuốn sách mới.

SV NP VP	PRON → [N6]
NP →PRON	ADV → [mới]
VP → ADVP V NP	V → [mua]
ADVP → ADV	CD → [hai]
NP → DET NOM	N → [cuốn]
DET → CD	N → [sách]
NOM → N N ADJP	ADJ → [mới]
ADJP → ADJ	

## e. Nam tặng Lan một cuốn sách.

$S \rightarrow NP VP$	PROPERN → [Nam]
NP → PORPERN	V → [tặng]
VP → V NP NP	PROPERN → [Lan]
NP → PROPERN	CD → [một]
NP → DET NOM	UN → [cuốn]
DET → CD	$N \rightarrow [sách]$
NOM → UN N	

#### 2. Rút gọn văn phạm CFG

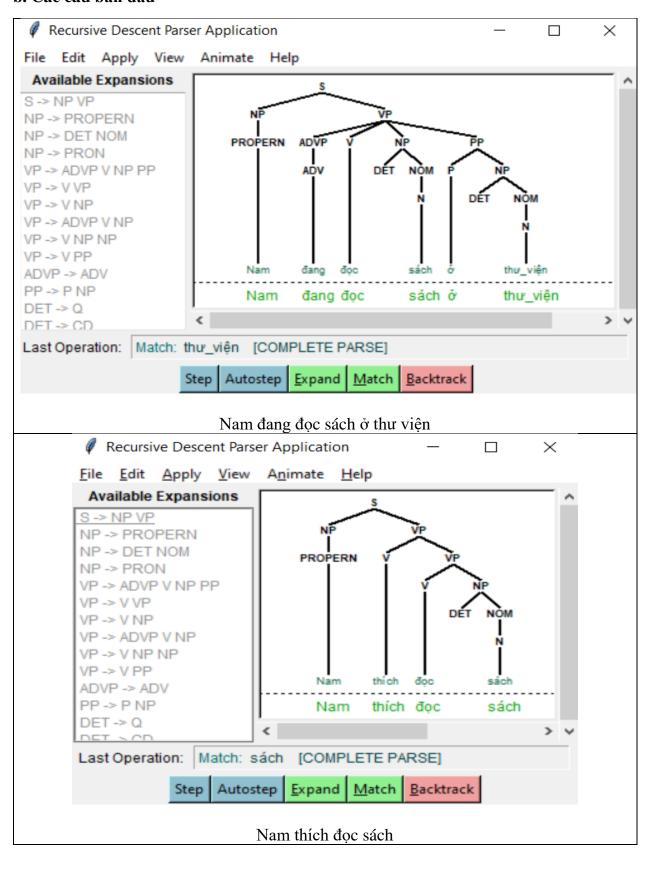
```
S \rightarrow NP VP
NP -> PROPERN | DET NOM | PRON
VP \rightarrow ADVP \ V \ NP \ PP \ | \ V \ VP \ | \ V \ NP \ | \ ADVP \ V \ NP \ | \ V \ NP \ NP \ | \ V \ PP
ADVP -> ADV
PP \rightarrow P NP
DET \rightarrow Q \mid CD \mid
ADJP -> ADJ | ADVP ADJ | ADJ ADVP
NOM \rightarrow N \mid N \mid N \mid ADJP \mid UN \mid N
PROPERN -> 'Nam' | 'Lan'
PRON -> 'Nó'
V -> 'đọc' | 'thích' | 'có' | 'mua' | 'tặng' | 'ở'
P -> 'ở'
N -> 'sách' | 'thư viện' | 'cuốn'
UN -> 'cuốn'
Q -> 'nhiều'
ADV -> 'rất' | 'mới' | 'đang' | 'hay' | 'mua'
ADJ -> 'hay' | 'mới' | 'nhiều'
CD -> 'một' | 'hai'
```

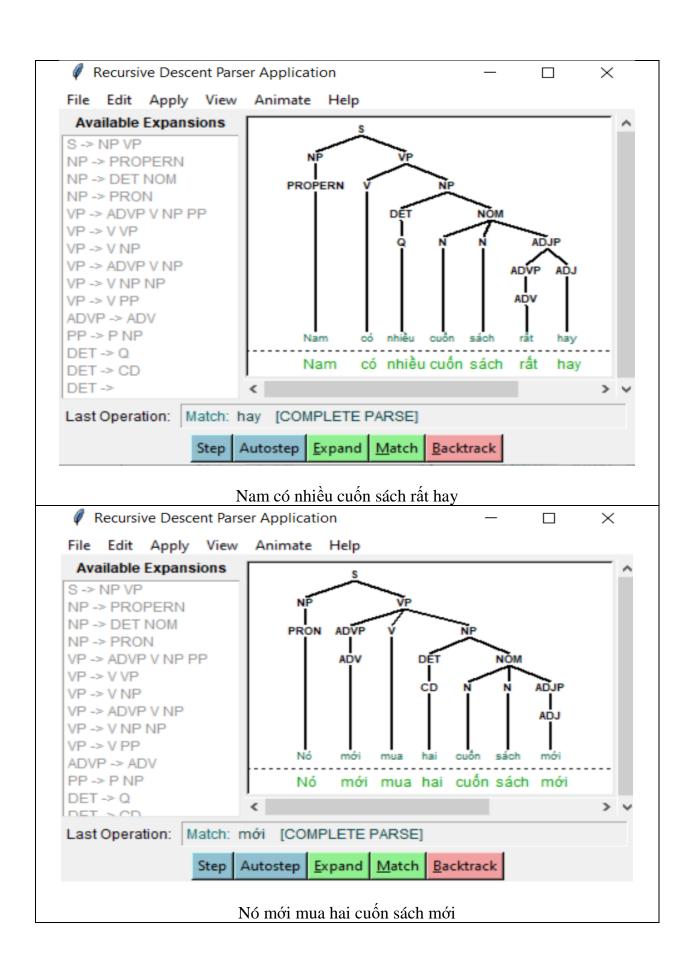
#### 3. Cài đặt thư viện NLTK trong Python để chạy chương trình

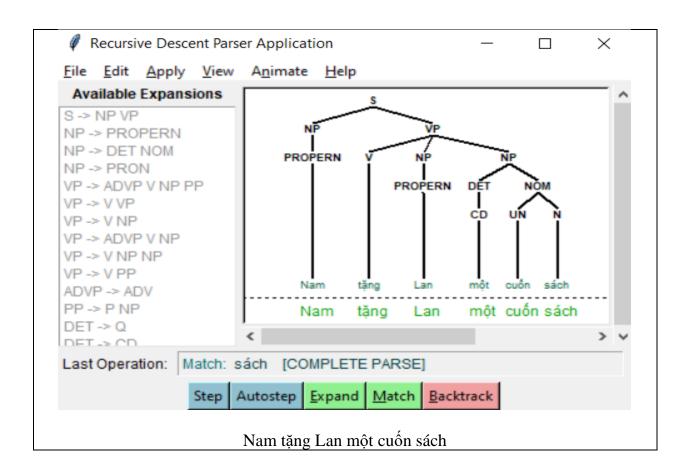
#### a. Code:

```
import nltk
grammer = nltk.CFG.fromstring("""
S -> NP VP
NP -> PROPERN | DET NOM | PRON
VP -> ADVP V NP PP | V VP | V NP | ADVP V NP | V NP NP | V PP
ADVP -> ADV
PP -> P NP
DET -> 0 | CD |
ADJP -> ADJ | ADVP ADJ | ADJ ADVP
NOM -> N | N N ADJP | UN N
PROPERN -> 'Nam' | 'Lan'
PRON -> 'Nó'
V -> 'đọc' | 'thích' | 'có' | 'mua' | 'tặng' | 'ở'
P -> 'ở'
N -> 'sách' | 'thư viện' | 'cuốn'
UN -> 'cuốn'
0 -> 'nhiều'
ADV -> 'rất' | 'mới' | 'đang' | 'hay' | 'mua'
ADJ -> 'hay' | 'mới' | 'nhiều'
CD -> 'môt' | 'hai'
""")
nltk.app.rdparser()
nltk.app.srparser()
file1 = open('baitap.txt', 'r', encoding="utf8")
Lines = file1.readlines()
for line in Lines:
    sentence = line.split()
    print(sentence)
    def parse(sent):
       a = []
        parser = new_func()
        for tree in parser.parse(sent):
            a.append(tree)
        return(a[0])
    def new_func():
        parser = nltk.ChartParser(grammer)
        return parser
    parse(sentence).draw()
```

#### b. Các câu ban đầu







#### c. Các câu phát sinh

