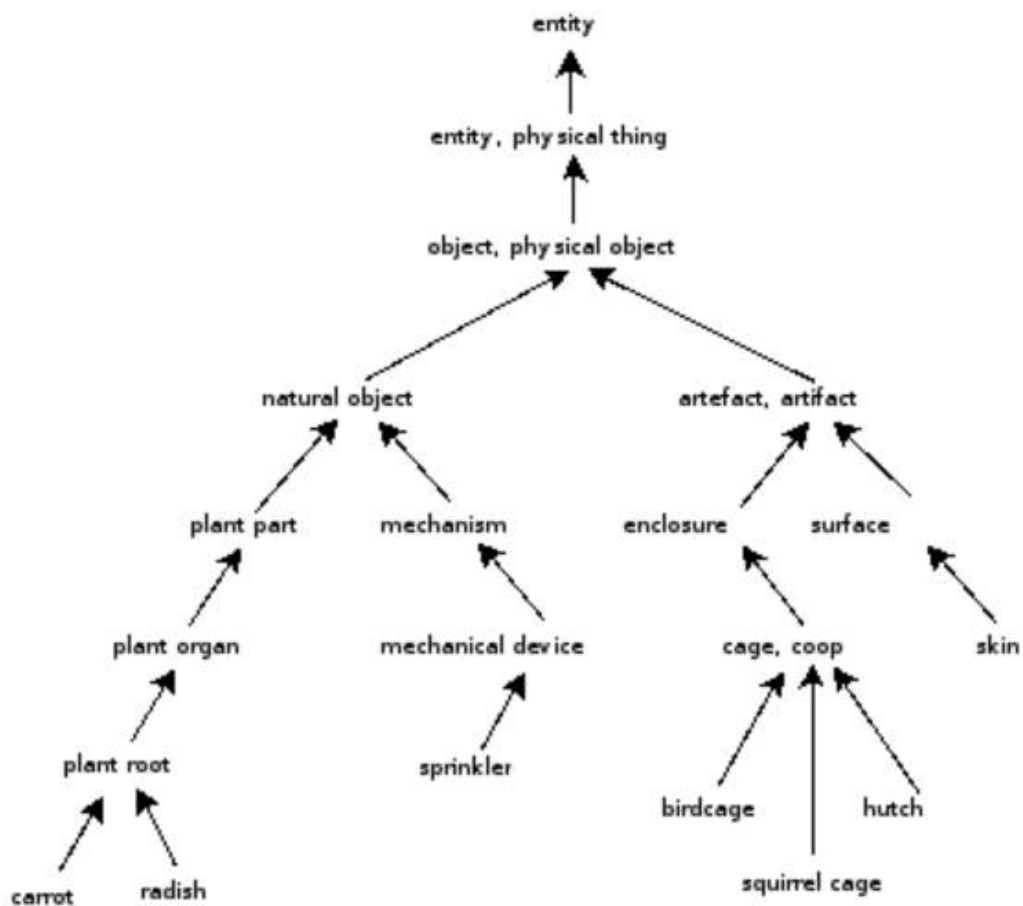


텍스트정보처리 중간고사

2021년 10월 20일 (수)

1. 자연어처리(텍스트정보처리)의 주요 task를 5개 나열하고, 각각의 task에 대해서 상세히 기술하시오.
2. 다음과 같은 WordNet 구조에서 상위어와 하위어를 설명하시오. 단, 그림에 나타난 사례를 이용하여 설명하시오.



3. 다음 코드의 실행 결과를 기술하시오.

```
rawText = "By 11 o' clock on Sunday, the doctor shall open the  
dispensary."
```

```
sTokenizer = SpaceTokenizer()
```

```
print( "Space Tokenizer 출력 :", sTokenizer.tokenize(rawText))
```

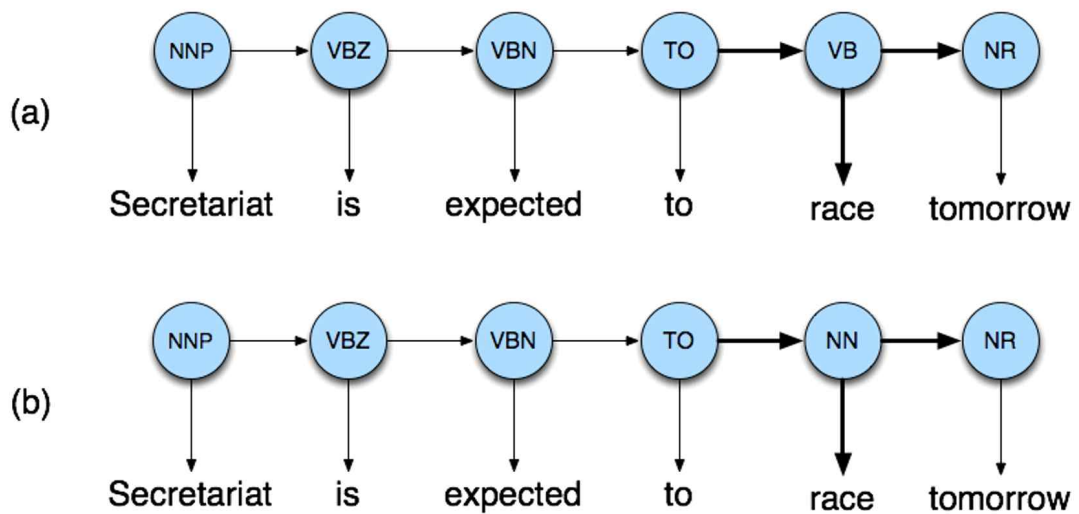
```
print( "Word Tokenizer 출력 :", word_tokenize(rawText))
```

4. 텍스트 전처리에서 필요한 Normalization, Lemmatization, Morphology, Stemming 이 각각 무엇을 의미하는지 사례를 들어 설명하시오.

5. 다음 코드의 실행 결과를 기술하시오.

```
raw = "I am big! It's the pictures that got small."  
print(re.split(r' +', raw))  
print(re.split(r' \W+', raw))  
print(re.findall(r' \w+!|\S\w*', raw))
```

6. 아래 그림에서 단어 'race'의 품사를 결정하는 과정을 기술하시오.

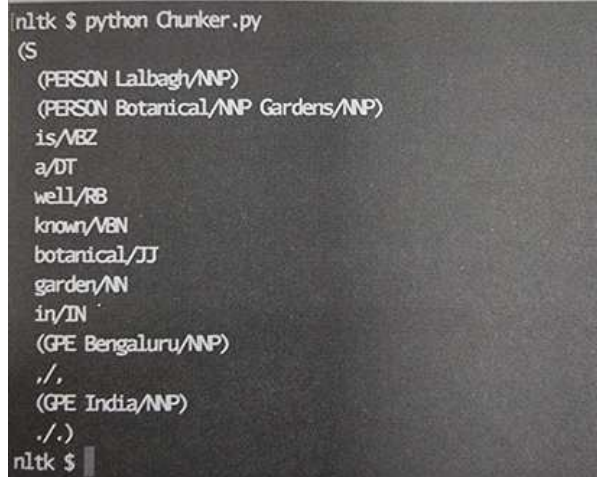


7. 다음 코드의 실행 결과를 기술하시오.

```
PCFG.py

1 import nltk
2 from nltk.parse.generate import generate
3
4 productions = [
5     "ROOT -> WORD [1.0]",
6     "WORD -> P1 [0.25]",
7     "WORD -> P1 P2 [0.25]",
8     "WORD -> P1 P2 P3 [0.25]",
9     "WORD -> P1 P2 P3 P4 [0.25]",
10    "P1 -> 'A' [1.0]",
11    "P2 -> 'B' [0.5]",
12    "P2 -> 'C' [0.5]",
13    "P3 -> 'D' [0.3]",
14    "P3 -> 'E' [0.3]",
15    "P3 -> 'F' [0.4]",
16    "P4 -> 'G' [0.9]",
17    "P4 -> 'H' [0.1]",
18 ]
19
20 grammarString = "\n".join(productions)
21
22 grammar = nltk.PCFG.fromstring(grammarString)
23
24 print(grammar)
25
26 for sentence in generate(grammar, n=10, depth=5):
27     palindrome = "".join(sentence).replace(" ", "")
28     print("String : {}, Size : {}".format(palindrome, len(palindrome)))
29
```

8. 다음과 같은 청킹 결과를 출력하기 위한 코드가 결과 그림 아래 있다. 괄호안에 적당한 코드를 입력하시오.



```
nltk $ python Chunker.py
(S
 (PERSON Lalbagh/NNP)
 (PERSON Botanical/NNP Gardens/NNP)
 is/VBZ
 a/DT
 well/RB
 known/VBN
 botanical/JJ
 garden/NN
 in/IN
 (GPE Bengaluru/NNP)
 ./,
 (GPE India/NNP)
 ./.)
nltk $
```

```
import nltk
text = "Lalbagh Botanical Gardens is a well known botanical garden in Bengaluru,
India."
sentences = nltk.(          )(text)
for sentence in sentences:
    words = nltk.(          )(sentence)
    tags = nltk.(          )(words)
    chunks = nltk.(          )(tags)
    print(chunks)
```

9. 아래와 같은 문법이 있을 때, 문장 “Book that flight”를 bottom-up 방식으로 구문분석하고 그 과정을 설명하여라.

Grammar	Lexicon
$S \rightarrow NP VP$	$Det \rightarrow that \mid this \mid a$
$S \rightarrow Aux NP VP$	$Noun \rightarrow book \mid flight \mid meal \mid money$
$S \rightarrow VP$	$Verb \rightarrow book \mid include \mid prefer$
$NP \rightarrow Pronoun$	$Pronoun \rightarrow I \mid she \mid me$
$NP \rightarrow Proper-Noun$	$Proper-Noun \rightarrow Houston \mid NWA$
$NP \rightarrow Det Nominal$	$Aux \rightarrow does$
$Nominal \rightarrow Noun$	$Preposition \rightarrow from \mid to \mid on \mid near \mid through$
$Nominal \rightarrow Nominal Noun$	
$Nominal \rightarrow Nominal PP$	
$VP \rightarrow Verb$	
$VP \rightarrow Verb NP$	
$VP \rightarrow Verb NP PP$	
$VP \rightarrow Verb PP$	
$VP \rightarrow VP PP$	
$PP \rightarrow Preposition NP$	

10. “I booked a morning flight”를 shift, left, right 연산자를 사용하여 의존 구문 분석하는 과정을 보여라. 하나의 상태는 stack, list, set 으로 구성된다.