

Traitement Automatique du Langage Naturel

TP 1 Analyse Morphologique (lexicale)

Analyse Syntaxique (syntactic parsing)



Grammar denotes the syntactical rules for conversation in natural language.

But in the theory of formal language, grammar is defined as a set of rules that can generate strings. The set of all strings that can be generated from a grammar is called the language of the grammar.

Context Free Grammar:

We are given a Context Free Grammar $G = (V, X, R, S)$, where:

V is a finite set of variables or non-terminal symbols,

X is a finite set of terminal symbols,

R is a finite set of rules,

S is the start symbol, a distinct element of V , and

V and X are assumed to be disjoint sets.

Chomsky Normal Form:

A Context Free Grammar G is in Chomsky Normal Form (CNF) if each rule of G is of the form:

$A \rightarrow BC$ [with at most two non-terminal symbols on the RHS]

or $A \rightarrow a$ [one terminal symbol on the RHS]

Exercice 1 : Nous voulons savoir si les phrases suivantes sont syntaxiquement correctes.

- la pomme mange le garçon
- pomme mange le garçon
- mange le garçon la pomme
- le garçon mange la pomme

on suppose que notre vocabulaire est limite {le : det, la : det, pomme : nom, garçon : nom, mange : verbe} et que notre grammaire est définie par les règles suivantes

$S \rightarrow NP VP$

$NP \rightarrow Det Nom$

$VP \rightarrow Verb NP$

$Det \rightarrow le \mid la \mid les$

$Verb \rightarrow mange$

$Nom \rightarrow pomme \mid garçon$

codez votre solution (from scratch) pour implémenter l'algorithme CYK afin de vérifier la syntaxe des phrases données.

Exercice 2 : en utilisant la bibliothèque Spacy, implémentez un correcteur orthographique et un correcteur grammatical (qui vérifie uniquement la syntaxe)

On ne doit pas confondre le correcteur orthographique et le correcteur grammatical

Le correcteur orthographique (spell checker **مدقق إملائي** analyse morphologique ou lexicale) compare les mots du texte aux mots d'un dictionnaire. Si les mots du texte sont dans les dictionnaires, ils sont acceptés, sinon une ou plusieurs propositions de mots proches sont faites par le correcteur orthographique.

Le correcteur grammatical (Grammatical error checking **المدقق النحوي** analyse syntaxique + analyse sémantique) vérifie que les mots du texte, bien qu'ils soient dans les dictionnaires, sont conformes aux règles de grammaire (accords, ordre des mots, etc.) et aux règles de la sémantique (phrase ayant un sens, absence de confusion d'homophones, etc.).

vous pouvez utiliser ce code pour commencer

```
!pip install contextualSpellCheck -q
#load the small English model, en_core_web_sm
!python -m spacy download fr_core_news_sm
import spacy
import contextualSpellCheck

nlp = spacy.load("fr_core_news_sm")
contextualSpellCheck.add_to_pipe(nlp)
doc = nlp("la pomme mange le garçon")
print(doc.text)
for token in doc: print(token.text, token.pos_, token.dep_)
from spacy import displacy
displacy.render(doc, style="dep", jupyter=True)

# Matcher is initialized with the shared vocab
from spacy.matcher import Matcher
# Each dict represents one token and its attributes
matcher = Matcher(nlp.vocab)
# Add with ID, optional callback and pattern(s)
pattern3 = [{"POS": "DET", "OP": "+"}, {"POS": "NOUN"}]
matcher.add('simple_phrase', [pattern3])
matches = matcher(doc)
# Matches are (match_id, start, end) tuples
for match_id, start, end in matches:
# Get the matched span by slicing the Doc
span = doc[start:end]
print(span.text)
```