Hacettepe University

Statistical Natural Language Processing

Assignment-2 : Context Free Grammar and Parsing

Author:
Efsun Sezer
N16244397

Instructor:
Assoc. Prof. Dr. Burcu CAN

April 11, 2019

1. Language Generation with CFG

A context-free grammar is a set of rules which describes a language by defining the way the strings are put together. It is stronger than finite automata or RE's, but still it cannot identify all possible languages.

CFG consists of terminals, non-terminals and production rules. Derivation of strings in the language of a CFG is achieved by starting with the start symbol (S) and repeatedly replacing some non-terminals by the right side of one of its productions.

1.1. Results

Example of sentences using given grammar:

```
-----Generate sentences and write it to the file-----
-----1 . sentence-----
the floor in every mouse with the sandwich with every president
    in every pickle with every pickle on a pickle under every
    mouse in the sandwich pickled every sandwich on a sandwich
    with every sandwich under a mouse under every mouse on a
    pickle under the mouse
-----2 . sentence-----
a floor pickled every sandwich
-----3 . sentence-----
a pickle under a president with the pickle under every sandwich
    with a pickle with a president in every president pickled the
     pickle
-----4 . sentence-----
every mouse on every mouse in every sandwich in every pickle in
    the floor with every sandwich with the sandwich with every
    pickle wanted every sandwich on the president on a sandwich
-----5 . sentence-----
a floor with the mouse in every pickle on a sandwich with a
    mouse with every sandwich with the president on the president
     under a floor ate a sandwich with every floor in a mouse in
    a mouse with every mouse in a mouse under the sandwich in a
    floor in a mouse with a sandwich under every sandwich on the
    floor with the president in every sandwich with a president
    with a sandwich in a mouse with a pickle in every sandwich on
     a floor on the pickle under the pickle under the pickle with
     every president with a president in the pickle with the
    floor with the president in every sandwich under the floor on
```

every floor in the mouse under the floor in a pickle in a president on a sandwich with every sandwich on the president with every president in the president in the president on a mouse under every president on the sandwich on every sandwich in the mouse in the sandwich on a president on the pickle under a floor in a sandwich on every sandwich on every pickle under every pickle on the sandwich under every pickle under every sandwich with every mouse on a president under the mouse with the sandwich on a mouse with the pickle in every pickle with every president in a floor on the sandwich in a sandwich with the president with a sandwich on the sandwich with the floor under the pickle under the sandwich with the floor with a president with the sandwich with every mouse on a floor with a mouse with the floor in a mouse on every pickle under every president on the pickle in every sandwich in every president on a floor in the pickle with a mouse under a floor under the sandwich in the floor in every pickle in a sandwich with the floor under the floor under a sandwich on a floor in a mouse under every president under every president in the floor in the president in every sandwich under every floor under the mouse in a mouse with a mouse under every pickle under every mouse under a floor in every floor with the pickle under every pickle under a floor in every president on every president with every president under every floor on a sandwich under the floor with a floor in a mouse on every pickle with a president with every floor under the sandwich in every president in a sandwich on the pickle on the president in every president in the president with the president under the sandwich under a president with every pickle with a pickle in every floor in every pickle on every floor in a sandwich with the sandwich in the mouse in the sandwich on every mouse with the floor in every sandwich on every pickle on a sandwich under the mouse in the mouse under every sandwich with every sandwich on every president on a sandwich

———6 . sentence———

the president kissed a pickle on a floor in the president on every mouse in every pickle under a president under a pickle in a mouse in a floor with a president in the pickle with the pickle with every mouse with the sandwich on a sandwich under every mouse in the pickle on every mouse under the sandwich on every floor in every floor on every sandwich under the president on a sandwich with the sandwich on the president under the floor under every sandwich on every pickle with every pickle with every pickle under the pickle

```
    with every floor on the mouse on the sandwich under every
    sandwich under the floor
----7 . sentence----
every sandwich washed the sandwich
----8 . sentence----
a floor pickled a mouse under a pickle on the pickle with every
    pickle under a sandwich in a pickle under a floor under every
     pickle under the sandwich under every mouse with the mouse
    under every mouse on a pickle in every floor in a pickle in a
     president on every mouse with a sandwich on every pickle
    under every sandwich in every sandwich under a sandwich in
    every floor under a president under a pickle on the floor
    under a sandwich in the mouse
----9 . sentence----
every pickle ate every sandwich
```

It is observed that the types of sentences in which the NP structure is dominant are produced. Results also showed that when the grammar types that allow branching are selected, longer sentences are produced. This situation explains why some sentences produced are longer than others.

2. Parsing Sentences with CYK Parser

The Cocke–Younger–Kasami-Algorithm which is known as CYK or CKY is a very popular parsing algorithm for context-free grammars. It is used to determine whether a particular word is part of a language. While doing that, it uses dynamic programming technique and builds a triangular table which causes the algorithm to be named Table Filling Algorithm. The CYK requires context free grammars given in Chomsky Normal Form (CNF) to perform.

2.1. Results

The program code takes sentences and checks whether given sentences belong to given grammar and generate possible parse trees. The results obtained by testing five different sentences are given below.

Example sentence-1 : "the sandwich wanted very old mouse"
Output:

```
Sentence 'the sandwich wanted very old mouse' belongs to the
    given grammar.

Parsing tree of the sentence :

                    Noun->mouse
                Noun
                    Adj->old
            NP
                Det->very
        VP
            Verb->wanted
S
            Noun->sandwich
        NP
            Det->the
```

Example sentence-2 : "a president with a sandwich kissed very floor"
Output:

```
Sentence 'a president with a sandwich kissed very floor' belongs
     to the given grammar.


Parsing tree of the sentence :

                Noun->floor
            NP
                Det->very
        VP
            Verb->kissed
S
                    Noun->sandwich
                NP
                    Det->a
            PP
                Prep->with
        NP
                Noun->president
            NP
                Det->a
```

Example sentence-3 : "very president on a president under very president with the floor with very sandwich with the mouse" wanted the mouse"

Output:

```
Sentence 'very president on a president under very president
    with the floor with very sandwich with the mouse wanted the
    mouse' belongs to the given grammar.
Parsing tree of the sentence :

                Noun->mouse
            NP
                Det->the
        VP
            Verb->wanted
S
                    Noun->mouse
                NP
                    Det->the
            PP
                Prep->with
        NP
                        Noun->sandwich
                    NP
                        Det->very
                PP
                    Prep->with
            NP
                            Noun->floor
                        NP
                            Det->the
                    PP
                        Prep->with
                NP
                                Noun->president
                            NP
                                Det->very
                        PP
                            Prep->under
                    NP
                                    Noun->president
                                NP
                                    Det->a
                            PP
                                Prep->on
                        NP
                                Noun->president
                            NP
                                Det->very
```

Example sentence-4 : " the president with a floor in very president under the mouse wanted a floor"
Output:

```
Sentence 'the president with a floor in very president under the
    mouse wanted a floor' belongs to the given grammar.


Parsing tree of the sentence :

                Noun->floor
            NP
                Det->a
        VP
            Verb->wanted
S
                    Noun->mouse
                NP
                    Det->the
            PP
                Prep->under
        NP
                        Noun->president
                    NP
                        Det->very
                PP
                    Prep->in
            NP
                            Noun->floor
                        NP
                            Det->a
                    PP
                        Prep->with
                NP
                        Noun->president
                    NP
                        Det->the
```

Example sentence-5 : " the floor pickled every mouse on the fine"
Output:

```
Sentence 'the floor pickled every mouse on the fine' does not
   belong to the given grammar.
```

References

[1] The CYK Algorithm
https://www.xarg.org/tools/cyk-algorithm/

[2] Context-Free Grammars
https://people.cs.clemson.edu/~goddard/texts/theoryOfComputation/6a.pdf