Context Free Grammars

Leon Rische

[2020-05-11 Mon 16:15]

Contents

Context free grammars are a concept originating in linguistics but widely used in (theoretical) computer science.

For this section, I'm going to use a bunch of mathematical notation, but I'll try to explain everything as it comes along.

A **set** is collection of elements where each element appears only once. The set of the number from one to four can be written as $\{1, 2, 3, 4\}$.

An **alphabet** is a set with a finite number of elements (the natural numbers 1, 2, 3, ..., would be an example for an infinite set).

A context free grammar has four components:

1. An alphabet T of **terminals** (read: "things where the expansion terminates")
2. An alphabet N of **non-terminals** (read: "things that can expand into other things")
3. A set P of **production rules**
4. The **start symbol** $S \in N$, an element of the set of **non-terminals**

It's important that the terminal and non-terminal alphabets are **disjoint**, i.e. there is no element that is part of both sets.

Production rules have the form $N \to (N \cup T)^\star$. $(N \cup T)$ can be read as "Combine the sets T and N", * is called **Kleene star** and can be read as "repeated zero or more times".

The part on the right of the $\to$ is called **right hand side**, or **RHS**, the part on the left **left hand side** or **LHS**.

A grammar is expanded starting with the start symbol and then replacing each non-terminal with the RHS of one of its rules.

Time for a simple example:

- $T = \{a\}$
- $N = \{A\}$
- $P = \{A \to BBa, B \to bb\}$
- $S = A$

Starting from S, the grammar expands into BBa, then each B expands into bb, so the end result is $bbbba$.

There can be multiple rules with the same **right hand side**, so there are (possibly infinitely) many sequences a grammar can expand to.

- $T = \{a\}$
- $N = \{A\}$
- $P = \{A \to Aa, A \to \varepsilon\}$
- $S = A$

ε is the notation for an empty sequence, so when expanding $A \to \varepsilon$, the A just disappears.

Expanding this grammar, we have two options, either expand A to Aa or to ε.

Expanding Aa further, we have two options again, either expand it to Aaa or to $\varepsilon a = a$.

This can be repeated infinitely many times, so the grammar produces sequences of a of any length. (Remember the Kleene star? We can use it to write that as $a^\star$ instead).