Formal Language (3B)

- Context Free Language

Young Won Lim
6/9/18

Please send corrections (or suggestions) to youngwlim@hotmail.com.

This document was produced by using LibreOffice and Octave.

Regular Language

a **regular language** (a **rational language**) is
a formal language that can be expressed
using a **regular expression**,
in the strict sense

Alternatively, a regular language can be defined
as a language recognized by a **finite automaton**.

The equivalence of **regular expressions** and
finite automata is known as **Kleene's theorem**.

Regular languages are very useful
in input parsing and programming language design.

https://en.wikipedia.org/wiki/Regular_language

Regular Language – Formal Definition

The collection of **regular languages** over an **alphabet** Σ
is defined recursively as follows:

The **empty language** Ø, and the **empty string language** {ε}
are **regular languages**.

For each a ∈ Σ (a belongs to Σ),
the **singleton language** {a} is a **regular language**.

If A and B are **regular languages**,
then A ∪ B (**union**), A • B (**concatenation**),
and A* (**Kleene star**) are **regular languages**.

No other languages over Σ are **regular**.

See regular expression for its syntax and semantics.
Note that the above cases are in effect the defining
rules of regular expression.

https://en.wikipedia.org/wiki/Regular_language

Equivalent Formalism

1. it is the language of a **regular expression** (by the above definition)
2. it is the language accepted by a **nondeterministic finite automaton** (**NFA**)
3. it is the language accepted by a **deterministic finite automaton** (**DFA**)
4. it can be generated by a **regular grammar**
5. it is the language accepted by an **alternating finite automaton**
6. it can be generated by a **prefix grammar**
7. it can be accepted by a read-only **Turing machine**

https://en.wikipedia.org/wiki/Regular_language

Regular Language Example

All **finite** languages are **regular**;

in particular the **empty** string language {ε} = Ø* is **regular**.

Other typical examples include the language
consisting of **all strings** over the **alphabet** {a, b}
which contain an even number of a’s, or
the language consisting of all strings of the form:
several as followed by several b’s.

A simple example of a language
that is **not regular** is the set of strings { $a^n b^n$ | n ≥ 0 }.

Intuitively, it cannot be recognized with a **finite automaton**,
since a **finite automaton** has **finite memory**
and it cannot remember the exact number of a's.

https://en.wikipedia.org/wiki/Regular_language

Context-free Language

In formal language theory, a **context-free language** (**CFL**) is a language generated by a **context-free grammar** (**CFG**).

Context-free grammar

Different context-free grammars can generate the same context-free language.

Intrinsic properties of the language can be distinguished from extrinsic properties of a particular grammar by comparing multiple grammars that describe the language.

https://en.wikipedia.org/wiki/Context-free_language

Context-free Language & Automata

The set of all **context-free languages** is identical
to the set of languages **accepted** by **pushdown automata**,
which makes these languages amenable to parsing.

Further, for a given **CFG**,
there is a direct way to produce a **pushdown automaton**
for the grammar (and thereby the corresponding language),
though going the other way is not as direct.
(producing a **grammar** given an **automaton**)

https://en.wikipedia.org/wiki/Context-free_language

Context-free Language Example

A model **context-free language** is $L = \{ a^n b^n : n \geq 1 \}$,
the language of all non-empty even-length strings,
the entire first halves of which are a's,
and the entire second halves of which are b's.

L is generated by the grammar $S \rightarrow aSb \mid ab$.

This language is not regular.

https://en.wikipedia.org/wiki/Context-free_language

Context-free Language Example

It is accepted by the pushdown automaton

$M = (\{ q0, q1, qf \} , \{ a, b \} , \{ a, z \} , \delta , q0 , z , \{ qf \})$

where δ is defined as follows:

$\delta(q0, a, z) = (q0, az)$
$\delta(q0, a, a) = (q0, aa)$
$\delta(q0, b, a) = (q1, \varepsilon)$
$\delta(q1, b, a) = (q1, \varepsilon)$
$\delta(q1, \varepsilon, z) = (qf, \varepsilon)$

https://en.wikipedia.org/wiki/Context-free_language

Context-free Language Example

Unambiguous CFLs are a proper subset of all CFLs:
there are inherently **ambiguous** CFLs.

An example of an inherently ambiguous CFL is
the union of $\{ a^n b^m c^m d^n \mid n , m > 0 \}$.

This set is **context-free**,
since the **union** of two context-free languages
is always context-free.

But there is no way to unambiguously parse strings
in the (non-context-free) subset $\{ a^n b^n c^n d^n \mid n > 0 \}$
which is the intersection of these two languages.

https://en.wikipedia.org/wiki/Context-free_language

References

[1] http://en.wikipedia.org/
[2]