Formal Languages and Compilers

Matteo Secco

March 2, 2021

Contents

1 Formal Language Theory

Alphabet Σ: any finite set of symbols $\Sigma = \{a_1, a_2, ..., a_k\}$

String: a sequence of alphabeth elements

Language: a set (possibly infinite) of strings

$$\Sigma = \{a, b, c\} \quad L_1 = \{ab, ac\} \quad L_2 = \{ab, aab, aaab, aaaab, ...\}$$

Sentences/Phrases: strings belonging to a language

Language cardinality: number of sentences of the language

$$|L_1| = |\{ab, ab\}| = 2 \quad |L_2| = |\{ab, aab, aaab, aaaab, ...\}| = \infty$$

Number of occurrences of a symbol in a string: $|bbc|_b = 2$, $|bbc|_a = 0$

Length of a string: number of its elements

$$|bbc| = 3 \quad |abbc| = 4$$

String equality: two strings $x = a_1a_2...a_h$ and $y = a_1a_2...a_k$ are equal $\iff$

- have same length: $|x| = |y| \iff h = k$
- elements from left to right coincide: $a_i = b_i \quad \forall i \in \{1..h\}$

1.1 Operations on strings

Concatenation $x = a_1a_2...a_h \wedge y = b_1b_2...b_k \implies x \cdot y = a_1a_2...a_hb_1b_2...b_k$

- associative: $(xy)z = x(yz)$
- length: $|xy| = |x| + |y|$

Empty string ϵ is the neutral element for concatenation: $x\epsilon = \epsilon x = x \, \forall x$.

- length: $|\epsilon| = 0$
- **NB:** $\epsilon \neq \emptyset$

Substrings: if $x = uyv$ then

- y is a substring of x
- y is a proper substring of $x \quad \iff u \neq \epsilon \vee v \neq \epsilon$
- u is a prefix of x
- v is a suffix of y

Reflection: if $x = a_1a_2...a_h$ then $x^R = a_ha_{h-1}...a_1$

- $(x^R)^R = x$
- $(xy)^R = y^R x^R$
- $\epsilon^R = \epsilon$

Repetition: $x^m = \underbrace{xxx...x}_{m \text{ times}}$. Inductive definition:

- $x^0 = \epsilon$
- $x^m = x^{m-1}x \quad \text{if } m > 0$

1.2 Operations on Languages

Reflection: $L^R = \{x | \exists y \left(y \in L \land x = y^R\right)\}$

Prefixes(L): $\{y | y \neq \epsilon \land \exists x \exists z \left(x \in L \land z \neq \epsilon \land x = yz\right)\}$

- **Prefix-free language:** $L \cap Prefixes(L) = \emptyset$

Concatenation: $L'L'' = \{xy | x \in L' \land y \in L''\}$

Power: inductive definition:

- $L^0 = \{\epsilon\}$
- $L^m = L^{m-1}L \quad \text{for } m > 0$
- Consequences:
  - $\emptyset^0 = \{\epsilon\}$
  - $L \cdot \emptyset = \emptyset \cdot L = \emptyset$
  - $L \cdot \{\epsilon\} = \{\epsilon\} \cdot L = L$

Universal language: over alphabet Σ: $L_{\text{universal}} = \Sigma^0 \cup \Sigma^1 \cup ...$

Complement: of L over Σ: $\neg L = L_{\text{universal}} \setminus L$