

2 Defining Languages with Regular Expressions

Basic Language Concepts

An *alphabet* for a language is a finite set of symbols used to form strings in that language.

A *string over an alphabet* is a finite sequence of characters from that alphabet.

Null string is the string with zero characters. Let's use the Greek letter lambda (λ) to represent the null string.

A *language* over an alphabet is any set of strings—finite or infinite—over that alphabet.

ab^3 represents the string abbb.

$(ab)^3$ represents ababab.

$\{a^i \mid i \geq 1\}$ the set of strings a^i for all i greater than or equal to 1.

Operations on Strings and Languages

Concatenation: ab , a^*b^*

Star: a^* , $(ab)^*$, $(a|b)^*$

Union: $a|b$, $a^*|b^*$

$$\begin{aligned} \{a^i b^i \mid i \geq 0\} &= \{\lambda, ab, aabb, aaabbb, \dots\} \\ a^*b^* &= \{\lambda, a, b, aa, ab, bb, \dots\} \end{aligned}$$

$A|B$, the *union* of A and B , is the set of the strings in A or in B or in both. Thus, $A|B = \{a, ab, bb\}$. We will sometimes refer to the union operator (“|”) as the “*or*” operator or the *vertical bar operator*. The union operator is usually represented with the symbol “ \cup ”, but we use “|” because it is available on the standard keyboard.

$A \cap B$, the *intersection* of A and B , is the set of the strings that are in both A and B . Thus, $A \cap B = \{ab\}$.

$\sim A$, the *complement* of A , is the set of all strings in the universe $\{a, b\}^*$ that are not in A . Thus, $\sim A$ contains λ , b , aa , ba , bb , and every string over $\{a, b\}$ of length 3 or more.

Regular Expressions

A *regular expression* is a well-formed finite-length expression in which we are allowed to use only

- \emptyset (which represents the empty set)
- λ (which represents the null string)
- any symbol in our alphabet
- the star, concatenation, and union operations
- parentheses (to show the scope of an operation)

Regular expression	Language defined
\emptyset	$\{ \}$
λ	$\{ \lambda \}$
a	$\{ a \}$
b	$\{ b \}$
ab	$\{ ab \}$
a^*	$\{ \lambda, a, aa, aaa, \dots \}$
b^*	$\{ \lambda, b, bb, bbb, \dots \}$
a^*b^*	$\{ \lambda, a, b, aa, ab, bb, \dots \}$
$a^* b^*$	$\{ \lambda, a, b, aa, bb, aaa, bbb, \dots \}$

No regular expression for $\{ a^i b^i \mid i \geq 0 \}$

$\lambda | ab | aabb | aaabbb \dots$ not a regular express