Languages

Original Slides by C. Busch, LSU.

Language: a set of strings

String: a sequence of symbols
from some alphabet

Example:

Strings: cat, dog, house, computer

Language: {cat, dog, house}

Alphabet: $\Sigma = \{a, b, c, \ldots, z\}$

Languages are used to describe computation problems:

$$PRIMES = \{2,3,5,7,11,13,17,\ldots\}$$

$$EVEN = \{0,2,4,6,\ldots\}$$

Alphabet: $\Sigma = \{0,1,2,\ldots,9\}$

Computation is translated to set membership

Example computation problem:

Is number x prime?

Equivalent set membership problem:

$$x \in PRIMES = \{2,3,5,7,11,13,17,\ldots\}?$$

Alphabets and Strings

An alphabet is a set of symbols

Example Alphabet: $\Sigma = \{a, b\}$

A string is a sequence of symbols from the alphabet

Example Strings

a

ab

$abba$

$aaabbbaabab$

Decimal numbers alphabet: $\Sigma = \{0,1,2,\ldots,9\}$

Example strings:

$102345 \qquad 567463386$

Binary numbers alphabet: $\Sigma = \{0,1\}$

Example strings:

$100010001 \qquad 101101111$

Unary numbers alphabet: $\Sigma = \{1\}$

Unary number:	1	11	111	1111	11111
Decimal number:	1	2	3	4	5

String Operations

$$w = a_1 a_2 \cdots a_n \qquad abba$$

$$v = b_1 b_2 \cdots b_m \qquad bbbaaa$$

Concatenation

$$wv = a_1 a_2 \cdots a_n b_1 b_2 \cdots b_m \qquad abbabbbaaa$$

$$w = a_1 a_2 \cdots a_n \qquad ababaaabbb$$

Reverse

$$w^R = a_n \cdots a_2 a_1 \qquad bbbaaababa$$

String Length

$$w = a_1 a_2 \cdots a_n$$

Length: $|w| = n$

Examples: $|abba| = 4$

$$|aa| = 2$$

$$|a| = 1$$

Length of Concatenation

$$|uv| = |u| + |v|$$

Example: $u = aab, \quad |u| = 3$

$$v = abaab, \quad |v| = 5$$

$$|uv| = |aababaab| = 8$$

$$|uv| = |u| + |v| = 3 + 5 = 8$$

Empty String

A string with no letters is denoted: ε or λ

Acts as a neutral element

Observations: $|\varepsilon| = 0$

$$\varepsilon w = w\varepsilon = w$$

$$\varepsilon abba = abba\varepsilon = ab\varepsilon ba = abba$$

Substring

Substring of string:

A subsequence of consecutive characters

String	Substring
$\underline{ab}bab$	ab
$\underline{abba}b$	$abba$
$ab\underline{b}ab$	b
$a\underline{bbab}$	$bbab$

Prefix and Suffix

string $abbab$

Prefixes	Suffixes
ε	$abbab$
a	$bbab$
ab	bab
abb	ab
$abba$	b
$abbab$	ε

$$w = \underbrace{u}_{\text{prefix}}\underbrace{v}_{\text{suffix}}$$

Exponent Operation

$$w^n = \underbrace{ww\cdots w}_{n}$$

Example: $(abba)^2 = abbaabba$

Definition: $w^0 = \varepsilon$

$$(abba)^0 = \varepsilon$$

The * Operation

Σ^* : the set of all possible strings from alphabet Σ

$$\Sigma = \{a, b\}$$

$$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

The + Operation

Σ^+ : the set of all possible strings from alphabet Σ except ε

$$\Sigma = \{a, b\}$$

$$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

$$\Sigma^+ = \Sigma^* - \{\varepsilon\}$$

$$\Sigma^+ = \{a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

Languages

A language over alphabet Σ
is any subset of Σ^*

Example: $\Sigma = \{a, b\}$

$$\Sigma^* = \{\varepsilon, a, b, aa, ab, ba, bb, aaa, \ldots\}$$

Languages: $\{\}$

$$\{\varepsilon\}$$

$$\{a, aa, aab\}$$

$$\{\varepsilon, abba, baba, aa, ab, aaaaaa\}$$

More Language Examples

Alphabet $\Sigma = \{a, b\}$

An infinite language $L = \{a^n b^n : n \geq 0\}$

$$\left.\begin{array}{l} \varepsilon \\ ab \\ aabb \\ aaaaabbbbb \end{array}\right\} \in L$$

$bbabb \notin L$

$abb \notin L$

Prime numbers

Numbers divisible by 1 and itself

Alphabet $\Sigma = \{0,1,2,\ldots,9\}$

Language:

$$PRIMES = \{x : x \in \Sigma^* \text{ and } x \text{ is prime}\}$$

$$PRIMES = \{2,3,5,7,11,13,17,\ldots\}$$

Even and odd numbers

Alphabet $\Sigma = \{0,1,2,\ldots,9\}$

Languages:

$$EVEN = \{x : x \in \Sigma^* \text{ and } x \text{ is even}\}$$

$$EVEN = \{0,2,4,6,\ldots\}$$

$$ODD = \{x : x \in \Sigma^* \text{ and } x \text{ is odd}\}$$

$$ODD = \{1,3,5,7,\ldots\}$$

Addition (of Unary Numbers)

Alphabet: $\Sigma = \{1, +, =\}$

Language:

$$ADDITION = \{x + y = z : \ \mathrm{x} = 1^{\mathrm{n}}, \mathrm{y} = 1^{\mathrm{m}}, \mathrm{z} = 1^{k}, \mathrm{n} + m = k, n \geq 1, m \geq 1\}$$

$$11 + 111 = 11111 \in ADDITION$$

$$111 + 111 = 111 \notin ADDITION$$

$$ADDITION = \{1+1=11, 1+11=111, 11+1=111, 11+11=1111, \ldots\}$$

Squares (of Unary Numbers)

Alphabet: $\Sigma = \{1, \#\}$

Language:

$$SQUARES = \{x\#y: \quad x = 1^n, y = 1^m, m = n^2\}$$

$$11\#1111 \in SQUARES$$

$$111\#1111 \notin SQUARES$$

$$SQUARES = \{1\#1, 11\#1111, 111\#111111111, \ldots\}$$

Two Special Languages

Empty Language

$$\{\ \} \text{ or } \emptyset$$

Language with Empty String

$$\{\varepsilon\}$$

Size of a language (number of elements):

$$|\{\ \}| = 0$$

$$|\{\varepsilon\}| = 1$$

$$|\{a, aa, ab\}| = 3$$

$$|\{\varepsilon, aa, bb, abba, baba\}| = 5$$

Note that:

Sets $\emptyset = \{\ \} \neq \{\varepsilon\}$

Set size $|\{\ \}| = |\emptyset| = 0$

Set size $|\{\varepsilon\}| = 1$

String length $|\varepsilon| = 0$

Operations on Languages

The usual set operations:

$$\{a, ab, aaaa\} \cup \{bb, ab\} = \{a, ab, bb, aaaa\} \quad \text{union}$$

$$\{a, ab, aaaa\} \cap \{bb, ab\} = \{ab\} \quad \text{intersecti on}$$

$$\{a, ab, aaaa\} - \{bb, ab\} = \{a, aaaa\} \quad \text{difference}$$

Complement: $\overline{L} = \Sigma^* - L$

$$\overline{\{a, ba\}} = \{\varepsilon, b, aa, ab, bb, aaa, \ldots\}$$

Reverse

Definition: $L^R = \{w^R : w \in L\}$

Examples: $\{ab, aab, baba\}^R = \{ba, baa, abab\}$

$$L = \{a^n b^n : n \geq 0\}$$

$$L^R = \{b^n a^n : n \geq 0\}$$

Concatenation

Definition: $$L_1 L_2 = \{xy : x \in L_1, y \in L_2\}$$

Example: $$\{a, ab, ba\}\{b, aa\}$$

$$= \{ab, aaa, abb, abaa, bab, baaa\}$$

Exponent Operation

Definition: $$L^n = \underbrace{LL\cdots L}_{n}$$

$$\{a,b\}^3 = \{a,b\}\{a,b\}\{a,b\} =$$
$$\{aaa,aab,aba,abb,baa,bab,bba,bbb\}$$

Special case: $$L^0 = \{\varepsilon\}$$

$$\{a,bba,aaa\}^0 = \{\varepsilon\}$$

Example

$$L = \{a^n b^n : n \geq 0\}$$

$$L^2 = \{a^n b^n a^m b^m : n, m \geq 0\}$$

$$aabbaaabbb \in L^2$$

Star (Kleene) Closure (*)

All strings that can be constructed from L

Definition: $L^* = L^0 \cup L^1 \cup L^2 \cup \cdots$

Example:

$$\{a, bb\}^* = \left\{ \begin{array}{l} \varepsilon, \\ a, bb, \\ aa, abb, bba, bbbb, \\ aaa, aabb, abba, abbbb, \ldots \end{array} \right\} \begin{array}{l} L^0 \\ L^1 \\ L^2 \\ L^3 \end{array}$$

Positive Closure

Definition: $$L^+ = L^1 \cup L^2 \cup L^3 \cup \cdots$$

Note that: $L^* = L^0 \cup L^+$

$$\{a,bb\}^+ = \left\{\begin{array}{l} a,bb, \\ aa,abb,bba,bbbb, \\ aaa,aabb,abba,abbbb,\ldots \end{array}\right\} \begin{array}{l} L^1 \\ L^2 \\ L^3 \end{array}$$