Strings and Languages

What is an Alphabet?

- An alphabet is a non–empty, finite set of symbols (usually denoted by Σ)
  - Examples of alphabets:
    - {0, 1}
    - $\{\alpha, \beta, \chi, \delta, \phi, \gamma, \eta\}$
    - {a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t , u, v, w, x, y, z}
    - {a}

What is a String?

- A *string over Σ* is a finite sequence (possibly empty) of elements of Σ.
- ϵ denotes the *empty string*, the string with no symbols.
  - Example strings over {a, b}
    - ϵ, a, aa, bb, aba, abbba
  - NOT strings over {a, b}
    - aaaa…., abca

The Length of a String

- The <u>length</u> of a string w, denoted $|w|$, is the number of symbols in the string
  - Example:
    - $|abbab| = 5$
    - $|a| = 1$
    - $|bbbbbbb| = 7$
    - $|\epsilon| = 0$

Concatenation of Strings

- For any alphabet Σ, the set of all strings over Σ is denoted as Σ^*.
- For w, x $\in \Sigma^*$
  - wx is the concatenation of w and x.
    - w = aba, x = bbb, wx = ababbb
    - For all w:
      - $\epsilon w = w \epsilon = w$
  - w^i for an integer i, indicates concatenation of w, i times
    - w = aba, w^3 = abaabaaba
    - For all w, $w^0 = \epsilon$

Reverse of a String

- For a string $w = w_1w_2\ldots w_{n-1}w_n$

  The reverse of the string w^R is the string with the symbols written in reverse:

  $$w^R = w_nw_{n-1}\ldots w_2w_1$$

Some String-related Definitions

- w is a <u>substring</u> of x if there exist y,z $\in \Sigma^*$ (possibly ϵ) such that x = ywz.
  - *car* is a substring of *<u>car</u>nage*, *des<u>car</u>tes*, *vi<u>car</u>*, *<u>car</u>*, but not a substring of charity.
- w is a <u>suffix</u> of x if there exists y $\in \Sigma^*$ such that x = yw.
- w is a <u>prefix</u> of x if there exists y $\in \Sigma^*$ such that x = wy.

What is a Language?

- A language is a set of strings made up of symbols from a given alphabet.
- A language over Σ is a subset of Σ^* (recall that Σ^* is the set of *all* strings over Σ)
  - Example
    - $\{a,b\}^* = \{\epsilon, a, b, aa, bb, ab, ba, aaa, bbb, baa, \ldots\}$
  - Example Languages over {a,b}
    - $\{\epsilon, a, b, aa, bb\}$
    - $\{w \in \{a,b\}^* \mid |w| = 8\}$
    - $\{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}$
    - $\{w \in \{a,b\}^* \mid n_a(w) = 2$ and w starts with b$\}$
    - Σ^* is a language for any alphabet Σ
    - $\varnothing$
    - $\{w \in \{a,b\}^* \mid |w|$ is odd$\}$
    - $\{\epsilon\}$

$n_a(w)$ is the number of a's in string w

Operations on Languages

- Since languages are simply sets of strings, regular set operations can be applied:
  - For languages L_1 and L_2 over Σ
    - $L_1 \cup L_2$ = all strings in L_1 or L_2
    - $L_1 \cap L_2$ = all strings in both L_1 and L_2
    - $L_1 - L_2$ = strings in L_1 that are not in L_2
    - $L' = \Sigma^* - L$

Concatenation of Languages

- If L_1 and L_2 are languages over Σ
  - $L_1L_2 = \{wx \mid w \in L_1 \text{ and } x \in L_2\}$
  - Example:
    - $L_1 = \{hope, fear\}$
    - $L_2 = \{less, fully\}$
    - $L_1L_2 = \{hopeless, hopefully, fearless, fearfully\}$

Concatenation of Languages

- If L is a language over Σ
  - L^k is the set of strings formed by concatenating elements of L, k times.
  - Example:
    - $L = \{aa, bb\}$
    - $L^3 = \{aaaaaa, aaaabb, aabbaa, aabbbb, bbbbbb, bbbbaa, bbaabb, bbaaaa\}$
    - $L^0 = \{\epsilon\}$

Kleene Star Operation

- The set of strings that can be obtained by concatenating any number of elements of a language L is called the Kleene Star, L^*

$$L^* = \bigcup_{i=0}^{\infty} L^i = L^0 \cup L^1 \cup L^2 \cup L^3 \cup L^4 \ldots$$

- ✓ Note that since L^* contains L^0, ε is always an element of L^*

Kleene Star Operation

- The set of strings that can be obtained by concatenating one or more elements of a language L is denoted L^+

$$L^{+} = \bigcup_{i=1}^{\infty} L^{i} = L^{1} \cup L^{2} \cup L^{3} \cup L^{4} \ldots$$

Note that $L^* = L^+ \cup \{\epsilon\}$

Specifying Languages

- How do we specify languages?
  - If language is finite, we can list all of its strings.
    - $L = \{a, aa, aba, aca\}$
  - Using basic language operations
    - $L = \{aa, ab\}^* \cup \{b\}\{bb\}^*$
  - Descriptive:
    - $L = \{w \mid n_a(w) = n_b(w)\}$

- We can also specify languages recursively…

Recursive Definitions

- Definition is given in terms of itself
- Example: factorial
  - 0! = 1
  - n! = n * (n–1)!

  - 4! = 4 * 3!
  - = 4 * (3 * 2!)
  - = 4 * (3 * (2 * 1!))
  - = 4 * (3 * (2 * (1 * 0!)))
  - = 4 * (3 * (2 * (1 * 1)))
  - = 24

Recursive Definitions and Languages

- Languages can also be described by using a recursive definition

1. Initial elements are added to the set (BASIS)

2. Additional elements are added to the set by applying a rule(s) to the elements already in the set (INDUCTION)

3. Complete language is obtained by applying step 2 repeatedly

Recursive Definitions and Languages

- Example: recursive definition of Σ^*

1) $\epsilon \in \Sigma^*$

2) For all $w \in \Sigma^*$ and all $a \in \Sigma$, $wa \in \Sigma^*$

3) Nothing else is in Σ^* unless it can be obtained by a finite number of applications of rules 1 and 2

Recursive Definitions and Languages

- Example: recursive definition of Σ^*
  - Suppose $\Sigma = \{a,b\}$
  - Consider applications of the recursive definition

1. $i=0: \Sigma^* = \{\epsilon\}$
2. $i=1: \Sigma^* = \{\epsilon, a, b\}$
3. $i=2: \Sigma^* = \{\epsilon, a, b, aa, ab, ba, bb\}$
4. $i=3: \Sigma^* = \{\epsilon, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, bbb\}$
5. ….

Recursive Definitions and Languages

- Example: recursive definition of L* (Kleene star of language L)

1. $\epsilon \in L^*$
2. For all $w \in L^*$ and all $x \in L$, $wx \in L^*$
3. Nothing else is in L* unless it can be obtained by a finite number of applications of rules 1 and 2

Recursive Definitions and Languages

- Example: recursive definition of L*
  - Suppose L = {aa,abba}
  - Consider applications of the recursive definition

1. i=0: L* = {ϵ}
2. i=1: L* = {ϵ, aa, abba}
3. i=2: L* = {ϵ, aa, abba, aaaa, aaabba, abbaaa, abbaabba}
4. …

Recursive Definitions and Languages

- Another Example: palindromes over Σ
  - Palindromes read the same way forward and backward
  - First half of the string is a mirror image of the second half
  - E.g. a, b, aa, aba, babbab, bbabb, ϵ

Recursive Definitions and Languages

- Another Example: palindromes over Σ

1. $\epsilon \in$ pal
2. For all $a \in \Sigma$, $a \in$ pal
3. For all $a \in \Sigma$ and all $w \in$ pal, $awa \in$ pal
4. Nothing else is in pal unless it can be obtained by a finite number of applications of rules 1–3

Recursive Definitions and Languages

- Another Example: palindromes over Σ
  - Suppose Σ = {a,b}
  - Consider applications of the recursive definition

1. i=0:pal = {ϵ, a, b}
2. i=1:pal = {ϵ, a, b, aa, bb, aaa, bab, aba, bbb}
3. i=2:pal = {ϵ, a, b, aa, bb, aaa, bab, aba, bbb,aaaa, baab, abba, bbbb, aaaaa, baaab, ababa, bbabb, aabaa, babab, abbba, bbbbb}
4. ...