Strings and Languages

- An alphabet is a finite set of symbols.
  - \{a, b, c, \ldots, x, y, z\}
  - \{\alpha, \beta, \gamma, \ldots, \chi, \psi, \omega\}
  - Ascii, Unicode, \{0, 1\}
• A string is a finite sequence of symbols from an alphabet written in juxtaposition.
  – thecatinthehat
  – ablewasiereisawelba
  – κωυσ
  – helloworld!
  – 01001000100001

• The string containing no symbols is the empty string, denoted by $\epsilon$.

• The set of all strings over alphabet $\Sigma$ is denoted by $\Sigma^*$.
String Operations

• The length of a string is its length as a sequence.
  – length of *abracadabra* is 11
  – denoted by absolute values:
    \(|abracadabra| = 11\)
  – \(|\epsilon| = 0\).

• Substring
  – A string *v* is a substring of string *w* if there are strings *x* and *y* so that
    \(w = xvy\).
  – If \(w = xy\), then
    * *y* is a suffix of *w* and
    * *x* is a prefix of *w*. 
• Concatenation:
  – Let $x = x_1x_2 \ldots x_n$
  – Let $y = y_1y_2 \ldots y_m$,
  – Then $xy = x_1x_2 \ldots x_ny_1y_2 \ldots y_m$.
  – if $x = "foot"$ and $y = "ball"$ then $xy = "football"

• Properties of Concatenation

  $|xy| = |x| + |y|$ \hspace{1cm} \text{length}

  $\epsilon x = x\epsilon = x$ \hspace{1cm} \text{identity}

  $(xy)z = x(yz)$ \hspace{1cm} \text{associativity}
• Exponentiation
  – If \( w \) is a string and \( n \geq 0 \),
  – then
  \[
  w^0 = \epsilon \quad \text{if } i = 0 \\
  w^i = w^{i-1}w \quad \text{if } i > 0
  \]

• Reversal
  – The reversal of string \( w \), denoted by \( w^R \), is defined as
  \[
  w^R = w \quad \text{if } w = \epsilon \\
  = au^R \quad \text{if } w = ua \ a \in \Sigma
  \]
Languages

- Let $\Sigma$ be an alphabet.
- A language over $\Sigma$ is a set of strings.
- Examples
  - $\Sigma^*$,
  - $\emptyset$,
  - the palindromes over $\{0, 1\}$,
  - the set of strings over $\{a, b\}$ that start with $a$. 
Operations on Languages

- union, intersection, difference as sets
- complement is with respect to $\Sigma^*$.
- concatenation

$$L_1L_2 = \{w_1w_2 \mid w_i \in L_i, i = 1, 2\}$$

- exponentiation

$$L^0 = \{\epsilon\}$$

$$L^{i+1} = L^iL \text{ for } i > 0$$
• closure
  If $L$ is a language, then $L^*$ is the set of all strings formed by concatenating zero or more strings from $L$.
  $$L^* = \bigcup_{i=0}^{\infty} L^i$$

• positive closure
  If $L$ is a language, then $L^+$ is the set of all strings formed by concatenating one or more strings from $L$.
  $$L^+ = \bigcup_{i=1}^{\infty} L^i$$