CS360 Theory of Computing

Alphabets, Strings, Languages, & Classes of Languages

Alphabets & Strings

Alpabet

- A finite set of symbols
- Usually denoted by ∑

Example: a binary alphabet, ∑ = {0, 1}

String (words)

- A sequence of alphabet symbols in an order

Example: 1010110

String Operations

Length of a string: $\boldsymbol{x} = |\boldsymbol{x}|$

- All our strings will have finite length

Empty string: denoted by ε (sometimes Λ)

- Length of empty string: $|\varepsilon| = 0$

Power Op:

$\boldsymbol{x}^k$ = k copies of a string concatenated

Example: $\boldsymbol{x} = 010$, $\boldsymbol{x}^3 = 010010010$

Convention: $\boldsymbol{x}^0 = \varepsilon$

Prefixes, Suffixes & Substrings

Prefix: first k chars of a string ***v*** for some value k

Suffix: last k chars of a string ***v*** for some value k

Substring: a string ***u*** that is found as consecutive characters in a string ***v***

Concatenation:

Formal: ***x*** is a prefix of ***y*** if ***y*** = ***x•z*** for some ***z***

- Often we will not write the "• "

More String Operations

Counting members:

$\boldsymbol{n_b}(\boldsymbol{x})$ = # of times alphabet symbol $\boldsymbol{b}$ is found in string $\boldsymbol{x}$

Reverse strings:

$\boldsymbol{rev}(\boldsymbol{x}) = \boldsymbol{x}$ written in reverse order (also, $\boldsymbol{x^r}$)

Palindrome:

$\boldsymbol{x}$ is a palindrome if $\boldsymbol{x} = \boldsymbol{rev}(\boldsymbol{x})$

Languages

$\sum^*$ is the set of all finite strings over alphabet $\sum$

A language **L** is a set of strings over $\sum$

- A subset of $\sum^*$, possibly $\sum^*$ itself
- Ø denotes the empty set over any $\sum$

Note: $\emptyset \neq \{\varepsilon\}$

Complement: **L'** = $\sum^*$ - **L**

Concatenation of Languages

If **L** and **M** are languages, then:

$\mathbf{LM} = \{\boldsymbol{xy} \mid \boldsymbol{x} \in \mathbf{L} \text{ and } \boldsymbol{y} \in \mathbf{M}\}$

Finite Power

$\mathbf{L}^k$ is k copies of **L** concatenated

Example: $\mathbf{L}^3 = \mathbf{LLL}$

- $\mathbf{L}^0 = \{\varepsilon\}$
- $\mathbf{L}^1 = \mathbf{L}$

Kleene Star

$\mathbf{L}^* = \bigcup_{i=0}^{\infty} \mathbf{L}^i$ and $\mathbf{L}^+ = \bigcup_{i=1}^{\infty} \mathbf{L}^i$

If **L** = {a, b} then:

L* = {ε, a, b, aa, ab, ...} i.e. **L*** includes ε

$\mathbf{L}^+$ = {a, b, aa, ab, ...}

Note: **L**** = **L***

L* is **closed** under Kleene Star.

Class of Languages

- A set of languages
- A set of sets (usually something in common)

Note: strings are members of languages, **not** members of classes of languages.

Idea is to describe other languages

- Also have the normal set operations: union, intersection, difference, complement, etc.

Example: **M** = (**L** ∩ **N**)***OP**

How do we test membership in **M**?

Each member of **M** consists of

- A finite # of strings (0 or more) from **L** ∩ **N** concatenated together
- Followed by a string from **O**
- Followed by a string from **P**