Compiler Construction

Lecture # 07

Mr. Usman Wajid

usman.wajid@nu.edu.pk

National University
of Computer & Emerging Sciences

Lexical Analysis: Specification of Tokens

Countable Set

It is a finite set such as rational numbers whose elements can be counted

Example, $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Uncountable Set

A set whose elements ares countable such as real numbers

Example, $\{\ 0, 0.1, 0.12, 0.13, \ldots, 5\}$ 0 and 1 is uncountable

Lexical Analysis: Strings and Languages

Alphabet

It is a finite set of symbols. It is represented by Σ

Example: $\Sigma = \{0,1\}$

String or Word

A string over an alphabet is a finite sequences of symbols from that alphabet.

Example: Strings over $\{0,1\}$: ϵ, 0,1,11,00, 11101010

A language

It is a countable set of strings over the alphabet

Lexical Analysis: Strings and Languages continued

Concatenation of Strings

The concatenation of strings `s` and `t` is the string formed by appending the string `t` to `s`,i.e., `st`

Length of a String

It is the number of symbols in the string

Lexical Analysis: Strings and Languages continued

A proper prefix:

A proper prefix of string S is any string obtained by removing symbols (except zero or S itself) from the end of S.

Example, ban is a prefix of banana

A proper suffix:

A proper prefix of string S is any string obtained by removing symbols (except zero or S itself) from the beginning of S.

Example, nana is a suffix of banana

Substring

It is obtained by deleting any prefix and any suffix from s

Example, ban and nana are a sub-string of banana

Lexical Analysis: Strings and Languages continued

A proper subsequence

Any string by deleting any symbols (except zero of S itself) from s

Example, baan is a subsequence of banana

Lexical Analysis: Operation on Languages

Union

The union of L1 and L2 consist of all strings in either L1 or L2

Example, L1={a,b}, L2={c} then L1 $\cup$ L2 ={ϵ, a, b, c}

Concatenation

The concatenation of L1 and L2 is the set of strings st where s ia string of L1 and t is string of L2

Example, L1={a,b}, L2={c} then L1 $\cup$ L2 ={ϵ, ac, bc}

Lexical Analysis: Operation on Languages continued . . .

- The ***Kleene* closure** of L, denoted L^* is $L^0 UL^1 UL^2 \ldots$
- The **positive closure** of L, denoted by L^+ is $L^1 UL^2 \ldots$
- Example,
  $\{a, b\}^* is \{\epsilon, a, b, aa, bb, ab, baaaa, bbb, \ldots\}$
  $\{a, b\}^+ is \{a, b, aa, bb, ab, baaaa, bbb, \ldots\}$

Lexical Analysis: Regular Expressions

- $(a|b)^*$ denotes the set of all strings consisting of zero or more instances of a or b
  Example: $\{\epsilon, a, b, aa, ab, ba, bb, aaa, \dots\}$

  Another regular expression for the same language is
  $(a^*b^*)^*$

- $a|a^*b$ denotes the language $\{a, b, ab, aab, aab, \dots\}$